MULTISCALE TOPOLOGICAL ANALYSIS FOR MAPPING MOLECULAR ENVIRONMENTS AND DYNAMICS

Luca Panconi

Supervisor: Dylan Owen

A thesis submitted to the University of Birmingham for the degree of DOCTOR OF PHILOSOPHY

Institute of Immunology and Immunotherapy
College of Medical and Dental Sciences
The University of Birmingham

September 2024

Declaration of Authorship

I hereby declare that the contents of this thesis are my own original works and have not been submitted in any capacity for a separate degree or qualification at the University of Birmingham or any other institution. This dissertation is comprised fully of my own work and contains nothing which is the outcome of collaborations with others, except where specified in the text and in the following instances:

- **Chapter 2** (**Figures 2.8, 2.9, 2.10, 2.13** and **2.15**): *C gattii* cell culture was performed by Robin C. May. HEK293 cell culture and *S. pombe* cell culture were undertaken by Maria Makarova. *C. gattii*, *HEK293* and *S. Pombe* confocal microscopy data were acquired by Maria Makarova. Jurkat T cell culture was by Alexander J. Collins. Jurkat T cell confocal microscopy data was acquired by Alexander J. Collins.
- **Chapter 3** (**Figures 3.9** and **3.10**): Jurkat T cell culture, CD3ζ DNA-PAINT data acquisition and qPAINT post-processing were undertaken by Olivia Dalby, Cecilia Zaza and Sabrina Simoncelli.
- **Chapter 4** (**Figures 4.6** and **4.10**): GUVs were prepared for di-4-ANEPPDHQ PAINT by Maria Makarova. Daniel J. Nieves optimised generation of membrane patches. Synthetic membrane di-4-ANEPPDHQ PAINT data was acquired by Jonas Euchner using an approach developed by Daniel J. Nieves. RAMA27 cells were cultured and prepared for di-4-ANEPPDHQ PAINT by Daniel. J. Nieves and Jonas Euchner. RAMA27 di-4-ANEPPDHQ PAINT data was acquired by Jonas Euchner. Membrane order mapping software was written by Daniel J. Nieves.

Acknowledgements

Firstly, I can't give enough thanks to my supervisor, Dylan Owen, who gave consistent and excellent advice, even though I often ignored it. He always seemed to find the balance between pushing me in the right direction and watching me make my own mistakes – which was exactly what kept me motivated. Having a space to explore my own research really made all the difference, and I couldn't imagine doing my PhD anywhere else.

I want to thank my many collaborators, reviewers, lecturers and colleagues, whose insights have shaped my work into what it is today. A big thank you to all members and honorary members of the Owen Lab, but in particular to Masha and Dan, who saw me through most of my studies. Many thanks to Juliette, and the brand new CMCB team, for getting me out of England and keeping me entertained in Stockholm.

I would like to thank the various institutes and funding bodies that supported me through my research. Most notably, to the Microscopy and Imaging Services at Birmingham University (MISBU) for microscope access and training. And of course, I must acknowledge funding from Oxford Nanoimaging (ONI) and the Engineering and Physical Sciences Research Council (EPSRC), who provided my studentship through the University of Birmingham CDT in Topological Design.

Further to that, I would like to thank all the committee members, lecturers, researchers, administrators, caterers and general staff at the CDT for keeping the whole operation running smoothly. And of course, to the software team at ONI for the mild (but well-deserved) bullying they imparted on me during my internship – sorry it wasn't enough to keep me humble. Despite my best efforts, they made me feel welcome with the team and in the company, and I couldn't have asked for a better group of people to show me what a real job feels like.

To my many biology and mathematics teachers, for whom I'm sure I was an insufferable student. I hope this thesis is evidence enough that I was, at least in part, paying attention. Special thanks goes to the baristas working in the hospital cafe across the road, who would often make my day that little bit easier by remembering my order for me, since they had all heard it enough times anyway.

To my friends, who insisted on sticking by me even as I disappeared into research – I can't overstate how important you all were in keeping me grounded, and keeping me going. To Joe, for showing me that you don't need to do a PhD to be intelligent, and to Jack, for showing me that you don't need to be intelligent to do a PhD. To Zak, for always being a smart guy with something stupid to say, and to Amy, for every one of our overpriced midday coffees. I never laughed quite as much before I met you all.

Of course, no acknowledgement would be complete without an obligatory thank you to my parents, without whom I literally wouldn't be here. In this case, however, it is the most genuine thanks I have to offer – since they are, actually, the best parents in the world. And of course, I want to thank all my family, for pretending to understand my ramblings and nodding along even though you (rightfully) had no idea what I was talking about. You all made me feel like I was genuinely doing something useful with my life. And for that, you were always my reason.

Abstract

Modern fluorescence microscopy techniques can visualise biological samples from micro- to nano-scale resolutions. Conventional microscopy produces images of cellular structures, while single molecule localisation microscopy (SMLM) localises individual molecules, which are represented as spatial or marked point pattern data. Depending on the imaging target and modality, the data arising from fluorescence microscopy can vary in architecture and present features with irregular or unpredictable geometries, which can be challenging to quantify. Topological data analysis is invariant of geometry and lends itself well to this form of feature extraction, but is yet to see widespread adaptation in quantifying biophysical properties of cellular structures. The cell plasma membrane is of particular interest, as the dynamic reorganisation of transmembrane proteins plays a crucial role in regulating cell signalling. Dysfunction in this process is associated with several human disorders, such as autoimmune diseases and cancer. Further, it is hypothesised that this reorganisation is influenced by the biophysical properties of the membrane, such as lipid composition. As such, this thesis concerns the development of novel topological data analysis techniques for feature extraction in image and point pattern data, with an emphasis on investigating membrane properties across acquisition scales. For conventional fluorescence microscopy, we present a topological image analysis tool (TOBLERONE) for cell and organelle segmentation. Then, we produce a framework for agent-based modelling of molecular aggregation on the plasma membrane (ASMODEUS), which simulates transmembrane protein dynamics. Furthermore, we introduce a software package (PLASMA) for partitioning marked point patterns and identify nano-scale lipid heterogeneity in RAMA27 SMLM data. These techniques may yield a promising avenue for mapping multiscale membrane properties.

List of Publications

The methods, data and results conveyed in this thesis have, at least in part, been communicated in the works below:

Journal Publications

1. **L. Panconi**, M. Makarova, E. R. Lambert, R. C. May, D. M. Owen. Topology-based fluorescence image analysis for automated cell identification and segmentation. *Journal of Biophotonics* 16 (3), 2023. DOI: https://doi.org/10.1002/jbio.202200199.
2. **L. Panconi**, C. D. Lorenz, R. C. May, D. M. Owen, M. Makarova. Phospholipid tail asymmetry allows cellular adaptation to anoxic environments. *Journal of Biological Chemistry* 299 (9), 2023. DOI: https://doi.org/10.1016/j.jbc.2023.105134.
3. **L. Panconi**, D. M. Owen, J. Griffié. Cluster analysis for localisation-based data sets: dos and don'ts when quantifying protein aggregates. *Frontiers in Bioinformatics* 3, 2023. DOI: https://doi.org/10.3389/fbinf.2023.1237551.
4. **L. Panconi**, A. Tansell, A. J. Collins, M. Makarova, D.M. Owen. Three-dimensional topology-based analysis segments volumetric and spatiotemporal fluorescence microscopy. *Biological Imaging* 3 (1), 2023. DOI: https://doi.org/10.1017/S2633903X23000260.

Books

1. **L. Panconi**, D. Nieves, M. Makarova, D. M. Owen. Methods for Imaging Cell Membranes (1st ed.). *CRC Press*, 2023. DOI: https://doi.org/10.1201/9781003273745.

Manuscripts in Preparation

1. **L. Panconi**, J. Euchner, M. Makarova, D-P. Herten, D.M. Owen, D. J. Nieves. Mapping membrane biophysical nano-environments. Manuscript under revision with *Nature Communications*.
2. **L. Panconi**, O. Dalby, C. Zaza, D.M. Owen, S. Simoncelli, J. Griffié. A generalised molecule distribution simulator for tracking transmembrane protein aggregation. Manuscript in preparation.

Code Availability

The original code and data conveyed in this thesis is available in the following repositories:

- Topological Boundary Line Estimation using Recurrence of Neighbouring Emissions (TOBLERONE): Panconi, L. (2024). toblerone (Version 1.0.0) [Computer software]. https://doi.org/10.5281/zenodo.12627127.
- Agent-based Spatiotemporal Molecular Distributions Evolving Under Simulation (ASMODEUS): Panconi, L. (2024). asmodeus (Version 1.0.0) [Computer software]. https://doi.org/10.5281/zenodo.12627125.
- Point Label Analysis of Super-resolved Marked Attributes (PLASMA): Panconi, L. (2024). PLASMA (Version 1.0.0) [Computer software]. https://doi.org/10.5281/zenodo.12627109.

Table of Contents

Chapter 1: Introduction 1
1.1 The plasma membrane 2
1.1.1 Plasma membrane structure and function 2
1.1.2 The lipid raft hypothesis 8
1.1.3 Interplay between lipid rafts and membrane biophysical properties 16
1.1.4 Techniques to study membrane biophysical properties and lipid order 20
1.1.5 Polarity-sensitive fluorescent probes 25
1.2 Microscopy 29
1.2.1 Introduction to fluorescence microscopy and justification for use 30
1.2.2 Single molecule localisation microscopy 34
1.2.3 DNA-PAINT 37
1.2.4 Localisation and resolution 40
1.2.5 Artefact correction 43
1.2.6 Super-resolution image reconstruction 45
1.2.7 Machine learning-assisted SMLM 46
1.2.8 Insights into lipid rafts from super-resolution microscopy 47
1.3 Spatial statistics 48
1.3.1 Spatial point patterns 48
1.3.2 Spatially-descriptive statistics 48
1.3.3 Overview of cluster analysis and unsupervised machine learning 51
1.3.4 Geometric cluster analysis 54
1.3.5 Density-based cluster analysis 55
1.3.6 Machine and deep dearning methods 57
1.3.7 Topological data analysis 59
1.3.8 Introduction to marked point patterns 64
1.3.9 Existing methods in MPP analysis 64
1.4 Outstanding questions in the field 66
1.5 Thesis objectives 68
Chapter 2: Topological image analysis 70
2.1 Introduction 71
2.2 Materials and methods 74
2.3 Results 87
2.3.1 2D segmentation of simulated images with TOBLERONE 87
2.3.2 2D segmentation of fluorescence microscopy data with TOBLERONE 89

2.3.3 3D segmentation of simulated z-stacks with 3DTOBLERONE 92
2.3.4 3DTOBLERONE for volumetric cell segmentation 95
2.3.5 Dynamic t-stack segmentation with tempTOBLERONE 97
2.3.6 TIA for cell tracking and trajectory mapping 99
2.4 Discussion 101
2.4.1 Summary of results 101
2.4.2 Considerations for the use of TOBLERONE in cell segmentation 103
2.4.3 Concluding remarks 104
Chapter 3: *In silico* protein aggregation dynamics 106
3.1 Introduction 107
3.2 Materials and methods 110
3.3 Results 122
3.3.1 Agent-based simulations conserve cluster properties 122
3.3.2 Modelling receptor nucleation 125
3.3.3 Simulating CD3ζ distributions in pre- and post-activation T cells 128
3.3.4 Modelling interacting molecular species 130
3.3.5 Simulating varied dynamics across multiple populations 132
3.4 Discussion 133
3.4.1 Simulating static protein maps with ASMODEUS 135
3.4.2 Modelling signal digitisation and molecular interactions with ASMODEUS 136
3.4.3 Concluding remarks 137
Chapter 4: Marked point pattern analysis 139
4.1 Introduction 140
4.2 Materials and methods 142
4.3 Results 152
4.3.1 Di-4-ANEPPDHQ-based ratiometric PAINT for generating marked point data 152
4.3.2 Detecting domains in simulated marked point patterns and determining statistical significance 153
4.3.3 Demonstration of P-Check on artificial membrane data 155
4.3.4 Detecting domains in marked point patterns from RAMA27 data 157
4.3.5 Quantitative mapping of domains in simulated marked point patterns ... 160
4.3.6 Measuring nanoscale membrane order in live cell membranes with PLASMA 164
4.4 Discussion 166

4.4.1 Results from ratiometric PAINT data of GUVs and RAMA27 cells 167
4.4.2 Applicability of PLASMA and advances in bioinformatics 168
4.4.3 Concluding remarks 168
Chapter 5: Discussion and conclusion 170
5.1 Summary of results 171
5.2 Suitability of methods 173
5.2.1 TOBLERONE 173
5.2.2 ASMODEUS 175
5.2.3 PLASMA 177
5.3 Long-term outlook 179
5.4 Concluding remarks 182
Bibliography 184
Appendix 206
7.1 Formulas 206
7.2 Schematics 210
7.3 Pseudocode 215
7.2.1 2DTOBLERONE 215
7.2.2 3DTOBLERONE 217
7.2.3 tempTOBLERONE 219
7.2.4 ASMODEUS 222
7.2.5 P-Check 224
7.2.6 JOSEPH 225

List of Figures

Figure 1.1: Lipids in the plasma membrane 3
Figure 1.2: Diagram of a cell plasma membrane 5
Figure 1.3: Diagram of cholesterol 6
Figure 1.4: Saturated lipids and cholesterol interact more favourably with each other than with unsaturated lipids 10
Figure 1.5: A lipid raft within a mammalian cell plasma membrane 12
Figure 1.6: The asymmetric lipidome of the plasma membrane 19
Figure 1.7: Alexa Fluor 488 24
Figure 1.8: An example polarity-sensitive probe, di-4-ANEPPDHQ 28
Figure 1.9: Optical diagram of a conventional fluorescence microscope 31
Figure 1.10: The workflow of STORM 37
Figure 1.11: An example of SMLM supplemented by machine learning 39
Figure 1.12: Schematic of filtering and localisation workflow 42
Figure 1.13: The impact of drift, and correction with fiducial markers 43
Figure 1.14: Multiple blinking 45
Figure 1.15: Example spatial point patterns 49
Figure 1.16: Plots of the Ripley's H function 50
Figure 1.17: A schematic of k-means clustering 55
Figure 1.18: A schematic of the DBSCAN cluster analysis algorithm 56
Figure 1.19: Schematic of persistent homology 60
Figure 1.20: A schematic of ToMATo clustering 62
Figure 1.21: Examples of marked point patterns 65
Figure 2.1: Schematic of TOBLERONE 81
Figure 2.2: Topological representations of a z-stack 83
Figure 2.3: The impact of changing persistence on the segmentation identified by TOBLERONE 84
Figure 2.4: Topological decompositions of videos, or t-stacks 85
Figure 2.5: Tracking component segmentation over time 86
Figure 2.6: Segmentation of exemplar image data 88
Figure 2.7: Impact of Gaussian blur and noise on sensitivity and specificity of TOBLERONE 90
Figure 2.8: Common failure cases for TOBLERONE 92
Figure 2.9: Results of segmentation algorithms on experimental data 93
Figure 2.10: TOBLERONE as a tool for identifying variations in membrane lipid order in GP images of *C. gattii* cells 94
Figure 2.11: Exemplar cases of 3D segmentation 95
Figure 2.12: Results of 3DTOBLERONE segmentation on simulated data 96
Figure 2.13: 3D segmentation algorithms on experimental data 97
Figure 2.14: Results of tempTOBLERONE segmentation on simulated data 98
Figure 2.15: Applications of tempTOBLERONE to cell data 100
Figure 3.1: Simulating PAD with ASMODEUS 116
Figure 3.2: Example simulation 119
Figure 3.3: Convergence of cluster properties 120

Figure 3.4: Simulation outputs under varying target cluster radii (r_t) and points per cluster (p_t)..... 121
Figure 3.5: The impact of increasing background to cluster ratio..... 123
Figure 3.6: Exemplar failure cases for ASMODEUS..... 124
Figure 3.7: Modelling receptor nucleation..... 126
Figure 3.8: Estimated probabilities of converging to clusters..... 128
Figure 3.9: Overview of DNA-PAINT imaging and qPAINT analysis..... 129
Figure 3.10: Simulated T cell activation via induced nucleation..... 131
Figure 3.11: Multiple population modelling..... 133
Figure 3.12: Proof-of-concept for modelling protein interactions..... 134
Figure 4.1: Schematic of P-Check..... 148
Figure 4.2: Workflow of JOSEPH..... 149
Figure 4.3: Techniques for parameter estimation..... 151
Figure 4.4: The results of P-Check on simulated MPP data sets with increasing overlap between the background and domain GP distributions..... 154
Figure 4.5: The results of P-Check on simulated MPP data sets with increasing domain radii..... 156
Figure 4.6: Applications to synthetic membranes..... 158
Figure 4.7: The results of JOSEPH on simulated MPP data sets with increasing overlap between the background and domain GP distributions..... 159
Figure 4.8: The results of JOSEPH on simulated data of varied domain size..... 162
Figure 4.9: The results of JOSEPH on simulated MPP data sets, varying the proportion of points assigned to domains..... 163
Figure 4.10: Live cell membrane results..... 165
Figure 7.1: Conceptual diagram of 2D and 3D TOBLERONE algorithm..... 210
Figure 7.2: Conceptual diagram of tempTOBLERONE algorithm..... 211
Figure 7.3: Conceptual diagram of ASMODEUS algorithm..... 212
Figure 7.4: Conceptual diagram of P-Check algorithm..... 213
Figure 7.5: Conceptual diagram of JOSEPH algorithm..... 214

List of Tables

Table 2.1: Excitation wavelength and fluorescence channels for each dye. 76
Table 3.1: Parameters used to simulate target distributions. 115
Table 4.1: Input parameters for simulated data sets and the range of values they were randomly generated from. 145

Abbreviations

2OHOA – 2-Hydroxyoleic acid.

ABM – Agent-based model.

AFM – Atomic force microscopy.

AI – Artificial intelligence.

ASMODEUS – Agent-based Spatiotemporal Molecular Distributions Evolving Under Simulation**.**

BOP – Biogenesis protein.

CD – Cluster of differentiation.

CRLB – Cramér-Rao lower bound.

CSR – Completely spatially random.

DBSCAN – Density-based spatial clustering of applications with noise.

DC – Dendritic cell.

DECODE – Deep context dependent.

DFT – Discretised Fourier transform.

DMEM – Dulbecco's modified Eagle medium.

DMSO – Dimethyl sulfoxide.

DOPC – Dioleoylphosphatidylcholine.

DPPC – Dipalmitoylphosphatidylcholine.

dSTORM – Direct stochastic optical reconstruction microscopy.

ECD – Extracellular domain.

ECM – Extracellular medium.

EM – Electron microscopy.

EM-CCD – Electron-multiplying charge-coupled device.

FBS – Fetal bovine serum.

FGFR – Fibroblast growth factor receptor.

FM – Fluorescence microscopy.

GP – Generalised polarisation.

GPMV – Giant plasma membrane vesicle.

GUV – Giant unilamellar vesicle.

HEK – Human embryonic kidney.

HILO – Highly inclined and laminated optical sheet.

HPAEC – Human pulmonary artery endothelial cells.

ITO – Indium tin oxide.

IoU – Intersection over union.

JOSEPH – Justification of separation by employed persistent homology.

KL – Kullback-Liebler.

LAT – Linker for activation of T cells.

MβCD – Methyl-β-cyclodextrin.

MINFLUX – Minimal fluorescence photon fluxes.

ML – Machine learning.

MLE – Maximum likelihood estimation.

ModLoc – Modulation localisation.

MPP – Marked point pattern.

NA – Numerical aperture.

NMR – Nuclear magnetic resonance.

PAD – Protein aggregation dynamics.

PAINT – Point accumulation for imaging nanoscale topography.

PALM – Photoactivated localisation microscopy.

PBS – Phosphate buffer saline.

PC – Phosphatidylcholine.

PCF – Pair correlation function.

PD1 – Programmed death-1.

PDL1 – Programmed death ligand-1.

PE – Phosphatidylethanolamine.

PenStrep – Penicillin and streptomycin.

PFA – Paraformaldehyde.

PH – Persistent homology.

PLASMA – Point label analysis of super-resolved marked attributes.

PLC – Peptide loading complex.

PS – Phosphatidylserine.

PSF – Point spread function.

RAMA27 – Rat mammary fibroblast.

ROI – Region of interest.

RPMI – Roswell Park Memorial Institute.

sCMOS – Scientific complementary metal-oxide semiconductor.

SM – Sphingomyelin.

SMLM – Single molecule localisation microscopy.

SNR – Signal-to-noise ratio.

STORM – Stochastic optical reconstruction microscopy.

TCR – T cell receptor.

TDA – Topological data analysis.

TIA – Topological image analysis.

TIRF – Total internal reflection.

TOBLERONE – Topological boundary line estimation using recurrence of neighbouring emissions.

ToMATo – Topological mode analysis tool.

YES – Yeast extract with supplements.

YPD – Yeast peptone dextrose.

Chapter 1: Introduction

The focus of this thesis is on the development of topological data analysis techniques for processing image and marked point pattern data, with applications to investigating the cell plasma membrane. This chapter serves as an introduction to the properties of the plasma membrane, the microscopy systems used for visualising these properties, and techniques for analysing the data which arise from those systems. Here, we discuss biophysical membrane properties, taking lipid order as a specific example, and introduce the concept of the lipid raft hypothesis. An overview of membrane phase separation and the role of spatiotemporal molecular organisation is given. Further, we explore the principles of fluorescence microscopy (FM), with an emphasis on SMLM and its applications in quantifying properties of the plasma membrane. We discuss methods for processing image and point pattern data arising from FM, focussing on spatially-descriptive statistics and methods of cluster analysis. Fundamentally, we explore existing topological data analysis techniques, and in particular persistent homology, which underpins the methods presented herein. Specifically, in **Chapter 2**, we present an implementation of persistent homology for segmenting images from conventional FM. Then, in **Chapter 3**, we develop a computational framework for modelling transmembrane protein aggregation. This model could serve a dual purpose: to simulate static protein maps which recapitulate the topological properties seen in experimental data, and to predict protein aggregate behaviour in dynamic systems. Finally, in **Chapter 4**, we introduce algorithms for identifying and extracting topological features in marked point pattern data. Primarily, these methods can partition marked point patterns into clusters defined by both a given spatial scale and a quantifiable distribution of marked values. However, in this chapter, we first explore the motivation of this work with a discussion on the plasma membrane.

1.1 The plasma membrane

1.1.1 Plasma membrane structure and function

Membrane composition and structure

The plasma membrane is the barrier that separates the aqueous compartments which form the cell interior from the surrounding extracellular medium. While the full extent of the plasma membrane's function varies across cell types, its primary role is for maintaining cell structure and survival[1]. The membrane is the principle site for conducting cell-cell communication and regulating interactions with the surrounding medium[1]. This is owed to its highly specialised nano-environment, comprising a heterogeneous ecosystem of biological molecules[1]. One of the most prevalent biological molecules in the plasma membrane is the phospholipid, which comprises a phosphate head, with an attached variable group, connected to two acyl lipid (or fatty acid) tails by a glycerol or alcohol backbone (**Figure 1.1a**)[2]. The lipidome of mammalian plasma membranes is primarily composed of four phospholipids, phosphatidylcholine (PC), phosphatidylinositol (PI), phosphatidylethanolamine (PE) and phosphatidylserine (PS), as well as the sphingolipid sphingomyelin (SM)[2]. Here, the variable group dictates phospholipid type – for example, PC presents a choline group, while PS presents serine[2]. In sphingolipids, such as SM, the glycerol component is replaced by a long-chain amino alcohol known as sphingosine, which is amide-linked to a fatty acid and phosphate group (**Figure 1.1b**)[2]. The membrane itself is composed of two layers of phospholipids, known as leaflets, which together form the phospholipid bilayer[3]. Here, the hydrophilic phosphate heads face towards the aqueous medium on either side of the membrane, while hydrophobic acyl lipid tails are sheltered behind them facing inwards (**Figure 1.1c**)[4]. Each acyl tail contains a hydrocarbon chain, which is composed of carbon and hydrogen atoms, held together

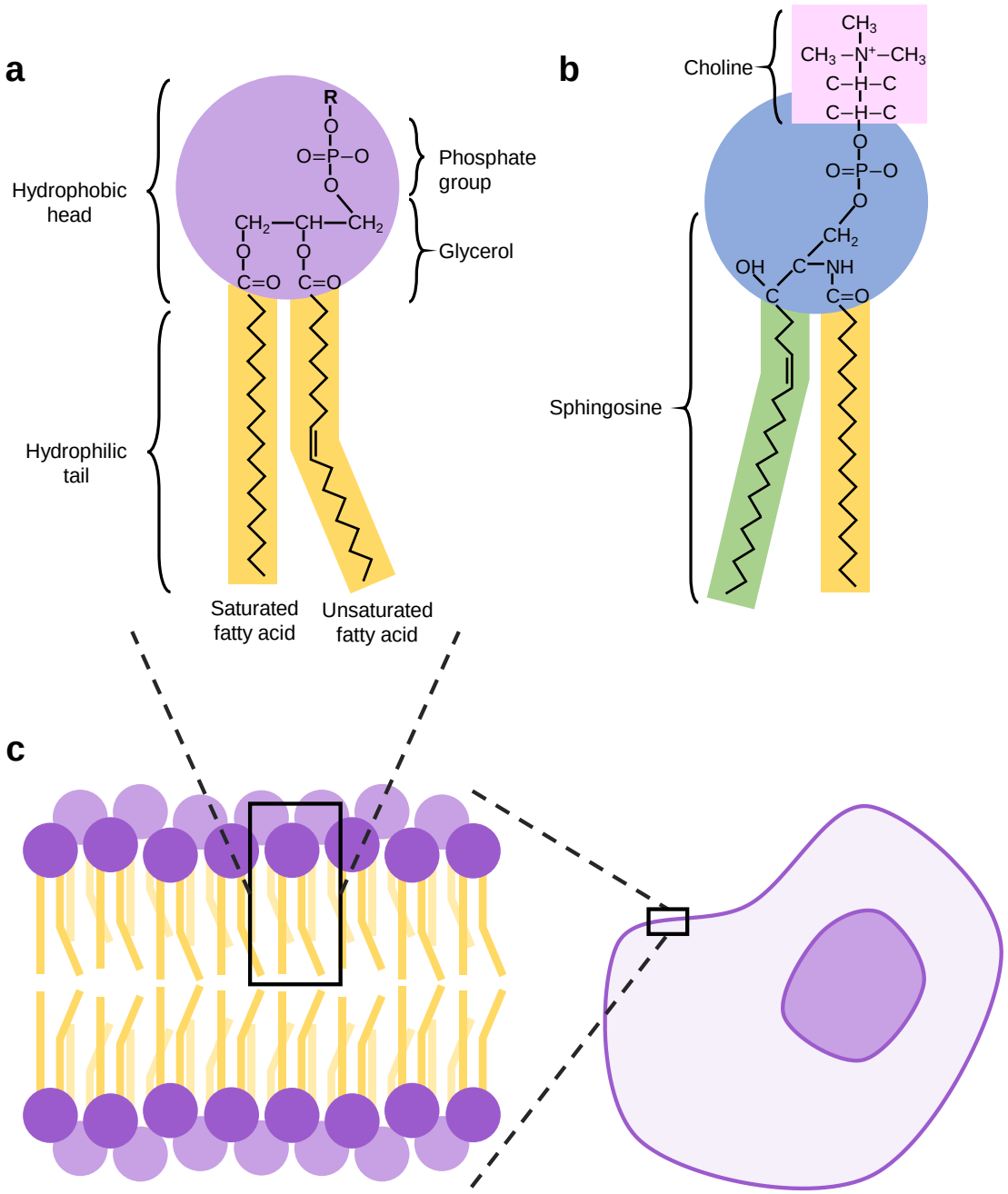


Figure 1.1: Lipids in the plasma membrane. **a** Diagram of a phospholipid. The hydrophobic head is comprised of a phosphate group attached to a glycerol moiety. Variable groups may be attached at the position labelled R. The hydrophilic fatty acid tails are hydrocarbons, which may be either saturated or unsaturated. **b** Diagram of sphingomyelin, which generally consists of a sphingosine base, with an 18-carbon chain and a double bond at position 4, attached to a phosphorylcholine fatty acid. **c** The position of the phospholipid within the phospholipid bilayer of a cell membrane.

by attractive forces called bonds[4]. In the context of acyl tail structure, adjacent carbon atoms may form single or double bonds – these tails are referred to as saturated and unsaturated fatty acids, respectively[4]. The fluid mosaic model, first coined by Singer and Nicolson in 1972, describes the bilayer as a two-dimensional fluid with associated biological molecules embedded (**Figure 1.2**)[5]. Eukaryotic cells contain glycolipids, another class of lipids, which are composed of two major moieties: a carbohydrate (a mono- or oligosaccharide group) and a lipid[6]. Glycolipids are found only on the outer (or exoplasmic) leaflet, the exterior surface of the membrane, acting as receptors to the extracellular medium (ECM) and inducing signal transduction[6]. The plasma membrane may also contain glycoproteins, which, like their glycolipid counterparts, also contain a carbohydrate group, but with the substitution of a peptide in place of a lipid[6].

Sterols (steroid alcohols) are another subtype of amphipathic lipids present in the mammalian plasma membrane. One such sterol is cholesterol, a 27-carbon compound composed of a hydrocarbon tail, four hydrocarbon rings (known as the sterol nucleus) and a polar hydroxyl group (**Figure 1.3**)[7]. Cholesterol is synthesised in the endoplasmic reticulum of the cell[8], but is noticeably sparse here, and most commonly localised with the plasma membrane (representing around 30% of all membrane-associated molecules)[7]. Owing to its hydrophilic head group and flat structure, cholesterol fits between the phospholipids in each monolayer, with its polar hydroxyl group close to the phospholipid head groups[9,10].

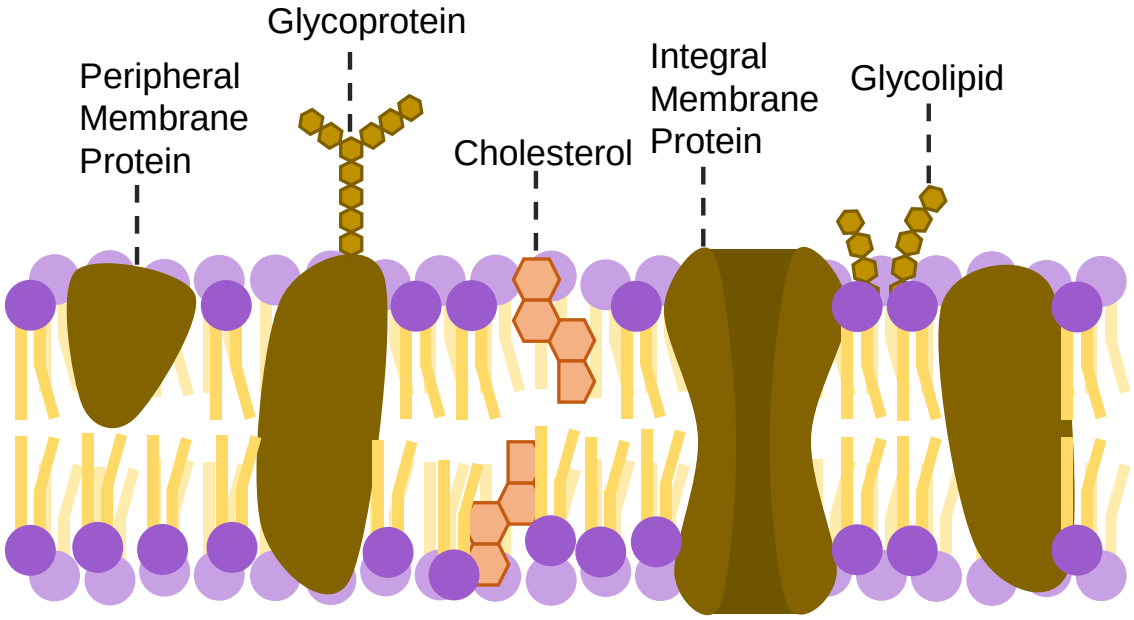


Figure 1.2: A diagram of a cell plasma membrane. The lipid bilayer provides a foundation in which peripheral proteins, integral (transmembrane) proteins, glycoproteins and glycolipids are embedded. Cholesterol is present within the bilayer of mammalian cells. This is representative of the Singer-Nicolson (fluid mosaic) model.

Proteins associated with the plasma membrane may be integral (embedded) or peripheral (attached to the surface via protein-protein interactions) and are largely responsible for regulating membrane processes and cellular function[11]. Some proteins, such as Lck or Ras proteins, may be affiliated with post-translational modifications[12]. This changes the structure and properties of the protein by proteolytic cleavage, adding a modifying group, such as phosphoryl, glycosyl or methyl, to one or more amino acids[12]. Integral proteins are typically transmembrane, with structures exposed on both the outer and inner leaflets. Such transmembrane proteins are held in the membrane by the hydrophobic amino acids present within their membrane-spanning component[11]. Membrane receptors are one class of integral protein, which are largely responsible for regulating cell signalling and communication[11]. Of particular note in this thesis is the T cell receptor complex, which mediates recognition of peptide

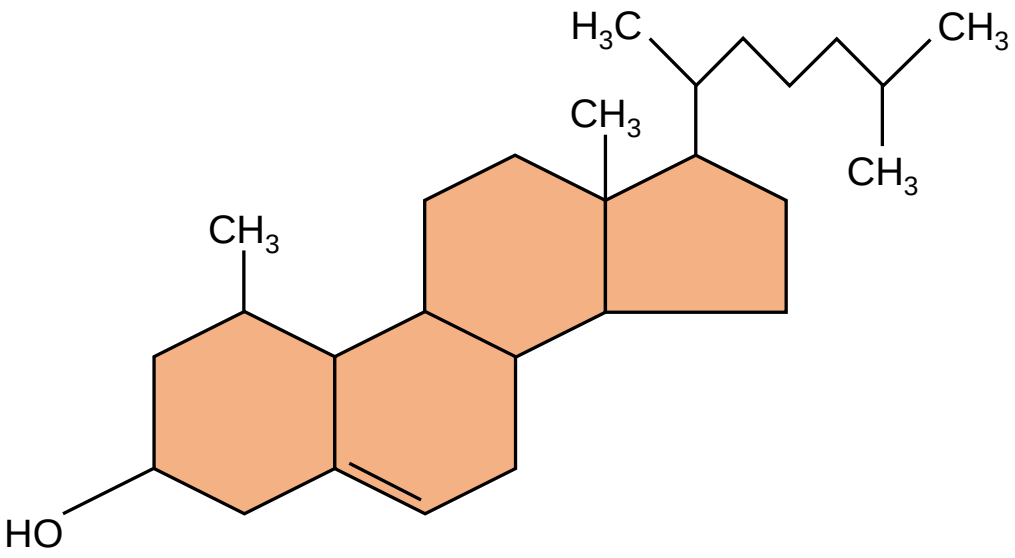


Figure 1.3: Diagram of cholesterol, a 27-carbon compound composed of a hydrocarbon tail, a sterol nucleus and a hydroxyl group.

fragments on antigen-presenting cells, and undergoes spatial reorganisation to elicit T cell activation[13]. Although the molecular composition differs on a cell-to-cell basis, all membranes possess lipids and proteins in roughly equal masses[14]. However, since protein molecules are larger than lipids, they only make up around 1-2% of all membrane molecules[14].

Membrane function

The structural role of the cell plasma membrane is to keep all cellular features, such as the cytoplasm and organelles, isolated from extracellular solvents and foreign biological structures[15]. It is also the primary site for complex biological processes, such as cell-cell recognition, cell communication, and active transport of molecules from the surrounding medium[15]. Cell signalling is the mechanism by which the plasma membrane receives or processes information from the cell's local environment[16]. Cell surface receptors, comprised of proteins or glycoproteins with highly specialised active

sites, bind to specific extracellular ligands[16]. This binding brings about activation of enzymes associated with the membrane, altering cell metabolism and ultimately giving rise to a cascade of chemical interactions[16]. Cell recognition, although functionally similar to cell signalling, is governed by binding of complementary molecules on opposing cell surfaces[17]. This mechanism allows the membrane to distinguish between neighbouring cells and is used to bring about behavioural changes of whole cells or tissues, such as during angiogenesis or adhesion[17].

The medium surrounding the cell contains nutrients required for cell survival, growth and proliferation, alongside other harmful or toxic substances, such as metabolic by-products[18]. The bilayer is populated with a range of specialised transmembrane proteins known as channels, which open in the presence of complementary substrates, and facilitate diffusion of small ions or molecules through the membrane[18]. Other proteins, termed pumps, force solutes across the membrane and into the cell via a process called active transport[19]. Both transporters may be gated by ligands, voltage or mechanical force[19]. Other membrane functions include endocytosis, in which a region of the extracellular space is internalised by the cell, and exocytosis, in which vesicles of molecules are brought to the inner leaflet and their contents are externalised[20].

Lipid packing and membrane order

Under the fluid mosaic model, the membrane exists in a state of thermodynamic equilibrium, afforded by the relative biophysical properties of the fluid membrane compartments[21]. One such property is membrane lipid order, which is characterised by lipid acyl chain structure, and can be measured directly through nuclear magnetic resonance (NMR) spectroscopy[22]. Steinkühler et al showed that membrane order is

proportional to membrane viscosity (resistance to change in shape or flow), and therefore inversely proportional to membrane fluidity (the reciprocal of viscosity)[23]. Heterogeneity in membrane order establishes a degree of varying fluidity across the membrane, which is essential for dynamics of lipids and proteins[21]. Several models have been proposed to describe these dynamics, including the picket fence theory[24,25], in which the motility of transmembrane proteins with large intracellular domains is obstructed by the dense cortical actin meshwork, the lipid raft hypothesis[26,27], in which protein affinity for ordered or disordered membrane regions dictates dynamics (see **Section 1.1.2**), and the size exclusion model[28], in which membrane protein distribution is influenced by extracellular domains interacting with the crowded extracellular space. Generally, transmembrane proteins move laterally across the plasma membrane, but motion may be restricted by the presence of intracellular proteins (such as membrane actin-myosin or the cytoskeleton)[29], contractile mechanisms of dynamic polar filaments[24,30], electrostatic interactions between proteins and lipids[31], obstruction by ligands[28], variation in membrane order[32], or post-translational modifications of proteins[33].

1.1.2 The lipid raft hypothesis

Membrane phases

The phospholipid bilayer can exist in a liquid-disordered, liquid-ordered or gel phase, although the latter is only observed in model membranes without cholesterol[34]. In the liquid-ordered phase, there is greater lipid packing, which reduces membrane permeability, obstructing the passage of ions and water-soluble molecules which otherwise would diffuse across the membrane and enter the cell[35]. But in the liquid-disordered phase, there is greater fluidity and lateral mobility in the plane of the bilayer.

These phases depend on phospholipid composition and are mainly dictated by the structure of the acyl lipid tails (**Figure 1.4**)[34]. From a biochemical perspective, the saturated tails generally adopt trans configurations and pack tightly together in an ordered array, promoting the liquid-ordered phase[36]. Unsaturated lipid tails are more likely to form cis configurations and become kinked[37]. In this case, the acyl tail loses its linearity and protrudes off-centre, making it less likely to fit neatly into the membrane beside other phospholipid molecules[37]. As such, the membrane becomes highly disordered and packing is less efficient, promoting the liquid-disordered phase[37].

Acyl tail length correlates with increased lipid order, which increases the melting temperature of the ordered phase over the disordered phase[38]. In thermodynamics, entropy is defined as the measure of a system's thermal energy per unit temperature that is unavailable for doing mechanical work[39]. In this context, entropy is interpreted as a measure of molecular disorder or randomness, and correlates with the number of geometric conformations (the three-dimensional arrangement of atoms) a molecule can attain[40]. Within unsaturated fatty acid molecules, an increase in temperature yields greater kinetic energy, which increases entropy and promotes lipid disorder[40]. Unlike unsaturated acyl tails, the four carbon rings which comprise the body of cholesterol give the molecule a rigid structure, which intercalates between phospholipid molecules[41]. This presents a dual functionality: at low temperatures, cholesterol prevents phospholipid molecules from clustering and stiffening, while at high temperatures, cholesterol reduces the entropy of acyl tails[41]. In either case, cholesterol plays a role in intra-cellular homeostasis by regulating membrane fluidity and counteracting the effect of changes in temperature on the nano-environment[42].

Lipids and membrane-bound molecules may present an affinity for a particular phase, and aggregate correspondingly[43]. Transmembrane proteins contain a hydrophobic

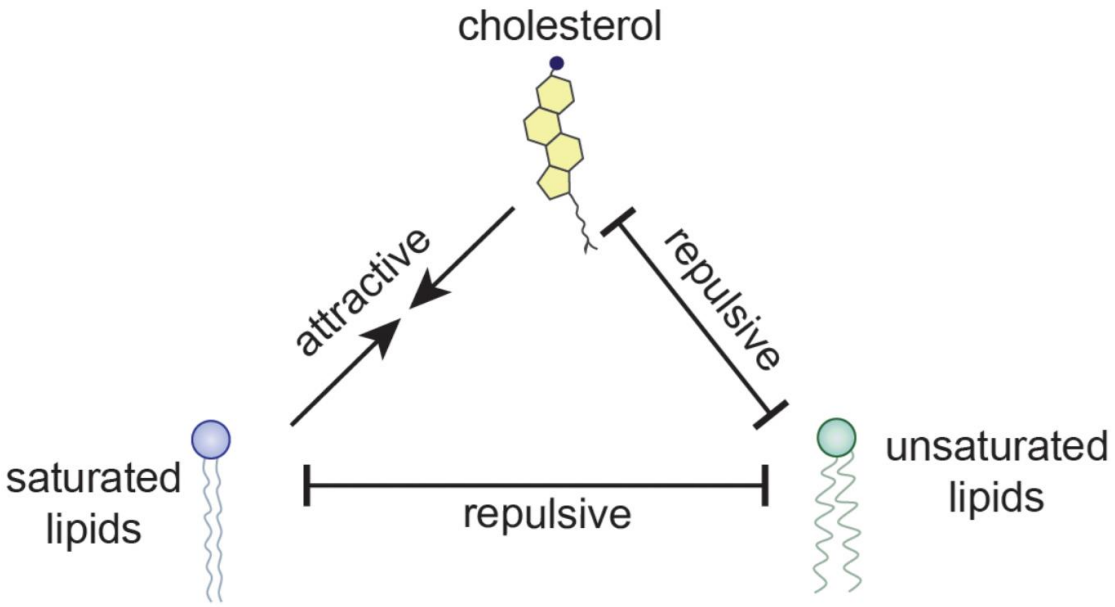


Figure 1.4: Saturated lipids and cholesterol interact more favourably with each other than with unsaturated lipids. Figure adapted from Levental et al[26].

region, known as a transmembrane domain, which embeds itself into the bilayer to evade water[43]. Lorent et al find that an increase in transmembrane domain length and total solvent-accessible surface area correlates with increased ordered phase affinity[44]. Recently, Gurdap et al showed that increasing the mass or glycosylation – a post-translational modification – of the extracellular protein domain (ECD) can decrease molecular diffusivity and affinity towards the ordered phase[45]. Initiation of cellular signalling has been attributed to the ECD, and many proteins undergo post-translational modifications or ligand binding during signalling events[45].

It is hypothesised that the membrane organises itself into liquid-ordered domains of sub-200nm size, enriched with clusters of cholesterol and saturated lipids (phospholipids and sphingolipids), that move laterally within the plasma membrane as a unit[35]. These domains, called lipid rafts (**Figure 1.5**), may underpin cell signalling and communication processes[46]. The pertinence of bio-membrane lipid order came to

light when Baumgart et al demonstrated spontaneous lipid phase separation in giant plasma membrane vesicles (GPMVs), in which co-existing liquid phases promoted plasma membrane organisation into “raft-like” lipid domains[47]. Sezgin et al reported greater lipid packing in this raft-like phase, which associated with higher viscosity and enrichment with saturated lipids, sterol analogues and lipidated proteins[48]. This, fundamentally, validated the principle that ordered, lipid-driven phases can elicit functional organisation of the plasma membrane[27]. From an evolutionary perspective, this supports the notion that the plasma membrane, with minimal energy expenditure, could concentrate specific reactants, exclude negative regulators, induce conformational changes, and regulate local membrane properties[26]. However, Levental et al determine an absence of most transmembrane proteins from lipid rafts in GPMVs, and find that even those proteins which are included are rarely enriched[49]. Further, Saka et al observe protein-depleted membrane domains in live cells under super-resolution microscopy[50]. The biophysical properties (such as size, lifetime and stability) and compositions of lipid rafts are context-dependent and reflect the specifics of the membrane in which they arise – the full impact of this heterogeneity is yet to be fully accounted for[51]. Adding to this complexity, Tulodziecka et al propose that immobilised order-preferring proteins could recruit a dynamic assembly of raft-forming lipids, rather than lipid-driven domains recruiting freely-diffusing proteins[52]. The interplay of lipids and proteins through cooperative clustering is fundamental to raft function, though the cause cannot be so easily distinguished from the effect.

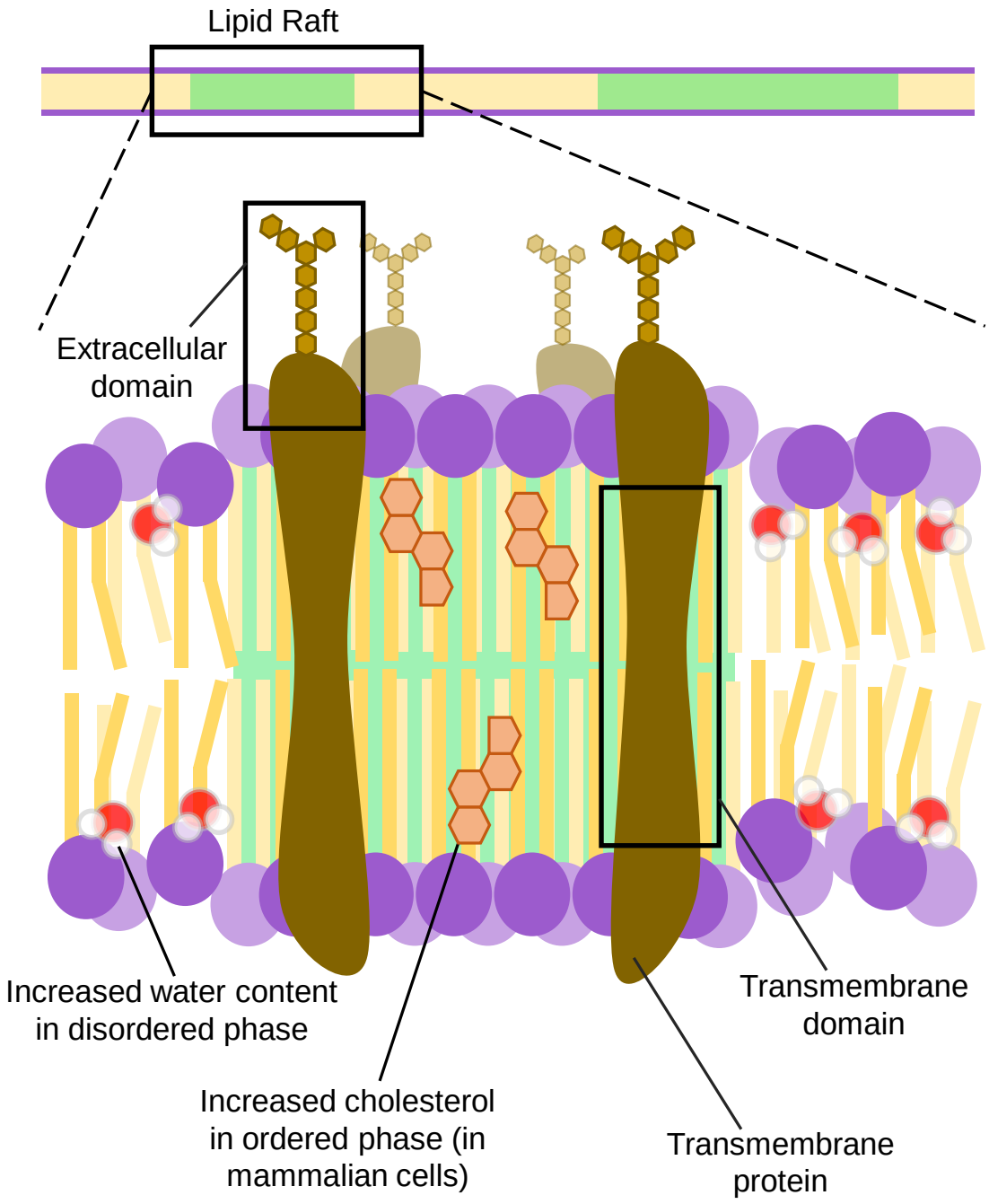


Figure 1.5: A lipid raft within a mammalian cell plasma membrane. Increased cholesterol promotes rigidity in the ordered phase and lipid packing excludes water molecules. As a result, water content is increased in the disordered phase. The transmembrane domain dictates protein affinity for lipid order, which may cause transmembrane proteins to aggregate into lipid rafts.

Phase separation and biomolecular condensates

Phase separation is the physicochemical process by which a homogenous mixture separates into distinct phases[53]. Phase separation in membrane lipids can result in the formation of lipid rafts[53]. Generally, phase separation among lipids is influenced by factors such as lipid composition, temperature and membrane curvature[53]. Biomolecular condensates are a class of membrane-less organelles and organelle subdomains – including, for instance, clusters of signalling molecules at the membrane – which concentrate specific collections of proteins and nucleic acids[54]. These dynamic assemblies arise from the condensation of cellular material through liquid-liquid phase separation and are hypothesised to play a role in protein assembly and signalling[55]. Furthermore, the crosstalk between membranes and condensates can promote phase separation in lipids and proteins[56]. Condensates often present as relatively small assemblies (ranging from nanometres to micrometres in scale) that do not coarsen into a single droplet despite their known proclivity to fuse[55]. Furthermore, membranes have been reported to control the size of intracellular condensates and modify their material properties[55]. It has been shown that the interaction between condensates and membranes can lead to the complex remodelling of the membrane-condensate interface, producing microscopic membranous protrusions[57]. Furthermore, Mangiarotti et al suggest that condensate wetting is a mechanism to tune lipid packing[56]. These observations offer some insight into the interaction between membranes and condensates at the micro-scale, but the underlying molecular mechanisms remain elusive and largely unexplored[55]. The relationship between the microscopic properties of the component molecules and the macroscopic properties of condensates is not fully understood[56]. Furthermore, it is not known how these

properties relate to biochemical and cellular functions, or if cells regulate these properties to achieve a functional effect[56].

Functional roles of lipid rafts

Lipid rafts modulate the aggregation and mobility of transmembrane proteins, through lipid-mediated interactions in membrane-hosted signalling events, and regulate the frequency of protein encounters to determine signalling outcomes[58]. These rafts play a role in several cell survival signalling pathways, motility of transporters and ion channels, and death receptor signalling[18,51]. Alteration of the composition or distribution of ordered domains can disrupt cell function, up to and including cell death[59,60]. For instance, Ros-Baró et al showed that lipid rafts underpin redistribution and translocation of the glucose channel GLUT4 in adipose cells[61]. This promotes insulin-regulated glucose uptake into muscle cells, which is foundational for physical movement[61]. Chamberlain et al found that SNARE proteins, which mediate fusion of vesicles during exocytosis, associate with glycosphingolipids and cholesterol in lipid rafts[62]. This association acts to concentrate SNARE proteins at defined exocytosis sites on the plasma membrane, and depends heavily on membrane lipid composition and organisation[62]. Analogously, Lajoie et al found that lipid rafts regulate caveolin-1 and dynamin-1 distribution, which is an essential prerequisite of vesicle formation in receptor-mediated endocytosis[63].

Lymphocytes, such as T cells, must strike a delicate balance between activating in response to signals from potentially pathogenic organisms and avoiding activation from stimuli emanating from the body's own cells[64]. It is well-documented that the raft-mediated oligomerisation, or clustering, of T cell receptors facilitates T cell activation and immunological specificity[65-67]. Upon activation, the T lymphocyte cell surface is

restructured, forming membrane domains at TCR-signalling foci and immunological synapses[68]. The plasma membrane condenses into the liquid-ordered phase at TCR activation sites, and TCR engagement promotes aggregation of lipid rafts[68]. The protein tyrosine phosphatase CD45 has been reported to both activate the Src family kinase Lck, which phosphorylates the TCR complex, and conversely to dephosphorylate TCR signalling motifs[69]. These mechanisms are hypothesised to grant CD45 both an immunoenhancing and immunosuppressive function, which enables graded signalling outputs while filtering weak or spurious signalling events[69]. CD45 and TCR bear large ECDs, and express affinity for disordered phase, whereas Lck is a small protein, which exhibits a strong affinity for the ordered phase[69]. Urbančič et al report that partial immobilisation of TCR, through aggregation or ligand binding, changes the complex's preference towards a more ordered lipid environment[70]. This leads to preferential partitioning of TCR with Lck, while segregating CD45, which promotes immunological activation[70]. Aggregation of lipid rafts results in colocalisation of Lck and raft-associated protein LAT (linker for activation of T cells) and exclusion of CD45[71]. Wang et al report that coupling of LAT protein condensates to ordered lipid domains is a prerequisite to functional membrane organisation in T cells[72]. Schiefer et al found that cholesterol depletion, and therefore lipid raft disruption, inhibited CD4+ T cell responses[73].

Analogously, receptor organisation plays a role in suppression of T cell signalling and effector functions. Programmed death ligand-1 (PDL1), expressed on antigen presenting cells, induces spatial reorganisation of programmed death-1 (PD1), a coinhibitory receptor expressed on activated T cells, into nanoscale clusters, which elicit an immunosuppressive response[74]. Fang et al showed that PDL1-presenting DNA origami flat sheets inhibited T cell signalling with two ligands separated by

200nm, but not at 13nm or 40nm[74]. The spatial organisation PDL1 regulates T cell signalling and is currently being investigated for intelligent nanotherapeutic design in immunotherapy[74]. Fernandes et al found that ligand discrimination depends crucially on individual contacts being ~200 nm in radius, matching the dimensions of the surface protrusions used by T cells to interrogate their targets, which suggests that cell topography and curvature influences signalling outcomes[75]. These results are not exclusive to just T cells and suggest that protein oligomerisation, lipid ordering and signal initiation are indeed linked via passive biophysical concepts[76]. Lipid nanodomains are generally heterogeneous and their characteristics depend on specific cell, lipid and environmental conditions, although it is generally regarded that protein clusters are tuned to optimise the transmission of signalling information[77,78].

1.1.3 Interplay between lipid rafts and membrane biophysical properties

Thermodynamics of membrane phases

Acyl tail length correlates with increased lipid order, which increases the melting temperature of the ordered phase over the disordered phase[38]. In thermodynamics, entropy is defined as the measure of a system's thermal energy per unit temperature that is unavailable for doing mechanical work[39]. In this context, entropy is interpreted as a measure of molecular disorder or randomness, and correlates with the number of geometric conformations (the three-dimensional arrangement of atoms) a molecule can attain[40]. Within unsaturated fatty acid molecules, an increase in temperature yields greater kinetic energy, which increases entropy and promotes lipid disorder[40]. Unlike unsaturated acyl tails, the four carbon rings which comprise the body of cholesterol give the molecule a rigid structure, which intercalates between phospholipid molecules[41]. This presents a dual functionality: at low temperatures, cholesterol

prevents phospholipid molecules from clustering and stiffening, while at high temperatures, cholesterol reduces the entropy of acyl tails[41]. In either case, cholesterol plays a role in intra-cellular homeostasis by regulating membrane fluidity and counteracting the effect of changes in temperature on the nano-environment[42].

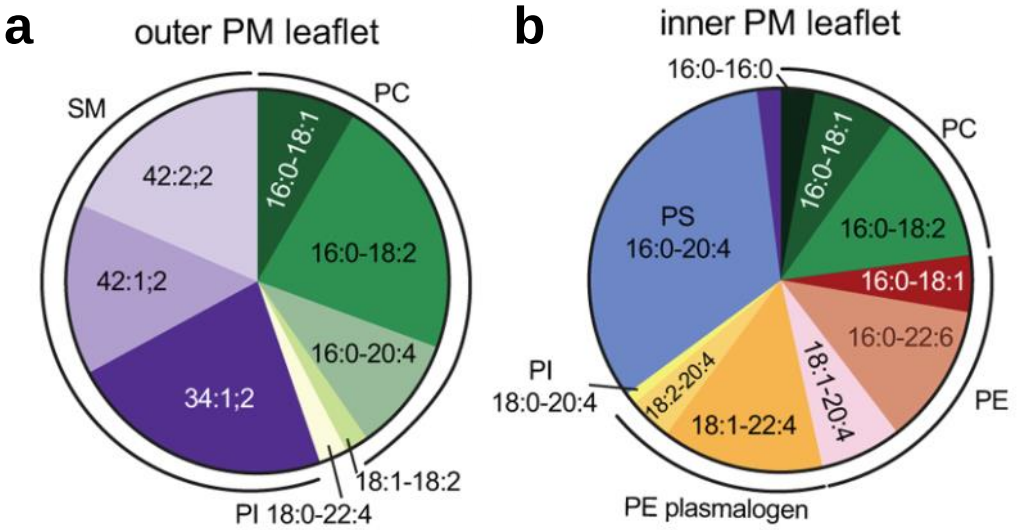
Lipid order and diffusivity

The diffusion coefficient will typically be lower in the ordered phase, due to increased viscosity, which correlates with tighter lipid packing[79]. Filipov et al found that the lateral diffusion coefficient decreases linearly with molecular cholesterol percentage and increases exponentially with temperature[80], with coefficient $D \approx 1.25 \times 10^{-13}$ m^2/s for 30% cholesterol content at 313°K. Dietrich et al reports that the ratio of coefficients between the liquid-disordered (D_d) and liquid-ordered (D_o) phase is measured as $D_d/D_o \approx 3$ in model membranes[81]. Further, Beckers et al found a near-linear correlation between membrane order and diffusion coefficient in GUVs (giant unilamellar vesicles) and GPMVs, which reconstitute lipid bilayers, but not in supported lipid bilayers, which present nanoscale differences in surface topology[82]. This suggests a relationship between lipid order, melting temperature, viscosity, diffusivity and membrane curvature, although a functional form has yet to be fully quantified[82]. Furthermore, these properties can differ significantly across and between leaflets.

Membrane asymmetry

The lipid subtypes PC and SM are found predominantly in the exterior leaflet while PE, PI and PS are typically associated with the inner leaflet[3]. This difference in leaflet composition, alongside a variation in curvature, characterises the property of

membrane asymmetry[21]. Such asymmetry is advantageous because it ensures that membrane proteins are correctly oriented for their specific functions[21]. Lorent et al performed comprehensive lipidomic analysis and determine lipid species concentrations across the inner and outer leaflets[3]. The authors found discrepancies between lipid concentrations pertaining to both the head groups and acyl chain formations[3], as summarised in **Figure 1.6**. Notably, they report that lipid tails in the cytoplasmic leaflet are approximately 2-fold more unsaturated than the exoplasmic leaflet[3]. Variation in cholesterol content can differ across the plasma membranes of mammalian cells, and may depend on cell line or primary cell origin[83]. For example, Buwaneka et al determine molecular cholesterol composition of 44.0% in the outer leaflet of HeLa cells and 3.6% in the inner leaflet[83]. In the same study, the authors find 49.0% content in the outer leaflet of Human Pulmonary Artery Endothelial Cells (HPAEC) and 0.6% in the inner leaflet[83]. It is, however, generally regarded that the exoplasmic monolayer comprises a higher molecular cholesterol content and exhibits greater lipid packing than the cytoplasmic monolayer[84]. Pinkwart et al found that the diffusion coefficient of cholesterol was approximately twice as large as the diffusion coefficients of phospholipids and sphingolipids in live cells, but only in the inner leaflet of the bilayer[85]. Buyan et al report increased cholesterol concentration in the outer leaflet, especially in regions of relatively high curvature, whereas PC and SM concentrations are reduced[86]. Owing to these differences, it is not necessarily the case that lipid domains or rafts must match up in position across the leaflets[3].



c

Outer (exoplasmic) leaflet		Inner (cytoplasmic) leaflet	
representative lipid	*mol% of outer leaflet*	*representative lipid*	*mol% of inner leaflet*
PC 16:0–18:2	22.2	PS 16:0–20:4	33.1
SM 34:1;2	22.4	PE O- 18:1–22:4	13.7
SM 42:2;2	18.5	PC 16:0–18:2	13.0
SM 42:1;2	14.4	PE 16:0–22:6	11.9
PC 16:0–20:4	9.7	PE 18:1–20:4	6.9
PC 18:0–18:1	8.4	PC 16:0–18:1	7.0
PC 18:1–18:2	2.0	PE 16:0–18:1	4.8
PS 18:0–20:4	1.5	PE O- 18:2–20:4	3.7
PI 18:0–22:4	0.9	PC 16:0–16:0	2.9
		SM 34:1;2	2.0
		PI 18:0–20:4	1.0

Figure 1.6: The asymmetric lipidome of the plasma membrane. Molecular content of lipid species present in **a** the exoplasmic and **b** the cytoplasmic plasma membrane leaflets. Numeric percentiles are given for each molecular species in **c**. Figure adapted from Lorent et al[3].

Girard and Bereau showed that membrane asymmetry may arise as a natural consequence of differences in phospholipid and sphingomyelin count across the

leaflets[87]. Asymmetry in composition is stabilised by the presence of a high free energy barrier against transverse diffusion of constituents (flip-flops) and the controlled abundance of flippase and floppase proteins, which translocate lipids from the outer leaflet to the inner leaflet, and vice versa, respectively[88]. Menon et al find that phospholipid scramblases, a class of proteins responsible for translocation of phospholipids between the monolayers of the membrane, are impaired upon cholesterol loading, and therefore cholesterol acts to stabilise asymmetry[89]. Levental et al find that mammalian membranes incorporate dietary polyunsaturated fatty acids, which induce a reduction in membrane order[4]. These effects are rapidly compensated through upregulation of saturated lipids and cholesterol, which promotes robust lipidomic remodelling and recovers homeostatic membrane packing and permeability[4]. Notably, inhibition of this response leads to cytotoxicity[4], which was also reported by Castillo et al, who showed that mitochondria-targeting cancer drugs can disrupt membrane asymmetry, alter membrane permeabilisation and elicit cell death[90].

1.1.4 Techniques to study membrane biophysical properties and lipid order

Non-optical methods for studying membrane order

The extent of lipidomic organisation, and even the existence of lipid rafts, is often drawn into question. Part of this scepticism arises from an absence of direct observation of lipid rafts in live cells, even with modern imaging modalities. The necessity for studying heterogeneity in membrane composition and biophysical properties has guided technological innovation. For example, electron microscopy (EM) is one such method for probing membrane order which makes use of electron optics[91]. When a focused beam of electrons is directed onto a coated sample, the metal ejects X-rays and backscattered electrons, which can be traced by a detector to

produce signal[91]. The wavelength of an electron used in EM depends on the voltage with which the microscope is operated, although this wavelength (~2-4pm) is typically far below the wavelength of visible light (~400-700nm)[91]. Since image resolution is proportional to imaging wavelength, this permits sub-1nm resolution[91]. However, EM only functions in samples coated with conductive metals, making the method unsuitable for live cell imaging[91]. Atomic force microscopy (AFM) can also probe membrane thickness by mechanically scanning the cell surface with a sharp probe and tracing motion[92]. Studies in artificial membranes have identified decreases in membrane thickness of up to 1nm in the liquid-disordered phase, when compared to the gel phase, using AFM[92]. However, AFM presents a slow rate of scanning which can lead to pronounced sample drift[92]. This is highly detrimental to live-cell image quality, especially when viewing lipid domains at the nanoscale.

Optical methods for studying membrane order

Optical imaging methods typically make use of fluorescence, which is the photophysical process of absorbing light and then producing light at a longer wavelength[93]. Quantitative fluorescence microscopy can be used for imaging lipid environments and measuring membrane lipid order in live and fixed cells, as well as in intact tissues (see **Section 1.2**). Substances which exhibit fluorescence are known as fluorophores[93]. The distribution of photonic emissions from a fluorophore is defined as the emission spectrum, which determines the relative intensity of each wavelength of light emitted[93]. In its ground state, a fluorophore rests in a relatively low-energy, stable configuration, and does not fluoresce[94]. When light from an external source hits the molecule, energy may be absorbed, and if the energy absorbed is sufficient, the molecule reaches a higher-energy state called an excited state[94]. The fluorophore

returns to the ground state and the excess energy is released, emitted as light – the length of time that the fluorophore remains in the excited state is called the fluorescence lifetime, which is typically around 10^{-8} seconds[93]. The excitation spectrum represents the relative emission of the fluorophore at each excitation wavelength[93]. Since the emission is of lower energy than the absorption, the light emitted is of longer wavelength[94]. Artificial fluorophores are usually designed for staining of specific cellular components or biological constructs[95]. Those used in imaging the plasma membrane must be amphipathic but not permeated by the membrane itself[95]. An example fluorophore, Alexa Fluor 488, is given in **Figure 1.7**. Each fluorophore presents inherent photophysical properties, which must be appropriately selected for, depending on target structure and imaging modality. This may include:

- Stokes shift – The difference between absorption and emission maximum, measured in either wavelength or wavenumber[96].
- Absorption cross-section (σ_{net}) – The probability that a photon will be absorbed by the fluorophore multiplied by the average cross-sectional area of the molecule[93].
- Quantum efficiency – The ratio of the number of photons absorbed to number of photons emitted, often measured as a percentage[97]. Probes are generally selected for high absorption cross section and quantum efficiency, as this increases contrast against background signals and noise[97].
- Switching cycle – The average number of times the fluorophore can switch between dark and fluorescent states[95]. This should be optimised to balance the trade-off between reducing the impact of imaging artefacts and ensuring sufficient fluorophore blinking[95].

- Photostability – In an excited state, fluorophores are more likely to react with oxygen in their environment, which may cause them to denature and lose their ability to fluoresce (photobleaching)[98]. The number of photons a fluorophore can emit before becoming photobleached varies between molecular species, but is typically between 10^4 and 10^5 for fluorescent proteins and between 10^5 and 10^6 for organic dyes[98].

Furthermore, the buffers, imaging conditions and laser intensity required by the fluorophore must not impact cellular physiology throughout the acquisition[95]. Anti-photobleaching solutions can obstruct photobleaching by reducing the oxygen available to fluorophores, but may be toxic to living cells[95].

Optical probes used in model membranes can discern the liquid-ordered and liquid-disordered phases by preferential partitioning[99]. These fluorescent molecules, known as partitioning probes, are lipophilic and therefore colocalise with specific lipid domains[99]. Such probes can be viewed under conventional fluorescence microscopy (see **Section 1.2.1**) in artificial membranes where domains form at the macroscale[99]. However, conventional portioning probes would need to be mixed into the lipids prior to membrane formation, making them unsuitable for imaging rafts in cells[99]. Traditionally, probes used in cell studies would bind to membrane components that aggregate in rafts[100]. This could be used to determine clustering of intrinsically raft-associated molecules, but would not demonstrate the physicochemical properties of the environment of the raft, unlike modern fluorescent probes[100]. In order to map the impact of lipid order on cellular function, dynamic visualisation and quantification of lipid packing in live cells is required. The aforementioned methods are either inapplicable to live-cell imaging or incapable of explicitly tracking lipid properties[99,100]. Further, the liquid-ordered phase is itself defined by structural arrangements of tight,

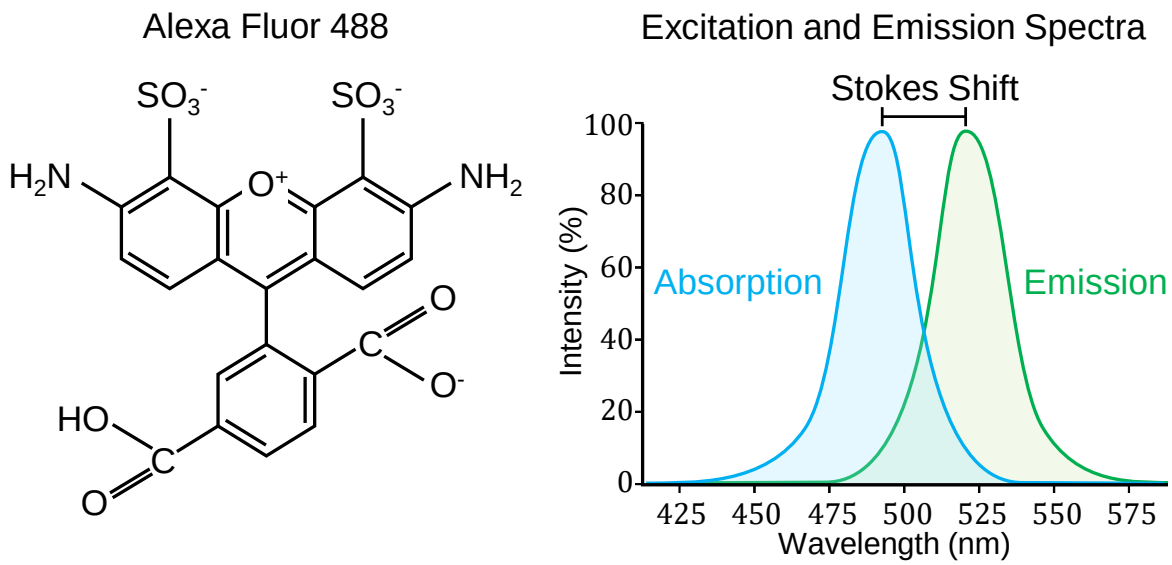


Figure 1.7: The chemical composition of the fluorescent probe Alexa Fluor 488 alongside the molecule's excitation and emission spectra. Stokes shift is defined as the difference between excitation and emission peaks.

highly specific lipid packing, and tagging of raft lipids with a bulky, hydrophilic label – such as those used in domain-specific probes – would reduce affinity for raft domains[101]. Accordingly, Sezgin et al finds that most fluorescent lipids are excluded from ordered phases[101]. However, smaller, quantitative, environmentally-sensitive probes offer a more practical method of probing live-cell membrane properties.

Environmentally-sensitive fluorescent probes

Fluorophores which alter their emission spectra, fluorescence intensity or fluorescence lifetime in response to the specific properties of their surrounding medium are known as environmentally-sensitive fluorescent probes[66]. Multi-channel fluorescent microscopy techniques image over several channels in sequence, detecting distinct ranges of wavelengths. The spatially-averaged intensity in each

channel can then be used as a proxy for the target membrane property. Such fluorescent probes include:

- Viscosity sensors, which measure the membrane's resistance to changes in shape. Some BODIPY dyes depict changes in viscosity by altering their fluorescence intensity and lifetime[102].
- pH sensors alter their fluorescence properties in response to acidity across the membrane, which is defined by the abundance of hydrogen ions (protons)[103]. pHrodo dyes are some of the most commonly used indicators of pH, for which green, red and deep red variants exist[103].
- Surface tension sensors are mechanosensitive. One such tension biosensor is Flipper-TR, which prefers a twisted molecular conformation in relaxed membranes[104]. Local lipid reorganisation stretches the probe into a planarised conformation as membrane tension increases – for this reason, Flipper-TR is called a push-pull probe[104]. This change of structure alters the distribution of electrons within the molecule, inducing a red-shift in excitation spectra and an increase in fluorescence lifetime[104].

1.1.5 Polarity-sensitive fluorescent probes

Efficient lipid packing in the ordered phase excludes polar water molecules from the otherwise non-polar bilayer, resulting in a change in local environmental polarity[105]. As a result, polarity-sensitive fluorescent probes alter their emission spectra depending on the polarity of the surrounding medium, and can therefore be used as a proxy for the degree of membrane lipid order[106]. These probes are solvatochromic and show an increase in charge separation when excited in polar solvents, which results in a larger dipole moment[107]. The transitions from locally excited state in non-polar solvents to

internal charge transfer state in polar solvents shifts the emission maxima[108]. This shift in emission profile between liquid-disordered and liquid-ordered phases allows for quantitative assessment of membrane order[42]. This is typically achieved by calculating a ratiometric measurement of the fluorescence intensity recorded in two spectral channels, known as a generalised polarisation (GP) value[42]. Leung et al showed that the order parameter derived from NMR correlates linearly with GP from fluorescence microscopy with polarity-sensitive dyes[22].

Nile Red

Nile Red is a solvatochromic dye, used for imaging lipid droplets and membranous structures in cells, which carries a quantum yield of 70% [109]. The structure of Nile Red includes a hydrophobic phenoxazine ring, which allows the molecule to integrate into the plasma membrane[109]. In the disordered phase, Nile Red fluoresces with an emission maximum of 630nm, which drops to approximately 550nm in the polar environment of the ordered phase[110]. This exhibits a blue shift of 80nm and represents a chromatic shift from yellow to deep red[111], however, Moon et al showed that the average Stokes shift of Nile Red molecules increased linearly with cholesterol content in model lipid vesicles[112]. Using spectrum and polarisation optical tomography, Zhanghao et al recorded the full spectra of Nile Red emissions, tracked dynamic lipid properties in subcellular membranes, and determined an increase in polarity from early to late endosomes[110]. Iaea et al used Nile Red-based probe NR12S to compare differences in lipid order between the cell plasma membrane and the endocytic recycling compartment of model cell line U2-OS[113].

Laurdan

The polarity-sensitive probe Laurdan comprises an aliphatic tail of 12 carbons[114]. As such, Laurdan embeds within the plasma membrane, exposing its fluorescent moiety to the surrounding aqueous medium[115]. As a derivative of the fluorescent hydrocarbon diphenylhexatriene, this dye has been shown to report local phase separation when inserted in GUVs and live cell membranes, resulting in a 50nm blue shift in the ordered phase, from 490nm to 440nm (blue to violet), with a quantum yield of 61%[116,117]. Laurdan has been used to demonstrate that HeLa and B35 neuroblastoma cell plasma membranes could be forced to undergo broad order-disorder phase transitions when subjected to variation in cholesterol and pH or treatment with anaesthetic[118]. Gaus et al pioneered the use of Laurdan to measure condensation of the plasma membrane at the site of T lymphocyte activation[68]. It was determined that membrane condensation occurs upon TCR stimulation but is prolonged by CD28 co-stimulation, which suggests a role of lipid rafts in mounting of the immune response[68].

Di-4-ANEPPDHQ

Di-4-ANEPPDHQ is a polarity-sensitive styryl dye, which has been used to visualise cholesterol-rich lipid domains (**Figure 1.8**)[119]. The molecular structure of di-4-ANEPPDHQ includes a hydrophobic polyene chain which nestles into the plasma membrane, parallel to lipid acyl chains, positioning its fluorescent core to interact with the local lipid environment[119]. Variations in polarity induce a 60nm blue shift in emission spectra (from 630 to 570nm) between membranes in the cholesterol-containing liquid-ordered state to the cholesterol-free liquid-disordered state[100,119]. Building on previous membrane order comparisons with Laurdan, Sengupta et al used di-4-ANEPPDHQ to show that cholesterol-dependent membrane order is critical for

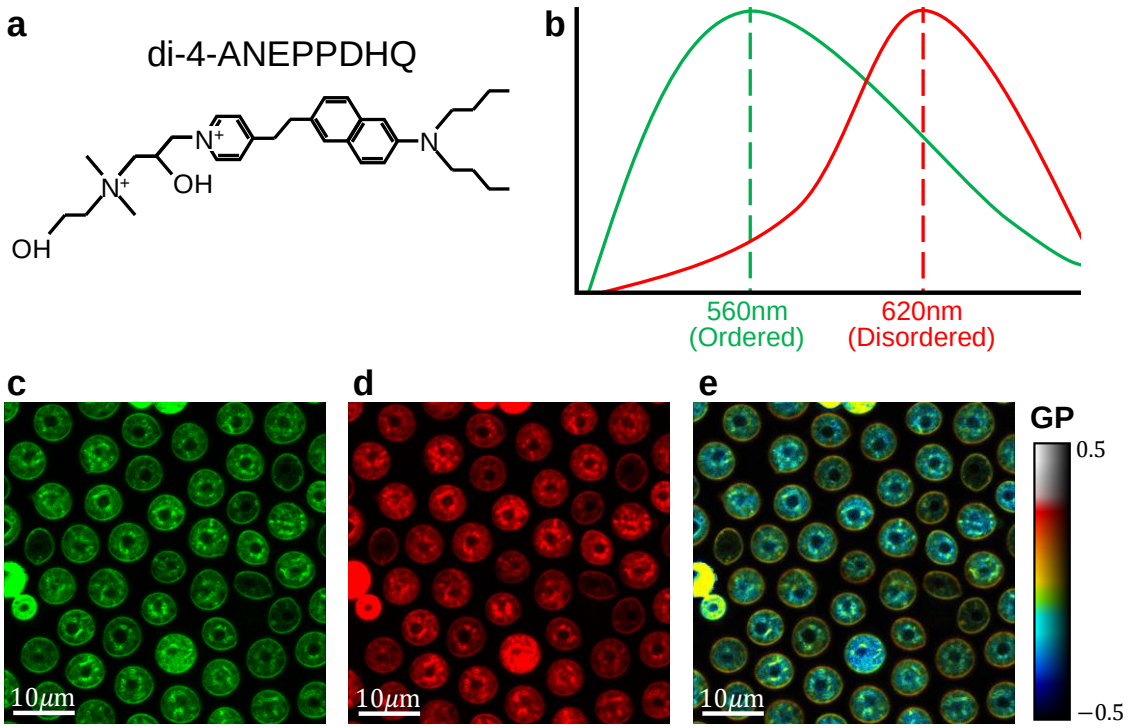


Figure 1.8: An example polarity-sensitive probe, di-4-ANEPPDHQ. **a** The chemical structure of di-4-ANEPPDHQ. **b** Schematic of emission spectra within the ordered (green) and disordered (red) phase. Variations in intensity are tracked across two image channels, corresponding to the **c** ordered and **d** disordered phase. **e** The GP image is taken as a ratiometric sum of **c** and **d**. *C gattii* cells stained with di-4-ANEPPDHQ were imaged under confocal microscopy as in Panconi et al[122].

responses generated by CD4+ T cells[120]. Waddington et al measured lipid order at the immune synapse with di-4-ANEPPDHQ and found that activation of liver X receptor (LXR), a key transcriptional regulator of cholesterol and phospholipid metabolism, significantly reduced lipid order and dampened proinflammatory T cell function[121].

Considerations for probe selection

As discussed, the collective biophysical properties which induce membrane order are determined by the structural diversity of lipid species within the membrane[123]. With GP

as a metric of lipid-packing, the next task is in quantifying the difference between domains and their surrounding bulk, and determining whether this difference is even large enough to be detectable. Feigenson et al report that model membranes express relatively small differences in cholesterol concentration between the ordered and disordered phases[124]. Furthermore, Sezgin et al has shown that differences in lipid packing are much smaller in natural membrane models than in the synthetic models which are typically used to calibrate analytical methods[101]. Even with state-of-the-art imaging modalities, the limitation of discernible lipid packing would require sensitive, or at least highly tuneable statistical analysis. Furthermore, it has been shown that polarity-sensitive dyes exhibit varying sensitivity to the biological parameters which underpin membrane order[125]. For instance, Ragaller et al determine that Pro12A (a Laurdan derivative) performs well at sensing cholesterol content, while NR12S (a Nile Red derivative) is preferable for differentiating the degree of acyl tail saturation and the properties of the phospholipid headgroup[125]. Alternatively, the authors find that NR12A (another Nile Red derivative) is best suited for differentiating between positions and configurations of the double bond in unsaturated lipids[125]. Unlike other polarity-sensitive dyes, di-4-ANEPPDHQ does not require multiphoton excitation and exhibits complex photophysics (sensitivity to multiple biophysical properties)[108,123]. Further, it has been shown that di-4-ANEPPDHQ is unaffected by membrane proteins, including inserted peptides, making it an ideal probe for assessing lipid packing in both live cells and artificial membranes[126].

1.2 Microscopy

Cells are generally only visible at sub-100μm scales, and therefore microscopy systems must be used to examine the structure, composition and function of the

plasma membrane[127]. Conventional fluorescence microscopy (FM) permits research into the biophysical properties of the plasma membrane at a diffraction-limited resolution of ~200nm[127]. With the introduction of super-resolution microscopy, it is possible to visualise and study structures at resolutions below the diffraction limit of light[128-130]. However, the output data of FM can range from simple grayscale images to highly multivariate datasets[131-135]. This data does not necessarily represent strictly spatial information about cellular structures and can encompass an array of membrane properties[136].

1.2.1 Introduction to fluorescence microscopy and justification for use

The basic fluorescence microscope

The term “fluorescence microscope” refers to any microscope which uses fluorescence to generate an image[137]. In practice, a fluorescence microscope is an optical instrument, which can perform selective excitation and detection of fluorescent molecules within organic and inorganic samples[138]. This process can be performed over a range of spatial and time scales, depending on the instrument setup[138]. A fluorescence microscope requires a near-monochromatic illumination source, which presents a higher light intensity than some widespread light sources such as halogen lamps[139]. Such light sources could include lasers, high-power LEDs, or xenon arc and mercury-vapour lamps with an excitation filter[139]. This near-monochromatic light may be filtered through dichroic mirrors, which are specifically designed to transmit particular wavelengths of light and reflect others[93]. A tube lens may be used to refocus non-converging rays from the objective lens to form an image, or collimate light, which is then collected by the objective lens and focused on a sample stained with fluorophores[135]. Light emitted from fluorescent molecules is then filtered back through

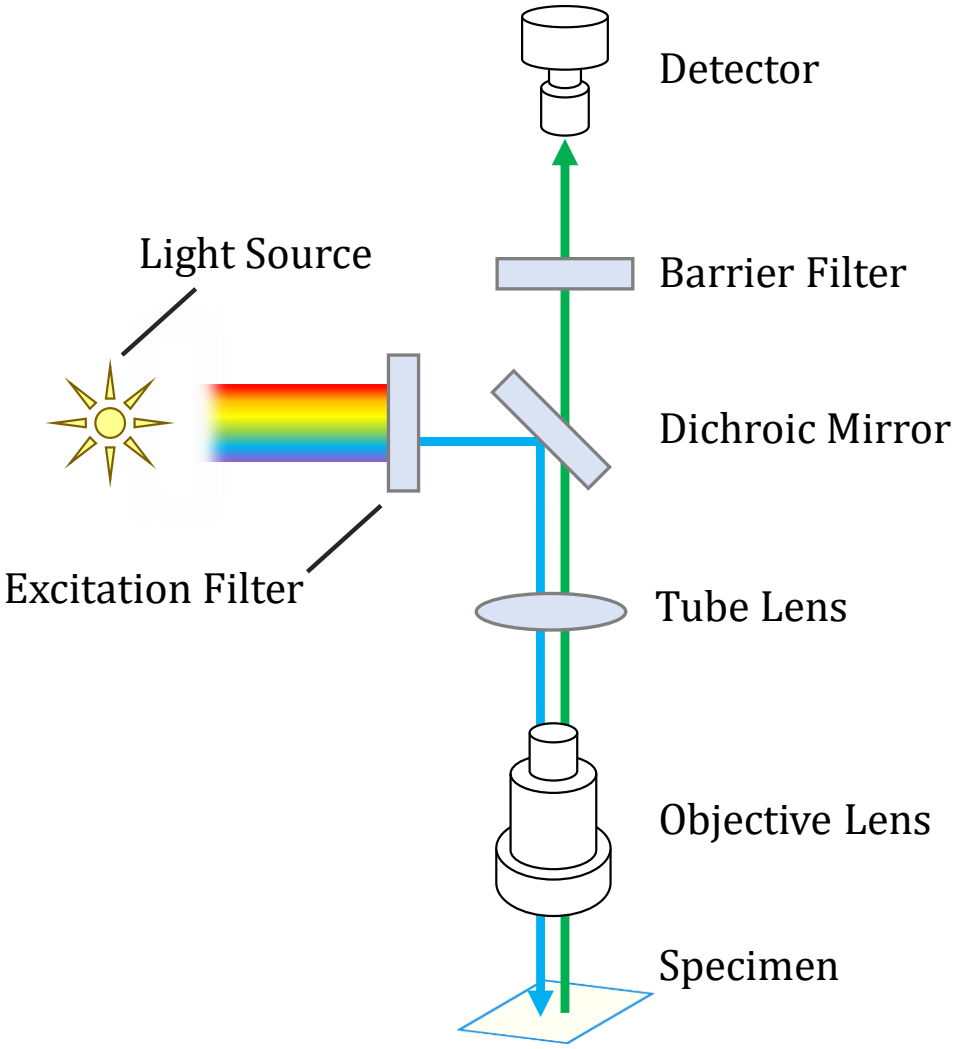


Figure 1.9: Optical diagram of a conventional fluorescence microscope. A light source is filtered through an excitation filter to isolate the excitation wavelength. The laser is reflected by a dichroic mirror, specific to the excitation wavelength, through the tube lens, the objective lens, and then onto the sample. Fluorophores are excited and emit light at a higher wavelength. Fluorescence is passed back through the dichroic mirror and transmitted across a barrier filter, which blocks reflected excitation light. Variations in intensity are recorded by a detector.

a barrier or emission filter, to block out any unwanted wavelengths, and onto a detector for image digitisation (**Figure 1.9**)[93]. In an idealised fluorescence microscope, only emitted light should reach the detector – the type of detector used, such as a charge-

coupled device (CCD) or scientific compound metal-oxide semiconductor (sCMOS) camera, will depend on the imaging hardware[135].

The diffraction limit and resolution

When a single molecule (a point source) fluoresces, photons are emitted isotropically in a spherical wavefront[140]. As the wavefront approaches the circular aperture of an objective lens, it experiences Fraunhofer diffraction, and the resultant intensity distribution at the back focal (Fourier) plane of the objective presents as a top-hat function in frequency space[140]. The inverse Fourier transform of a top-hat function produces an amplitude which is a sinc function, and the intensity is a sinc squared function at the detector – this is known as an Airy disk[138]. As a result, fluorescing molecules appear as a diffraction pattern of light instead of an infinitely small point, which creates blur when several fluorophores are imaged in close proximity[141]. The functional form of the three-dimensional intensity distribution of this diffraction pattern is termed the point spread function (PSF)[141]. The resolution of FM is defined as the smallest distance at which two point sources can be distinguished as independent objects and is fundamentally limited by the diffraction of light[127]. The diffraction limit of an objective is dependent on its ability to collect light and resolve detail at a fixed distance from the sample – this property is known as the numerical aperture[142] (NA), typically quantified between 0.5 and 1.5. The diffraction-limited resolution of an optical system was characterised by Ernst Abbe in 1873 and is given by,

$$d_{xy} = \frac{\lambda}{2\mathrm{NA}},$$
$$d_z = \frac{2\lambda}{\mathrm{NA}^2},$$

where λ is the wavelength of light, d_{xy} is the lateral resolution, and d_z is the axial resolution[142]. Note that the PSF is elongated along the optical axis (z), owing to the

asymmetry of the wavefront originating from the microscope objective and, as a result, axial resolution is typically worse than lateral[138]. Even with an idealised NA of approximately 1.5, visible light is constrained to a wavelength of at least 380nm[93], which places a theoretical lower bound on the axial resolution of approximately 130nm. In practice, this is rarely achieved, with modern optics systems consistently resolving at 200nm or more across the focal plane[143].

Variants of conventional fluorescence microscopy

To capture the depth of a specimen, the sample must be imaged over several thin slices[144]. Optical sectioning is a system for achieving 3D imaging which works by removing fluorescence detected from regions outside of the focal plane[144]. By repeating this process over a subset of axial positions, a sample can be split into several planes and reconstructed into a scan or z-stack[144]. FM can also be used for live-cell imaging provided phototoxicity is accounted for and a hospitable environment is provided for cells to carry out their metabolic and physiological functions[145]. Lowering the laser intensity will counteract the effect of phototoxicity, but also deteriorate the visible distinction between the sample and the background[145]. This may reduce the signal-to-noise ratio (SNR) of the system, which can lower image contrast[145]. Further, built-in image processing techniques such as line averaging may also require additional computation time, which increases the acquisition time between frames[145]. As such, temporal resolution generally comes at the expense of spatial resolution, and can range between microseconds and minutes[145].

FM systems will typically be selected to optimise particular optical or experimental parameters. For instance, laser scanning confocal microscopy illuminates one diffraction-limited volume of the image at a time, accepting signal only from that region,

which may then be averaged over a given period of time to improve SNR[146]. This excludes secondary fluorescence from regions outside the focal plane while achieving maximum resolutions of ~180nm laterally and ~500nm axially[146]. However, confocal suffers from longer acquisition times and may be inappropriate for imaging with pronounced cellular motility or drift[99]. As light transitions from a medium of greater refractive index (such as the sample) to one of lesser refractive index (such as the glass coverslip) at a specific incident angle, it undergoes reflection instead of refraction[92]. Total internal reflection fluorescence (TIRF) microscopy exploits this principle to produce an evanescent field, which penetrates no more than 100nm into the sample and diminishes rapidly with distance, confining fluorophore excitation to those in close proximity[92]. This enhances axial resolution (~100nm isotropically) and improves SNR, but is limited by a shallow penetration depth and stringent sample preparation requirements[92].

1.2.2 Single molecule localisation microscopy

Conventional FM is a fast and relatively cheap imaging modality, which gives suitable spatial and temporal resolution for studying microscale cellular processes[147]. However, colocalisation and co-clustering of membrane proteins typically occurs at scales below the diffraction-limited resolution[148]. Methods in super-resolution microscopy, such as single molecule localisation microscopy (SMLM), can overcome this limit[148]. Instead of the grid-like images derived from conventional FM, SMLM can provide a series of molecule localisations as output[96]. These localisations are aggregated into point clouds which may also express additional information, such as localisation precision, photon count, or data regarding quantitative properties of the fluorophore environment[96].

The principles of single molecule localisation microscopy

The fundamental principle of localisation-based super-resolution microscopy is to temporally separate individual PSFs, avoiding ensemble activation of multiple fluorophores, estimate each molecule position, and reconstruct the underlying structure by painting fluorophores in a pointillist fashion [138]. By exploiting the fluorescence intermittency or blinking properties of point sources, it is possible to image individual fluorescing point sources[149]. Furthermore, it has been shown that the pixelated PSF of an individual point source can be recapitulated by a Gaussian or Airy function, whose central position estimates the spatial position of the emitter, and whose standard deviation is a measure of the localisation uncertainty[150]. By stochastically activating sparse subsets of fluorophores in each image frame, SMLM can achieve temporal separation of PSFs and limit the probability of PSF overlap[151]. Once acquisition is complete, each frame is processed individually to localise all fluorescing molecules[152]. Molecule positions (taken as point coordinates) and localisation uncertainties are recovered from each frame, with all points overlaid and aggregated into one distribution[128].

Methods in SMLM

Diversity in target structures and experimental conditions has given rise to a range of methods in SMLM, in which each modality presents its own advantages and disadvantages, depending largely on the type of acquisition required. For example, photoactivated localisation microscopy (PALM) makes use of photoactivatable fluorophores, which switch to an active state under UV illumination and then become irreversibly photobleached[153,154]. By exciting a small region at low UV power, only a

sparse subset of fluorophores will reach emission simultaneously – this process is repeated until a sufficient number of localisations has been recorded[153]. Labelling in PALM is typically undertaken by genetically encoding fluorescent proteins, such as EosFP, mEos, GFP or PS-CFP2[154]. Since only low power irradiation is used, PALM is suitable for live cell imaging and can achieve resolutions of ~20nm laterally and ~50nm axially[96]. In fluorescence, the process of transitioning from an active (emitting) state to an inactive (dark) state, or vice versa, is known as photoswitching[95]. Stochastic optical reconstruction microscopy (STORM) uses photoswitchable fluorophores, which are forced into a stable dark state, in the presence of buffers containing thiol reducing agents and oxygen scavengers, by laser illumination[95]. Then, in the presence of redox buffers, a low powered UV laser activates a small subset of fluorophores[95]. Pairs of dyes are often used, with one molecule acting as a primary fluorophore, while the other is used as a facilitator, which induces photoswitching of the primary[95]. In STORM, labelling is often achieved with immunolabelling (**Figure 1.10**). Direct STORM (dSTORM) is a variant of STORM which uses a single photoswitchable fluorophore that reversibly switches between active and inactive states, depending on the wavelength of irradiation. dSTORM can image cellular structures with resolutions of approximately 20nm, without the use of a facilitator[155]. In this instance, laser illumination induces an inactive dark state in fluorophores, which then stochastically return to their fluorescent ground state[155].

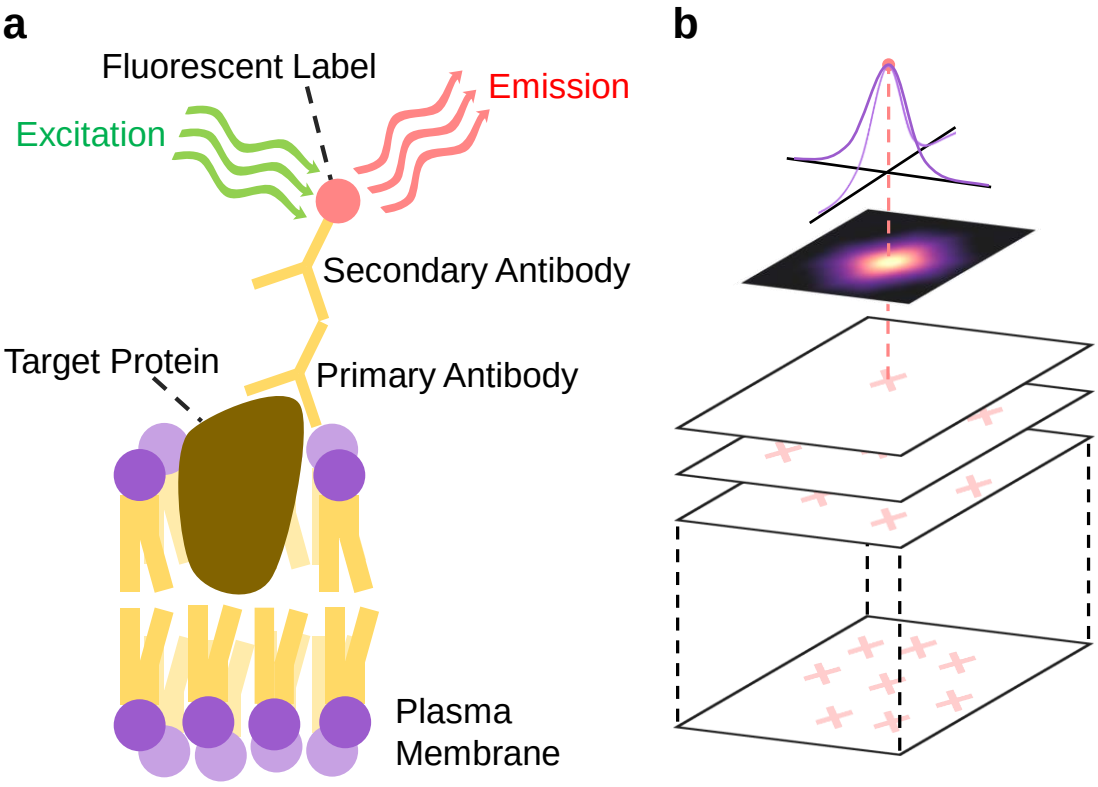


Figure 1.10: The workflow of stochastic optical reconstruction microscopy (STORM). **a** Fluorophores are attached to membrane proteins by immunofluorescence labelling. **b** Individual PSFs are recorded over several frames with Gaussian fitting used to localise molecules in each image. Localisations are combined to give a point cloud.

1.2.3 DNA-PAINT

Point accumulation for imaging nanoscale topography (PAINT) is a form of SMLM which makes use of fluorescent ligands which diffuse freely unless bound to a target molecule and immobilised[141]. A standard variant of PAINT, known as DNA-PAINT, uses a hybridisation probe comprised of a fluorescently-labelled DNA fragment[156]. These fragments are termed as imager strands and bind specifically to their complements, known as docking strands, which are attached to target molecules[156]. Dye-labelled ligands diffuse rapidly, and may cross a space corresponding to several pixels between acquisition frames, giving them the appearance of low-intensity blurred

streaks, which can be filtered out of image frames by thresholding[156]. When the probe is immobilised for a prolonged period of time, the detector can accumulate a emitted photons for detection, which produces a stable PSF[141]. PAINT microscopy is advantageous for circumventing photobleaching, since stable fluorophores can be continuously replenished in the sample, but suffers from high background fluorescence signal and long acquisition times in dense samples[157]. DNA-PAINT depends on stochastic binding and unbinding of imager and docking strands – these binding events are predictable, and imager strands exhibit characteristic fluorescence on- and off-times[158]. PSF separation is therefore regulated by controlling the concentration of the dye, rather than by exploiting the complex photophysical properties of fluorophores[158].

Advancements in DNA-PAINT

In 2014, Jungmann et al introduce Exchange-PAINT, a multiplex imaging approach wherein orthogonal imager strands are sequentially applied to the same sample[159]. This allows for sequential imaging of multiple targets using only a single dye and a single laser source[159]. This was succeeded by the development of qPAINT, which uses the predictable second-order association kinetics of imager strands to their docking strands to obtain absolute quantification of molecule numbers[158]. More recently, Schueder et al developed proximity PAINT (pPAINT), an extension to DNA-PAINT, which uses a split-docking site configuration to detect spatial proximity between target biomolecules[160]. This modality can used to detect interacting protein pairs with both high sensitivity and accuracy[160]. Ostersehlt et al demonstrated an analogous DNA-labelling principle for use with MINFLUX (minimal fluorescence photon fluxes) SMLM for 3D nanoscale (<5nm) imaging of multiple molecular

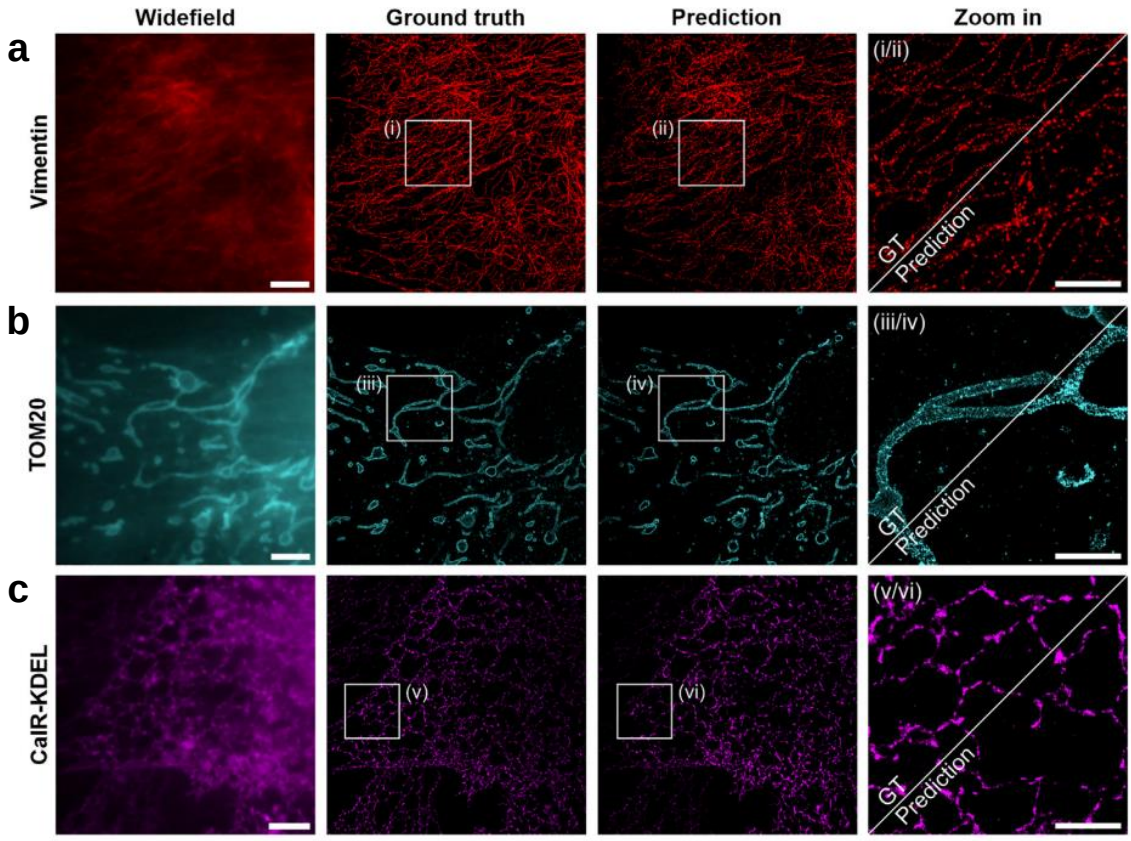


Figure 1.11: An example of SMLM supplemented by machine learning. Fast DNA-PAINT of cellular proteins **a** Vimentin, **b** TOM20 (mitochondria) **c** CalR-KDEL (endoplasmic reticulum), imaged in fixed U-2 OS cells. Cells and organelles were labelled with fluorescent probes JF_{635}-S5 (**a** and **c**) and 5 nM SiR-Hy4 (**b**). Wide-field and conventional PAINT (ground truth) images are given for each data set. Predicted super-resolution images were generated by neural network-assisted DNA-PAINT (HT-PAINT). Zoomed regions for each image are marked with corresponding numerals and show the relative similarities between ground truth and HT-PAINT output. HT-PAINT images were acquired ~25-fold faster than conventional PAINT. Scale bars are 5µm (overview) and 2µm (zoom-in). Figure adapted from Jang et al[165].

targets[161]. Recently, Narayanasamy et al developed a neural network-based approach which can predict fluorophore positions from high emitter density DNA-PAINT and reduce image acquisition time[162]. Further, Kessler et al introduced a short-distance self-quenching mechanism for fluorophore dimers to reduce background fluorescence

signal in DNA-PAINT[163]. Schueder et al report 10-fold faster imaging speeds, without compromising image quality or spatial resolution, by designing optimised DNA sequences and buffer conditions[164]. Jang et al built upon these principles by incorporating direct protein label, HaloTag7, in combination with exchangeable ligands and the DeepSTORM neural network (**Figure 1.11**) to reduce DNA-PAINT acquisition time (up to ~25-fold, or in the order of seconds)[165]. This permits fast, live-cell compatible imaging which bypasses photobleaching by providing a constant signal over time[165]. High-resolution and precise molecular quantification make DNA-PAINT a particularly advantageous form of SMLM. In **Chapter 3**, we use DNA-PAINT to localise the TCR-CD3 complex on T cell membranes, and in **Chapter 4**, we introduce an adapted ratiometric PAINT modality for use with solvatochromic probes.

1.2.4 Localisation and resolution

During pre-processing of raw SMLM data, each image frame is analysed independently (**Figure 1.12a-b**). First, correction methods such as rolling ball algorithms, difference of Gaussians, thresholding, or wavelet filtering may be used to increase contrast between active emitters and the background (**Figure 1.12c**)[166]. The approximate position of emitters is then determined – this may be achieved by taking local maxima in intensity values within a connected 8 neighbouring pixel range[96]. A localisation algorithm is used to estimate the spatial coordinates of the emitters, typically by fitting of PSF models to the intensity distributions of point sources (**Figure 1.12d**)[167]. This may be achieved by Maximum Likelihood Estimation (MLE), which localises each molecule by identifying the positions for which the likelihood of obtaining the observed image is maximised[167]. The algorithm is computationally more expensive than non-iterative algorithms, but advances in graphics processing units mean modern

imaging systems can perform MLE in real time[167]. The precision of MLE depends on both the SNR and the accuracy of the model PSF, although PSFs can be calibrated experimentally and often outperform models in this case[167]. The accuracy of PSF fitting may be further improved by implementing dynamic spline PSF models[168]. The localisation uncertainty (sometimes denoted localisation precision) refers to the degree of uncertainty surrounding true emitter position, and is quantified as the difference between the true position and the coordinates returned by the localisation algorithm[169]. This directly influences the achievable spatial resolution and primarily depends on setup-dependent experimental conditions (e.g. optics, objectives and hardware), sample-dependent experimental conditions (e.g. SNR, photon count and blinking) and the registration potency of the algorithm used (**Figure 1.12e**)[169]. For a PSF with standard deviation σ_0 and photon count N, defined in an image with pixel size α and mean background intensity β, the amended Cramér-Rao lower bound (CRLB) is given by,

$$\sigma_{loc} \geq \sqrt{\left(\frac{12\sigma_0^2 + \alpha^2}{12N}\right)\left(\frac{16}{9} + \frac{8\pi\sigma_0^2\beta^2}{\alpha^2 N^2}\right)},$$

where σ_{loc} is the uncertainty of the corresponding localisation[170]. Theoretically, MLE approaches the CRLB as SNR increases[167]. To estimate the precision of an SMLM modality experimentally, multiple images can be taken of a fluorescing fiducial marker (e.g. TetraSpeck beads), which can be localised across all frames – the spread of all localisations then serves as an estimate for localisation uncertainty[171].

The resolution of an SMLM acquisition depends on such parameters as localisation precision, labelling density, fitting performances and temporal shifts[172]. One method of estimating resolution is to apply the SMLM modality to a data set where molecules

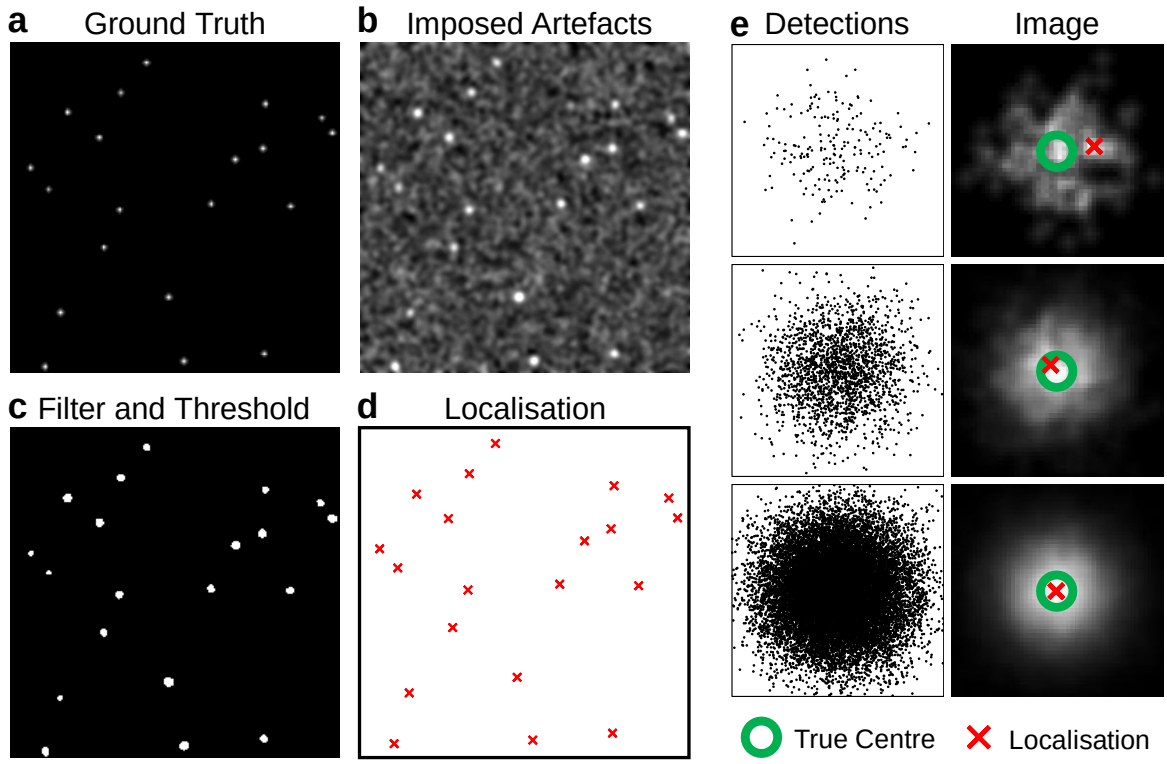


Figure 1.12: Schematic of filtering and localisation workflow. **a** A simulated data set of ground truth localisations. **b** The impact of blur and noise. **c** A filtering algorithm reduces background, and thresholding binarises the image to highlight PSFs. **d** Localisations are derived from the filtered image to negate the effect of imaging artefacts. **e** The impact of photon count on SNR and the observed PSF. The higher the photon count, the more easily identifiable the PSF, and the greater the precision of the localisation. Localisation estimates (red) move closer to the true molecule position (green) as photon count increases.

are separated by a known distance, such as in nanorulers or nuclear pores. Computationally, resolution can be estimated using the Discretised Fourier Transform (DFT), which maps the image into a Fourier space[173]. Taking the resulting space at the logarithmic scale highlights an approximately circular central mass, with radius equal to the scale at which information content disappears into background noise – this serves as an estimate of resolution for the image[173]. This process is known as Fourier Spectral Analysis, and has the advantage that it does not depend on SNR[173].

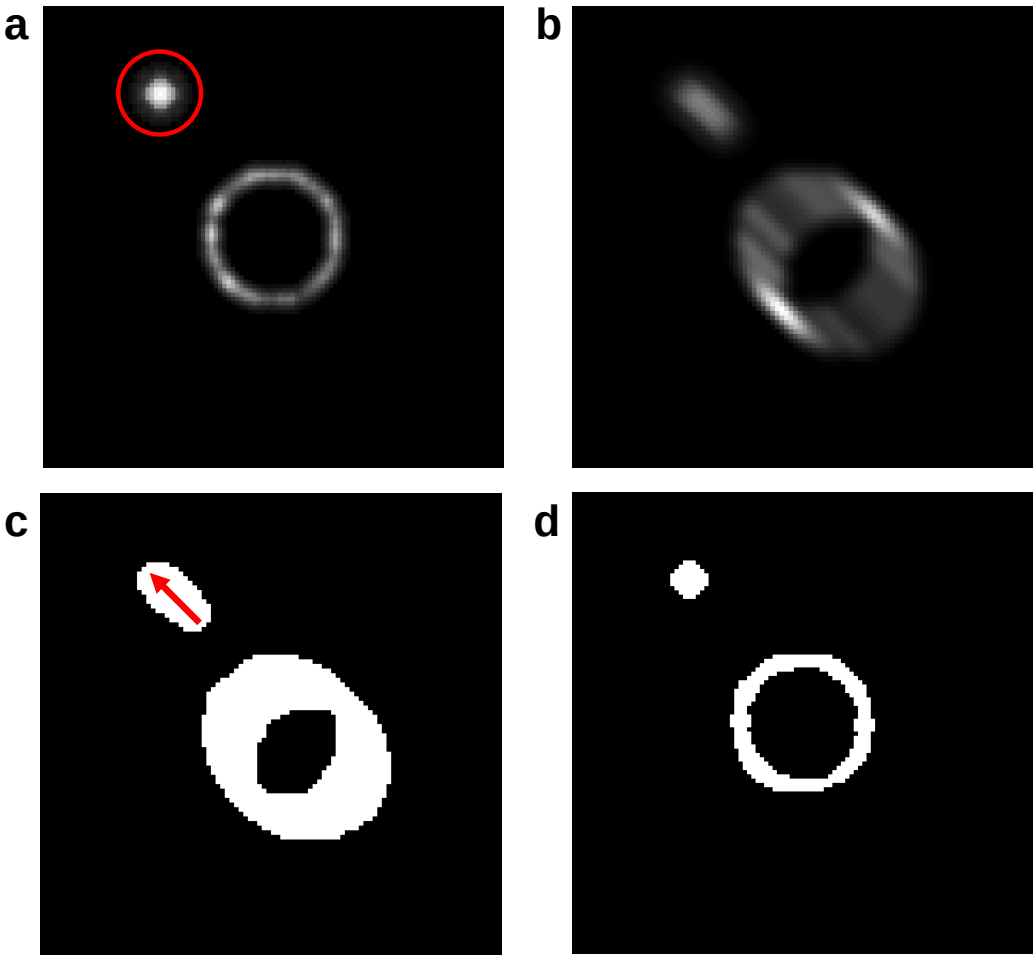


Figure 1.13: The impact of drift, and correction with fiducial markers. **a** A fiducial marker (circled in red) is added to a sample with a fluorescing ring-like structure. **b** As a result of drift, the structure appears blurred and stretched in the acquired image. **c** Binarising the image highlights the impact of the artefact on the observe structure. The difference in position of the fiducial marker can now be traced backwards (red arrow). **d** By tracing the dynamics of the marker, localisations can be offset to their true position.

1.2.5 Artefact correction

In microscopy, a visual artefact is a structural feature of the data which is not a legitimate feature of the sample and instead arises from experimental conditions, data acquisition or processing methods[96]. Artefacts in SMLM may distort the representation

of molecular structures and dynamics. At the nanoscale, small variations in specimen position can produce large localisation offsets – this phenomenon, known as drift, depends on the speed of the imaging system and typically occurs at the range of tens of nanometres (**Figure 1.13**)[174]. The offset of drift can be estimated by tracking the displacement of fiducial markers across all image frames[175]. Emitter positions can then be more accurately determined by subtracting drift. Alternatively, piezoelectric actuators can automatically recalibrate the position of the sample throughout acquisition[175].

Stochastic blinking of individual fluorophores across multiple frames may cause individual emitters to be localised in multiple times – this process, known as multiple blinking, may induce pseudoclusters which skew spatial statistics (**Figure 1.14**)[176]. To correct this, multiple fluorescent probes may be used to localise the same target molecule, then two-colour colocalisation can discriminate against false-positive detections[97]. Alternatively, computational methods may be used to aggregate adjacent localisations across frames into a single spatial localisation[176]. These techniques may yield an increase in precision but can sacrifice sample density if too many emitters are merged – if labelling density is too sparse, fine structural properties will not be resolved, irrespective of precision[147]. However, the higher the labelling density, the greater the probability of PSF overlap, which may perturb localisation uncertainty and alter the number of emitters detected[177]. This can be difficult to avoid, especially in cases when a high activation probability is required for fast imaging[177]. Depending on the imaging modality used, a trade-off between fluorophore density and laser intensity can be finetuned[177]. Alternatively, localisations may be filtered out by multi-emitter fitting algorithms, designed to detect overlap and calculate differences between consecutive frames[178].

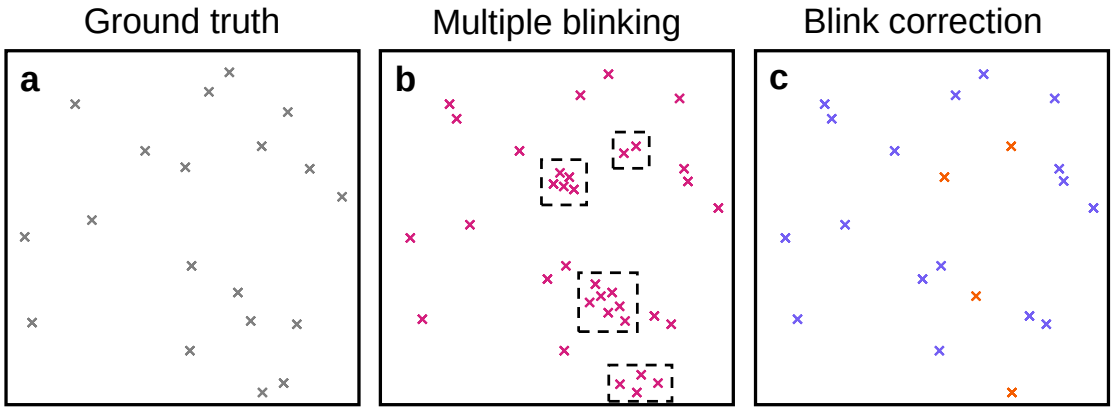


Figure 1.14: Multiple blinking artefacts that were not present in the ground truth data (**a**) may produce an excess of localisations which may induce pseudoclusters (highlighted in **b**). This arises as a result of the stochastic nature of fluorescence, which causes a single fluorescent molecule to emit over consecutive acquisition frames. Blink correction algorithms detect instances of blinking (**c**) and aggregate false positive localisations (orange).

1.2.6 Super-resolution image reconstruction

Once all localisations have been mapped and corrected for artefacts, they may be aggregated into a point pattern representing molecule coordinates, for which each point has an associated localisation uncertainty. From here, a super-resolution image can be reconstructed for the purpose of data visualisation or prior to applying further image analysis. One way of achieving this is to generate pixels with size equal to the average localisation precision. The intensity of each pixel may be proportional to the number or total photon count of all localisations that fall within its volume. Alternatively, the image plane may be divided into a fine grid of pixels (typically 5-20nm) and each localisation may be represented as a Gaussian with mean position equal to the estimated molecule centre, amplitude proportional to the photon count, and standard deviation equal to the localisation uncertainty[179]. Labelling density has a pronounced effect on the fidelity and resolution of the reconstructed image[138]. If we consider the

super-resolution image as a series of digitised points which sample a true continuous structure, then by the Nyquist-Shannon sampling criterion, the sampling interval must be greater than at least twice that of the highest spatial frequency of the sample to accurately reconstitute the continuous structure and preserve spatial resolution[180,181]. Therefore, if the labelling density is lower than the Nyquist interval, the structure will be under-sampled and the image will appear fragmented.

1.2.7 Machine learning-assisted SMLM

The magnitude and dimensionality of SMLM data sets makes them appropriate targets for augmentation with machine and deep learning methodologies. Speiser et al developed DECODE (deep context dependent), a deep learning-based tool which assists in localising single molecules at high emitter density[182]. More recently, Saguy et al introduced DBlink, a spatiotemporal interpolation method for generating super-resolved video reconstructions from SMLM data[183]. Such techniques have, however, garnered concern – as interpolation, although potentially accurate, is often synthetic and prone to hallucinations[184]. Gómez-de-Mariscal et al argue that machine and deep-learning should be used to extract insights from “gentle” imaging and prioritise minimising photodamage instead of recovering compromised data from harsh, phototoxic illumination[185]. The authors propose that this will permit observation of undisturbed living systems, and not just synthetic data retrieval[185]. Notably, a large, international consort of researchers, under von Chamier et al, released an entry-level platform for training deep learning networks, known as ZeroCostDL4Mic[186]. This has allowed researchers to train analytic tools for microscopy by leveraging free cloud-based computational resources, in an effort to overcome accessibility barriers[186].

1.2.8 Insights into lipid rafts from super-resolution microscopy

Super-resolution microscopy has been used to visualise lipid-mediated protein clustering and dynamics, especially in the context of proteins which exhibit ordered-phase affinity. Sengupta et al used PALM with pair-correlation analysis to image the nanoscale organisation of GPI-anchored proteins in COS-7 cells and found preferential localisation in raft domains[14]. The authors determined that these proteins form clusters in the ordered phase and that the number of proteins per cluster decreased after depletion of cholesterol with methyl-beta cyclodextrin (MβCD)[14]. Eggeling et al used STED-FCS (Stimulated Emission Depletion Fluorescence Correlation Spectroscopy) to detect single diffusing lipid molecules in the plasma membrane of living epithelial cell line PtK2[187]. The authors determine that sphingolipids and glycosylphosphatidylinositol (GPI)-anchored proteins, which are enriched in lipid rafts, are transiently trapped in cholesterol-mediated molecular complexes (for ~10–20ms) within <20nm diameter areas[187]. On the contrary, Sevcsik et al used protein microtracking and single particle tracking with PALM to show that phase partitioning is not a fundamental element of GPI-anchored protein organisation in the plasma membrane[188]. Sako et al demonstrated single-molecule imaging and tracking of epidermal growth factor receptor (EGFR), which is known to be a raft-associated protein[189], on A431 cell surfaces[190]. The results suggest that lipid rafts are not static entities, but exhibit transient and dynamic behaviour[190]. Owen et al utilised PALM and direct STORM (dSTORM) to quantify the spatial heterogeneity of raft-associated proteins LAT and Lck in T cells[191]. The authors find that these proteins form distinct nanoclusters within lipid rafts, which is a prerequisite for T cell signalling[191]. Sezgin et al suggest that enabling the efficient use of polarity-sensitive probes in super-resolution microscopy will be an important future development in the

study of lipid rafts[48]. In this vein, Bongiovanni et al propose spectrally-resolved PAINT microscopy, which works with environmentally-sensitive probes like NileRed to acquire both the nanoscale spatial coordinates and emission spectrum of single fluorophores[171]. However, in order to draw further conclusions from super-resolution data, appropriate quantification methods must be used, as discussed in the following section.

1.3 Spatial statistics

1.3.1 Spatial point patterns

The output of SMLM takes the form of a table of values which may include the spatial positions of each localisation, photon counts, estimated precision, and information derived from environmentally-sensitive dyes[147]. The spatial coordinates alone generate a spatial point pattern – a list of estimated molecule positions which may be visualised as a scatter plot or point cloud (**Figure 1.15**). This information allows for the visualisation, quantification and interpretation of molecular organisation on the plasma membrane. Analysis methods specific for SMLM extract quantitative information from the primary coordinate-based data, including underlying geometry, topology, and number of proteins[147]. Spatially-descriptive statistics can be used to probe the underlying geometric properties of the point pattern[192].

1.3.2 Spatially-descriptive statistics

Spatially-descriptive statistics do not necessarily provide a partitioning of the data, but can quantify spatial organisation in the point distribution and inform parameters for cluster analysis algorithms[192]. The Ripley's functions are spatially-descriptive statistics

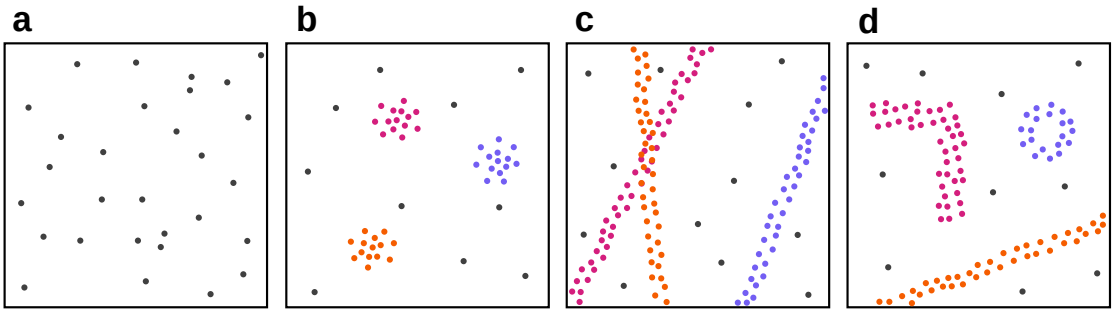


Figure 1.15: Example spatial point patterns displaying a range of cluster geometries, including **a** a completely spatially random or uniform distribution, **b** Gaussian clusters, **c** fibrous clusters, and **d** a mixed (compound) distribution of cluster geometries and topologies.

which detect deviations from spatial homogeneity[192]. For an ROI of area A, containing n points, the Ripley's K function is defined as,

$$K(r) = \frac{A}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \delta_{ij}(r).$$

Here, $\delta_{ij}(r)$ evaluates to 1 if the distance between points i and j is less than r, and 0 otherwise. This is equivalent to counting the number of neighbours at each radius, averaging over all points and scaling by area. This result is normalised to the L function,

$$L(r) = \sqrt{\frac{K(r)}{\pi}},$$

which acts as an intermediary for the H function,

$$H(r) = L(r) - r.$$

Positive values of $H(r)$ suggest greater point density and clustering at radius r, while negative values suggest dispersion and homogeneity. Values close to 0 suggest a completely spatially random (CSR) distribution. The argmax of the H function is known

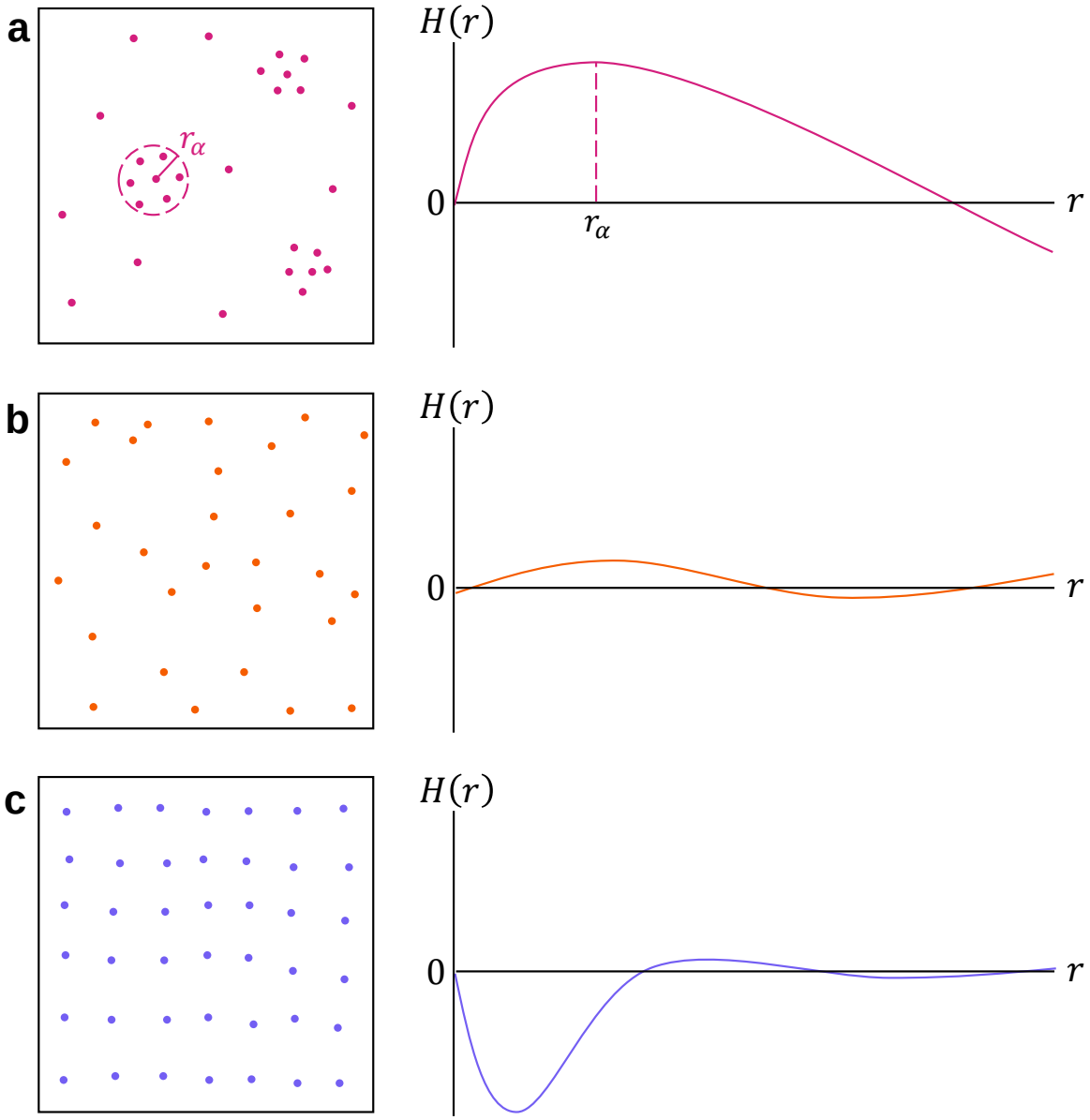


Figure 1.16: Plots of the Ripley's H function for varying data geometries. **a** A clustered distribution (magenta) presents as a notable peak in the H function. The argmax of this peak is known as the radius of maximal aggregation, r_α, and serves as an estimate for cluster radius. **b** A CSR distribution (orange) yields a relatively flat H function with no discernible peak, while **c** a homogeneous distribution (purple) creates a prominent dip.

as the radius of maximal aggregation[191] and serves as an estimate for true cluster size, if clusters are present and known to be circular (**Figure 1.16**). Owen et al used

image analysis based on Ripley's functions to quantify the distribution and heterogeneity of raft-associated proteins LAT and Lck in PALM and dSTORM data[191]. The Pair Correlation Function (PCF) is a derivative of the Ripley's function which defines the probability of finding a point at a specific distance from another point, relative to what would be expected in a completely random distribution[193]. This is achieved by considering a reference point and counting the number of points which fall within distinct concentric rings at increasing distances[193]. The count in each ring is then normalised by the area of the annulus and the average density of points across the ROI[193]. The functional form of the PCF is given by,

$$g(r) = \frac{A}{\pi n(n-1)r^2} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \frac{\delta\left(\Delta r/2 - \left|(r + \Delta r/2 - r_{ij}\right|\right)}{\Delta r},$$

where r_{ij} is the distance between points i and j, and Δr is the ring thickness. Here, δ evaluates to 1 if r_{ij} is within a range Δr over the inner annular radius r, and 0 otherwise. This provides a relative probability of finding a point at a specific distance from a reference point, instead of within a given radius. A higher value of the PCF at a particular radius indicates a clustering of points at that distance, while a lower value suggests avoidance. In practice, Sengupta et al used pair-correlation to analyse GPI-anchored protein distributions in PALM data and showed changes in cluster properties after depletion of cholesterol with MβCD[14].

1.3.3 Overview of cluster analysis and unsupervised machine learning

The term cluster refers to a non-empty subset of a point pattern in which all members share a degree of similarity[194]. The notion of distance is central to data clustering algorithms and represents this degree of similarity between data in a given feature space[194]. A cluster analysis algorithm takes point data as input and aims to output a

partitioning of the data such that each point is assigned to a cluster based on some pre-defined metric[194]. Since both the input and output data is unlabelled, cluster analysis is considered a form of unsupervised machine learning[194]. The choice of algorithm may depend on the assumptions of the method and the parameters required. Quantitative studies of molecule clustering from noisy SMLM data requires accurate, unbiased and precise estimation tools[195], but it is not always clear which, if any, algorithm will be most appropriate for the data set. This decision will usually be made on qualitative grounds; however, quantitative comparisons can be made for pre-determined data geometries, which justify the use of particular algorithms. There are generally two forms of cluster analysis: global clustering, which provides spatially-descriptive statistics regarding the data (e.g. Ripley's functions or PCF), and complete clustering, which deterministically partition each data point into a specific cluster[194]. Complete clustering methods cannot explicitly detect the presence of clusters and can give erroneous results if applied to data without spatial structure[196]. It is therefore recommended to first undertake global clustering to determine whether clusters can be identified, then apply complete clustering to partition the data. Once a partition is identified, within- and between-cluster analysis can be performed. Within-cluster analysis may incorporate probing of the size, shape and density of individual clusters, while between-cluster analysis can determine relative differences between these properties among different clusters from the same data set. A multitude of cellular processes are driven by the assembly of biomolecules into clusters, and partitioning can highlight the important biological phenomena underpinning molecular organisation[148,197].

Cluster properties and estimation methods

Cluster analysis algorithms may require information about the expected properties of clusters in the data space[177]. These properties include, but are not limited to, the number of clusters, cluster geometry and topology, cluster size, interpoint distances, and typical point density within clusters[196,198,199]. If assumptions are imposed on the geometry or organisation of clusters, fewer parameters may be required[196]. For instance, k-means clustering[200] assumes that there is a finite number of distinct Gaussian clusters present in the data and requires only one parameter, k. This parameter, representative of the number of clusters, may be estimated by Monte Carlo simulation, non-parametric slope statistics, MLE, prediction-based resampling, or persistent homology[201,202]. However, no choice of parameter will circumvent the hyperparameters – that is, the restriction to Gaussian geometry. Other cluster analysis methods may require information regarding the expected size or density of clusters, if not both. Assuming circularity, the radius of maximal aggregation derived from the Ripley's H function can serve as an estimate for cluster radius, although this has been shown to vary non-linearly with true cluster size, bringing increasing inaccuracy at larger scales[195]. Density, in the context it is required for cluster analysis, may be harder to quantify[203]. Even with the true cluster radius known, the density may vary across the ROI, in which case the average density may not be a suitable estimate[203]. Furthermore, depending on the algorithm used, both a minimum and maximum density may be required[199]. Density is typically estimated from known experimental outcomes, such as the expected number of molecules per cluster[203]. In terms of outputs, there are several statistics used to evaluate the performance of a clustering algorithm on simulated data, such as the Adjusted Rand Index, which assesses the proportion of points which have the same cluster classification in both the ground truth and output,

adjusted for chance[204,205]. Alternatively, the intersection over union metric determines the ratio of the intersection of the convex hulls of output clusters and the convex hulls of ground truth clusters, compared to their union[196,206].

1.3.4 Geometric cluster analysis

Geometric clustering methods circumvent parameterisation by imposing assumptions on cluster geometry to absorb variability in data. One such method, as discussed, is the k-means algorithm[200], which assumes a circular or Gaussian distribution for each cluster (**Figure 1.17**). Under this assumption, each cluster pertains a central mean position around which each associated point is distributed subject to a quantifiable variance. Given an expected number of clusters k, the algorithm starts by initialising k centre points, randomly distributed across the data space, and assigning each neighbouring point to a centre based on which it is closest to[200]. Once all points have been assigned, a new mean centre is calculated from their positions and the process is repeated iteratively until the centres stabilise – the point assignment at convergence is taken to be the overall clustering[200]. However, the method is sensitive to the choice of initial cluster centres, and convergence is not always guaranteed. Further, all points will eventually be assigned to clusters, including outliers, and so pre-processing may be required. In the context of SMLM, k-means clustering has been used, for example, to analyse cluster properties of surface receptor HER2 on breast cancer cells imaged with PALM and dSTORM[207].

An advanced geometric approach known as Bayesian clustering[208] makes use of the probabilistic law of Bayes theorem to actively determine the probability of each point belonging to a specific cluster. Here, each point is assigned an initial cluster allocation known as the prior. This allocation may be random, but can also be chosen

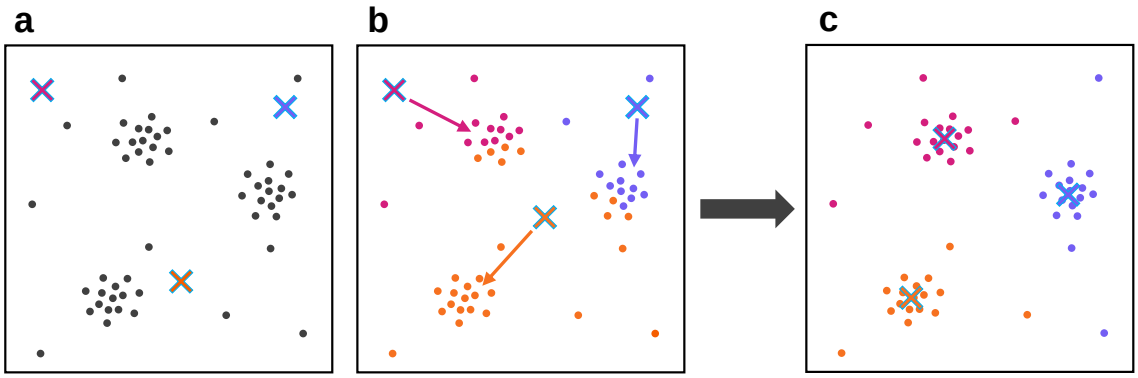


Figure 1.17: A schematic of k-means clustering. **a** Cluster centres are initialised randomly at k positions across the ROI. Here k = 3. **b** Points are assigned to the cluster whose centre they are spatially closest to. The centre is updated to the average position of all points within its cluster. **c** This process is repeated iteratively until no change in centre position is produced.

deterministically to improve convergence time[208]. Given an expected geometric distribution for each cluster, each allocation is systematically examined to determine whether a more suitable alternative exists – that is, an alternative allocation which maximises the probability of correct cluster choice[208]. After finite iterations, a posterior distribution is formed which maximises the probability of appropriate cluster allocation. This distribution is then selected as the new partitioning[208]. In practice, Bayesian clustering has been used for analysing distributions of TCR-CD3 subunit CD3ζ in STORM data[209] and LAT in 3D PALM data[210].

1.3.5 Density-based cluster analysis

Density-based cluster analysis stems from the principle that clusters produce higher than average point density relative to the background. Clusters are therefore constructed by sequentially aggregating the densest points within the ROI, which may be derived deterministically or through the use of kernel density estimation[211]. While

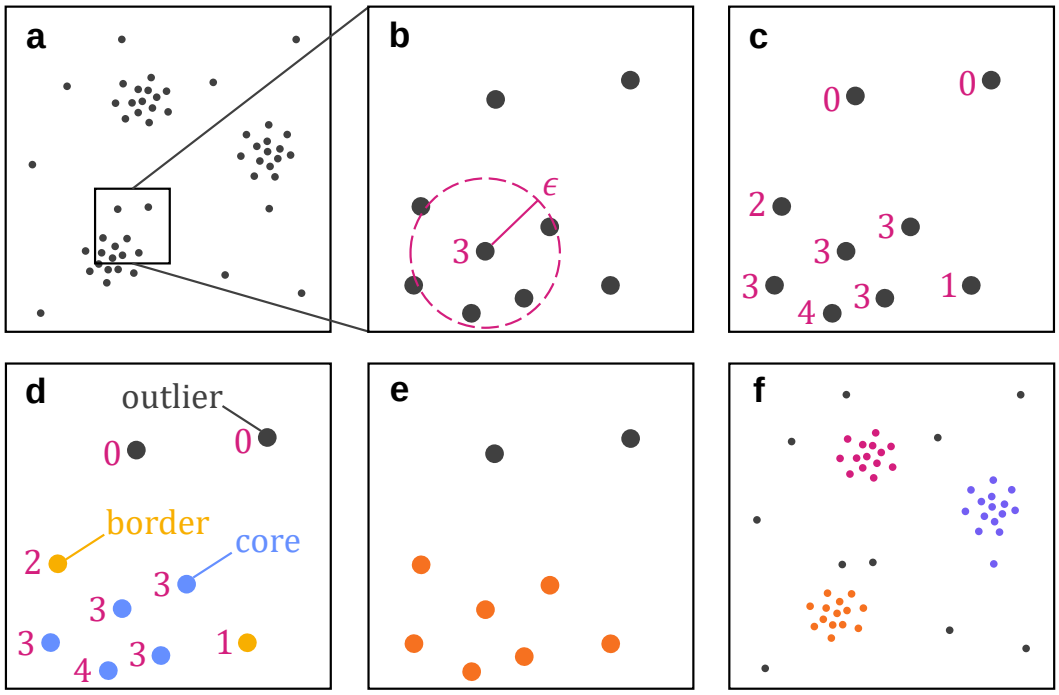


Figure 1.18: A schematic of the DBSCAN cluster analysis algorithm. **a** An example data set with Gaussian clusters. **b** Zoom in of the square region in **a**. **c** For each point, the number of neighbours within a search radius ϵ is recorded. **d** Points which have a number of neighbours greater or equal to *minPts* are denoted as core points, while any neighbours of core points are denoted as border points. Here, *minPts* = 3. All remaining points are designated as outliers. **e-f** Clusters are formed by aggregating neighbouring core and border points.

these methods are generally invariant of geometry, they come at the expense of additional parameterisation – typically requiring at least one parameter to interpret each of distance and density[198,212]. One such method is known as Density-Based Spatial Clustering of Applications with Noise[198] (DBSCAN), and requires both a distance parameter, ϵ, and density *minPts* (**Figure 1.18**). Here, any points with at least *minPts* neighbours in a radius ϵ are denoted as core points and any other points within that radius are resigned as border points – clusters are then formed by aggregating all neighbouring core and border points[198]. DBSCAN was shown to achieve the highest

IoU scores on simulated data when compared to other common clustering algorithms[196]. DBSCAN has been used, for example, to quantify GPCR oligomerisation from DNA-PAINT data[213]. Topographic approaches measure the relative densities of peaks over their surroundings. By interpolating point densities into a continuous surface, local maxima, which may correspond to clusters, can be identified[214]. In the context of SMLM, topographic prominence has been used to compare lymphocyte function-associated-antigen 1 (LFA-1) distributions in stimulated and non-stimulated T cells, imaged via dSTORM[214].

Voronoï tessellation is a method of subdividing space into a number of polygonal regions, based on the Euclidean distance, with each region centered on a data point[215]. These regions are denoted as seeds, and each seed defines a region of space closer to its point than any other[215]. As the Voronoï edges are equidistant from the two nearest seeds, the simplest way of generating a diagram of seeds is to compute the perpendicular bisectors between them – this ensures that there is no intersection between any Voronoï polygons[215]. SR-Tesseler is a density-based cluster analysis algorithm which partitions data sets based on the size of these polygons[216]. Adjacent seeds are sequentially connected up until a given area threshold, which produces a segmentation of the tessellation, and each connected component is analogous to a distinct cluster[216]. SR-Tesseler, in particular, has been used to analyse receptor organisation on neuron cell surfaces, imaged via dSTORM[216].

1.3.6 Machine and deep dearning methods

Machine and deep learning-based methods for cluster analysis are built on pre-trained models, which learn and then identify features of clusters in the data space. This is achieved by training models on simulated point patterns with known ground truth

clusters using frameworks such as self-organising maps, support vectors, or neural networks[194,217]. While the term “machine learning” traditionally encompasses both supervised and unsupervised learning, cluster analysis is inherently defined as unsupervised, and so the methodologies described here are restricted as such[194]. One such method takes as input an array of values derived from each point's nearest-neighbour distances and as outputs a binary label indicating the cluster the point belongs to[218]. This model has undergone training via a conventional neural network to maximise accuracy by adjusting internal values and comparing the given output with the expected output[218]. The model was tested on novel simulated data, including scenarios not encountered during training, and maintained high accuracy[218]. Deep Embedded Clustering (DEC) is a deep learning-based method which simultaneously learns feature representations and cluster assignments using neural networks[194]. This method devises a centroid-based probability distribution and aims to minimise the Kullback-Liebler (KL) divergence to an auxiliary target distribution to simultaneously improve clustering assignment and feature representation. This probability distribution is iteratively derived by minimising the KL divergence with respect to a pre-calculated probability distribution. Here, the KL divergence serves as a loss function and is minimised via Stochastic Gradient Descent[194]. Methods such as these address the limitations of existing computational approaches, such as handling large-scale data sets and overparameterisation. However, the resulting partitions will be limited by the variability of the simulated input data and, as a black-box approach, it is impossible to determine the criteria by which clusters are assigned.

1.3.7 Topological data analysis

Topology is the field of mathematics concerned with the discrete properties of shape[219]. In the context of point patterns, the term “topology” may be used to describe the qualitative features of spatial data[219]. Topological data analysis (TDA) lends itself to the theory of density-based clustering, as it presents methods which are inherently free of geometric constraints[219]. One method of performing TDA is to generate a simplicial complex over the data – this is a mathematical construct which may be used to characterise the topology of a point set[220]. This complex takes the form of a triangulation (**Figure 1.19a**) and comprises a set of nodes (0-cells), connected by edges (1-cells) with spaces between three adjacent edges filled by a face (2-cells), and (in three dimensions or higher) with spaces between three adjacent faces filled by a volume (3-cells)[220]. For example, the Vietoris-Rips complex is a subtype of simplicial complex which considers two nodes to be adjacent if the distance between them (typically Euclidean) does not exceed a given maximum radius parameter (**Figure 1.19b**)[220]. Here, we restrict our discussion to nodes and edges alone, as these suffice to generate higher-dimensional constructs[220].

One cluster analysis method, known as Topological Mode Analysis Tool[199] (ToMATo), builds on the TDA technique of persistent homology (PH). In theory, PH can make use of any simplicial complex[221], but since our focus here is on spatial point data, we consider only the Vietoris-Rips complex[220]. This complex can be constructed across a continuous range of spatial scales up to a pre-defined maximum, r (**Figure 1.19b**). Throughout this process, each connected component is formed, and then later absorbed by another component, save for the final mode which connects every point in the cloud. As such, each root is assigned a birth scale, the scale at which the connected component is created, and a death scale, the scale at which the component

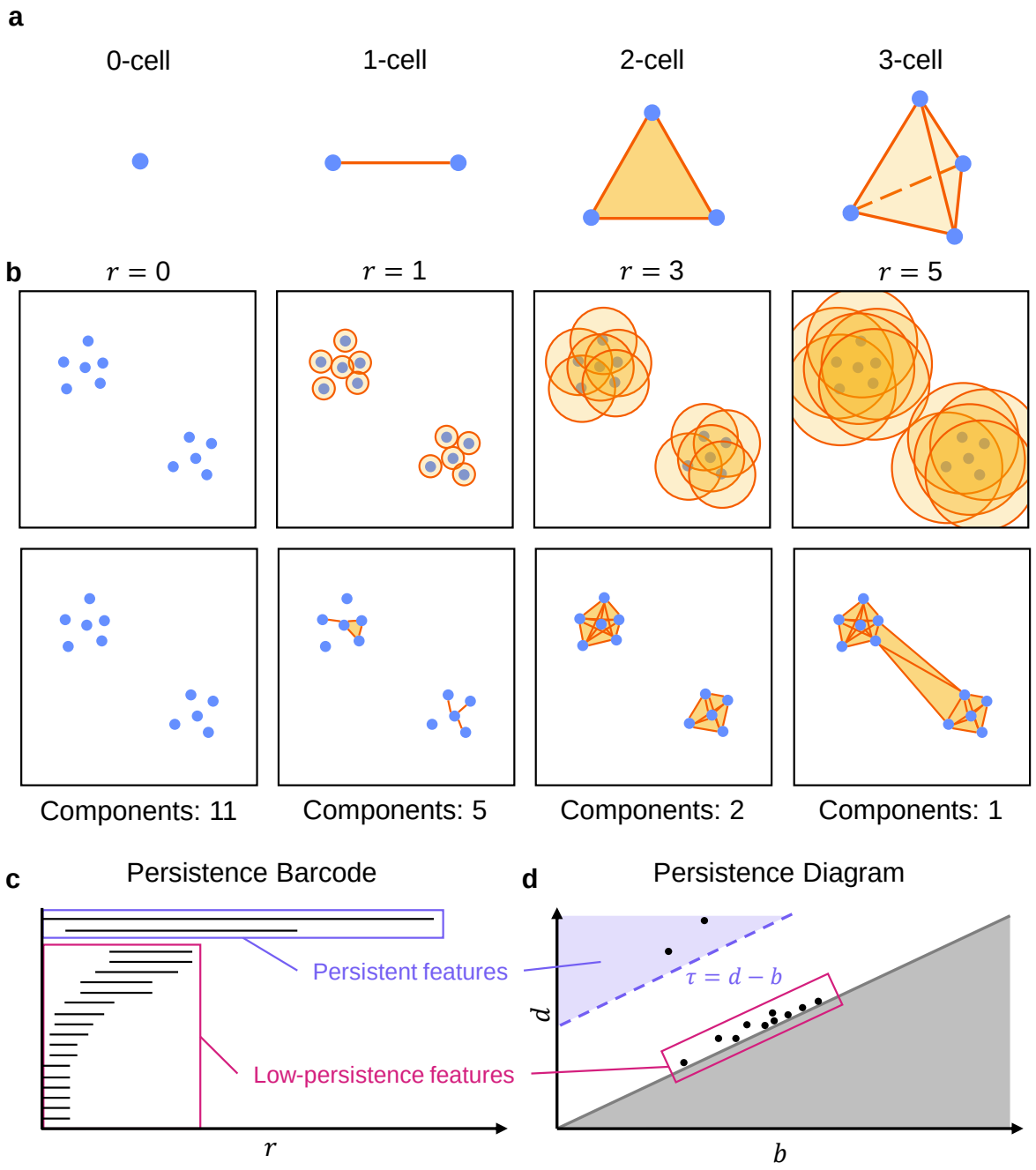


Figure 1.19: Schematic of persistent homology and the Vietoris-Rips complex. **a** The k-cells which comprise a simplicial complex. Each k-cell is the generalisation of a triangle to k dimensions. A 0-cell represents a node, a 1-cell represents an edge between 2 adjacent nodes, a 2-cell represents a face between 3 adjacent edges, and a 3-cell represents a volume between 4 adjacent faces. **b** The construction of a Vietoris-Rips complex across spatial scales. This is analogous to considering a circle of radius r around each node and taking any two nodes to be adjacent if their circles intersect. The radius at which a connected component is created is called the birth

scale, and the radius at which it is connected to a larger component is called the death scale. **c** A persistence barcode characterises each topological feature as a line with length equal to the persistence. Longer lines represent persistent features, while shorter lines represent low-persistence sub-components. **d** Birth scales, b, can be plotted against death scales, d, in a persistence diagram. The further a point lies from the bisector $b = d$, the higher feature's persistence. A persistence threshold, $\tau = d - b$, can be used to filter out low-persistence features. In both cases, two persistent features are identified, corresponding to the clusters in **b**.

connects to another. The difference between the death and birth scale is denoted as the persistence[222], which is recorded for each connected component. Topological features may be characterised by a persistence barcode (**Figure 1.19c**), in which each component is represented by a straight line initialised at the birth scale, ending at the death scale, and plotted on a 1D axis[223]. Alternatively, a persistence diagram (**Figure 1.19d**), which plots the birth scale against the death scale for each connected component, may be used[223]. In both cases, components with greatest persistence (low birth scales and high death scales) correspond to regions of the space which form at low radii and do not merge with the background until substantially high scales are reached. A persistent threshold, τ, can be introduced to filter out features which are not persistent. It is here that persistence homology derives its namesake, as only connected components with persistence greater than τ will be returned.

For ToMATo (**Figure 1.20**), PH is used to construct a gradient field and identify basins of attraction, which may signify the presence of a connected component or topological feature within the data[224]. The gradient field is derived from a density field, which may be interpreted as the local density of a point, as determined by kernel density estimation or simply counting the number of neighbours within a defined radius[222]. An ordering of points by density produces what is known as a filtration scheme[225].

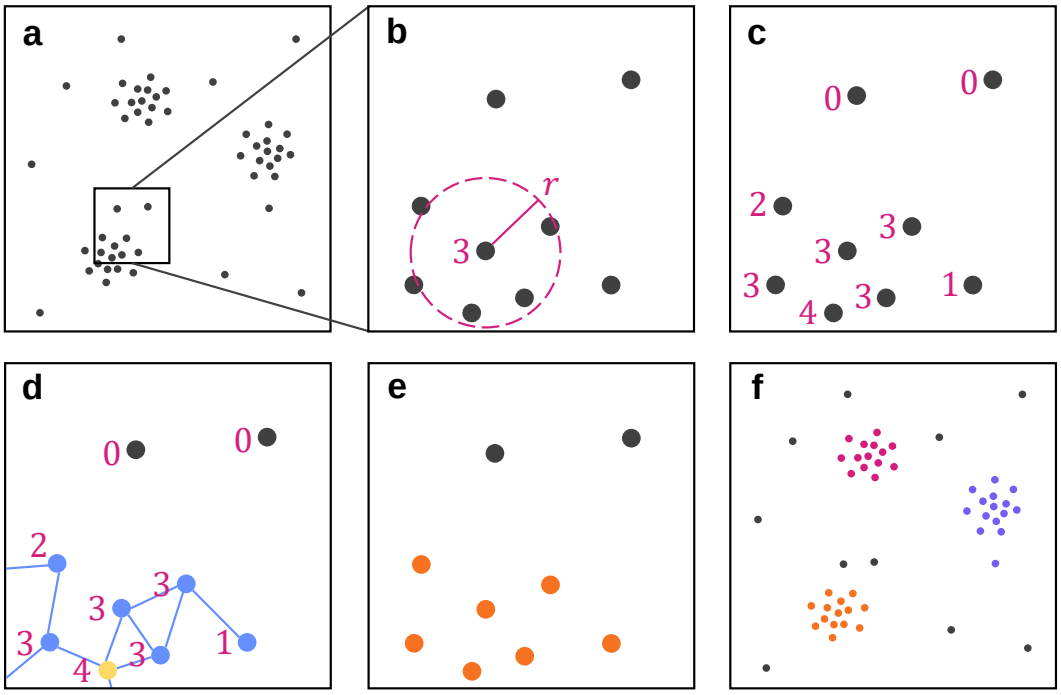


Figure 1.20: A schematic of ToMATo clustering. **a** An example data set with Gaussian clusters. **b** Zoom in of the square region in **a**. **c** The initial steps of ToMATo are primarily the same as in DBSCAN, except for a slight change of notation, in which the search radius is denoted as r. **d** The point with highest density is taken to be a root (yellow) and points are sequentially connected to any neighbours whose root density is at most τ different to their own, where τ is the persistence threshold. In this instance, the persistence threshold is taken to be 3. **e-f** Any clusters found with root density less that τ are filtered out and designated as outliers.

Connected components are constructed by iteratively analysing each filtration entry in sequence to determine local maxima, while simultaneously attaching points to their neighbours[226]. In each connected component, the point of greatest density is denoted the root or mode and any point adjacent to the component may be attached provided its density is no less than τ different from the mode[227]. Two adjacent connected components may be aggregated into one provided the difference between their root densities is no greater than τ. This process is known as mode-seeking and is the

principle by which each connected component is identified[227]. At the conclusion of the algorithm, any object with root density less than τ is also filtered out to remove background[227].

Fundamentally, the construction of simplices on data points induces a graph or network. Topological descriptors are numeric quantities which describe structural properties of these graphs. This includes the Betti numbers, which count the number of topological features (e.g. connected components or holes) at any dimension, and the Euler characteristic, a topological invariant which summarises the simplicial complex[219]. These statistics have been used to build machine learning pipelines for feature extraction in FM data[228]. Mapper is an alternative to persistent homology, which creates a simplified representation of high-dimensional data by clustering data points and mapping these clusters into a graph structure[229]. First, the data is transformed to a lower-dimensional space by methods of dimensionality reduction (typically principal component analysis). Then, the data space is covered by a set of overlapping intervals, which divides the space into bins. The choice of these intervals is user-defined and may be selected for equal size or equal point density. Within each bin, a cluster analysis algorithm is used to partition data points – the choice of this algorithm depends on user preference. The clusters are mapped to nodes in a graph, and edges are formed between nodes if they share any data points across the assigned intervals. In culmination, this reduces high-dimensional data into a lower-dimensional representation that preserves its topological features, which allows for some visualisation of data shape and structure. Mapper has been used to analyse neuroimaging data[230], genetic data in breast-cancer patients[231] and biomolecular folding pathways[232]. TDA methods are inherently scalable and can handle large, high-dimensional data sets[219]. This is advantageous given that the output of super-

resolution microscopy is often characterised as big data[128]. Furthermore, TDA does not require a priori knowledge of the data structure, and does not impose strict assumptions on data geometry[230]. PH in particular is robust to noise, making it suitable for data that is noisy or incomplete, which may arise from imaging artefacts[224]. Furthermore, PH is relatively stable under small perturbations to the persistence threshold, so minor variations will not generally alter the discrete properties of features identified[224].

1.3.8 Introduction to marked point patterns

Marked point patterns (MPPs), also known as marked point processes, are an extension to traditional point clouds in which each localisation is marked with additional values or labels[233]. Each point in an MPP consists of an an m-tuple in which one or more dimensions correspond to spatial coordinates, and all other dimensions represent non-spatial data (**Figure 1.21**). Such patterns can be either categorical, in which each mark arises from a finite set, or continuous, in which marks adopt scalar numerical values. In SMLM, categorical MPPs may arise from overlaying localisation maps of several molecular species, where the mark corresponds to the molecule type[207]. Continuous MPPs may be sourced through the use of environmentally-sensitive dyes, in which the mark quantifies a property of the probe's environment[42]. Under polarity-sensitive probes, each localisation is assigned a quantitative label in the form of the GP value, which represents the degree of membrane order[42].

1.3.9 Existing methods in MPP analysis

For categorical MPPs, colocalisation analysis refers to any analysis of the spatial proximity between different points of distinct categories[234]. Co-clustering, on the other hand, is the phenomenon by which molecules from distinct categories form spatial

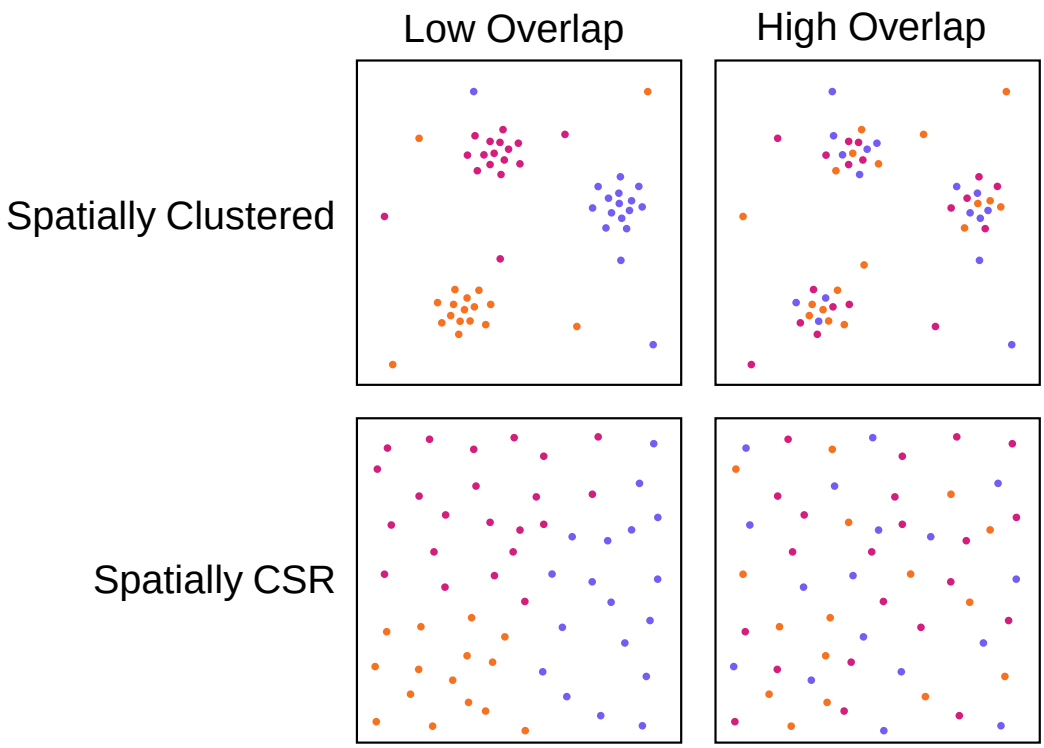


Figure 1.21: Examples of marked point patterns. Spatial organisation may vary between MPPs, with some displaying distinct spatial separation in clusters, and others presenting as CSR. Points with similar marks may be spatially well-separated, or may display a high degree of overlap.

clusters together[234]. Quantification of colocalisation and co-clustering is the primary strategy for determining the potential interactions of dynamic protein complexes in biological samples[235]. Dual-colour FM can be used to detect colocalisation of two labelled molecules[236]. In image data, this can be quantified by the degree of overlap in separate colour channels (i.e. by Mander's coefficient) or by correlation of pixel intensities (i.e. by Pearson's correlation coefficient)[166,237]. However, colocalisation analysis conducted on conventional fluorescence microscopy images may suffer as a result of chromatic errors and a limited spatial resolution[238]. In point pattern data from SMLM, two molecules will never occupy the precise same position, and therefore colocalisation is instead defined as a measure of intermolecular distance or spatial

association[239]. The simplest approach to determining point pattern colocalisation is to use a nearest-neighbour distance analysis[237]. However, this approach may yield false positives in the case of one species being randomly distributed[237]. The coordinate-based colocalisation (CBC) value is based on the proportion of points in category B within the radius of a point in category A compared to the total count of category B in a pre-defined maximum radius[239]. This is normalised over area and used to calculate the Spearman's rank coefficient, to which the colocalisation index is proportional[239]. This method is, however, highly sensitive to the strategy of labelling, especially when the same structure is labelled by two different techniques that may interfere with each other. State-of-the-art analysis methods, such as SODA, Clus-DoC and LAMA, probe for colocalisation across a range of spatial scales[234,240,241].

While traditional cluster analysis methods can be used for probing the spatial organisation of an MPP, they do not typically take account of the mark each point has been assigned[242]. As a result, these techniques cannot be used to determine whether proteins of the same value tend to colocalise or co-cluster, and therefore do not offer any information about the existence of underlying domains[243]. Some second-order characteristics have been adapted for investigating correlations between the locations and labels of MPPs. These represent functions of the distance between two points of the pattern, rather than the point pattern itself. Such characteristics include Isham's mark correlation function and Stoyan's k_{mm} function (normed), as well as the mark variogram and the mark covariance function (unnormed)[233].

1.4 Outstanding questions in the field

With the advent of super-resolution microscopy, among other probing techniques, more information about the composition and organisation of the plasma membrane is

becoming available. There is increasing evidence surmounting to the use of lipid domains not just as structural components of the membrane, but as functional molecular rafts. However, many of the mechanistic principles guiding membrane organisation and behaviour are unknown. Tracking the exact dynamic processes of individual molecules is not generally trivial, and requires imaging over several frames taken at sufficiently short intervals of time. Therefore, protein and lipid diffusion can only be measured under microscopy techniques with high spatial and temporal resolution. Current research efforts continue to focus on interpreting dynamic and marked point pattern data. Outstanding questions regarding experiments into lipid organisation revolve around determining whether lipid rafts are present, how best to identify them, and what the link between their organisation and cell function may be. Further, there is uncertainty about how lipid domains may vary within cells, between cells, and amongst organelles. Understanding how membrane compartmentalisation impacts cellular processes requires thorough, reproducible analysis at the nano-, micro- and meso-scale. Depending on the imaging target and modality, microscopy data can vary in architecture and present features with irregular or unpredictable geometries, which can be challenging to quantify. Traditional geometric analyses may fail to capture the structure of the data, but TDA could provide a framework to study these complexities and perform feature extraction effectively. However, TDA has not seen widespread use in analysing super-resolution data, largely owing to difficulty in its interpretation, especially for biologists or microscopists who may not be familiar with topological concepts. Here, we aim to generalise TDA towards this biological context.

1.5 Thesis objectives

In this work, we present a series of analytic tools tailored to multifaceted microscopy data. Each methodology is designed to provide specific insights into the biophysical properties of a distinct data type, allowing for a comprehensive analysis of cellular membranes at varying resolutions and scales. The first of these methods is known as Topological Boundary Line Estimation using Recurrence of Neighbouring Emissions (TOBLERONE), an image analysis tool which allows for the segmentation of 2D images, 3D z-stacks, and spatiotemporal t-stacks. In **Chapter 2**, this tool is applied to identifying cell membranes and organelles in confocal microscopy data, which may aid understanding of cellular architecture and membrane properties. Then, magnifying to the nanoscale, we introduce a protein aggregation dynamics simulator which learns from spatially-descriptive statistics taken directly from SMLM data. The model serves a dual purpose: to simulate static protein maps which capture the topological properties of point patterns derived from experimental data, and to predict protein aggregate behaviour in dynamic systems. This approach, known as Agent-based Spatiotemporal Molecular Distributions Evolving Under Simulation (ASMODEUS), is explored in **Chapter 3**. Finally, in **Chapter 4**, we introduce a software package for interpreting marked point pattern data. In this context, we use this package to probe for evidence of lipid nanodomains in MPPs arising from ratiometric DNA-PAINT. The package, Point Label Analysis for Super-resolved Marked Attributes (PLASMA), contains two algorithms for processing SMLM data types, including point patterns with non-spatial dimensions. Most notably, we introduce a topological cluster analysis technique for partitioning of marked point patterns, which is, to our knowledge, the first of its kind.

We show that the analytic tools introduced in this work are capable of distinguishing the biophysical properties associated with lipid composition and order in these data types. Furthermore, we find evidence of regions of homogeneous membrane order at nanoscale resolutions. In summary, these methods, combined with multimodal fluorescence microscopy and SMLM, may yield a promising avenue for mapping membrane biophysical properties at nano- and micro-scale resolutions. This broad resolution spectrum facilitates a more comprehensive understanding of cellular environments, which may yield new discoveries in cell biology and further study into the micro- and nano-environment of the plasma membrane.

Chapter 2: Topological image analysis

The focus of this chapter is on the implementation of persistent homology for segmenting visible structures in conventional fluorescence microscopy data. A thorough review of associated literature suggests that this is a novel application of topological image analysis and, in particular, depicts the first application of persistent homology to processing fluorescence microscopy data.

Contributions

Contributing authors

L. Panconi[1,2,3], M. Makarova[4,5], R. C. May[4], A. J. Collins[6] & D.M. Owen[1,3,7].

Affiliations

[1]Institute of Immunology and Immunotherapy, University of Birmingham, Birmingham, UK.

[2]College of Engineering and Physical Sciences, University of Birmingham, Birmingham, UK.

[3]Centre of Membrane Proteins and Receptors (COMPARE), University of Birmingham, Birmingham, UK.

[4]School of Biosciences, College of Life and Environmental Science, University of Birmingham, Birmingham, UK.

[5]Institute of Metabolism and Systems Research, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK.

[6]Department of Chemistry, University of Cambridge, Cambridge, UK.

[7]School of Mathematics, College of Engineering and Physical Sciences, University of Birmingham, Birmingham, UK.

Author contributions

L. P. conceptualised the project and methodology, wrote simulation and TOBLERONE code, performed simulations and data analysis. *C. gattii* cell culture was performed by R. C. M. HEK293 cell culture and *S. pombe* cell culture were undertaken by M. M. *C. gattii*, HEK293 and *S. Pombe* confocal microscopy data were acquired by M. M. (**Figures 2.8, 2.9, 2.10** and **2.15**). Jurkat T cell culture was performed by A. J. C. Jurkat T cell confocal microscopy data was acquired by A. J. C. (**Figure 2.13**). D. M. O. provided project supervision.

2.1 Introduction

Advances in fluorescence microscopy have allowed for highly-resolved planar, volumetric and spatiotemporal imaging, producing 2D images, 3D scans and live-cell videos respectively[134,144,244,245]. Image segmentation algorithms can be used to separate fluorescing objects from the background and each other[133-135]. Without such techniques, researchers must undertake time-consuming manual segmentation and quantification, which can yield subjective results and impact reproducibility[246-248].

State-of-the-art segmentation algorithms depend on supervised machine or deep learning approaches, which learn to interpret patterns in training data to make informed predictions on new images[249]. Such methods include CellSeg, StarDist and CellPose, which carry the advantage that they are highly adaptable[249-251]. With the advent of cloud-based models such as Segment Anything[252], semi-automatic 2D segmentation can be achieved with minimal user input – however, this approach requires manual annotation of each object in each image (by clicking on them in a user interface) and is not yet suitable for automated image analysis. As supervised learning

approaches, these techniques require annotated image repositories for training, which must be labelled manually and can be subjective. Furthermore, there is no generalised method of extending 2D machine learning algorithms to incorporate variations in geometry or higher dimensional data sets without completely retraining new algorithms[253,254]. Developments in convolutional and recurrent neural networks show promise for 3D cellular and biomedical image segmentation, but are yet to achieve the accuracy of 2D methods[255,256]. A fundamental drawback of these machine learning models is that they are black box approaches, which means there is no insight into the algorithm's quality, maintainability or internal structure[253]. Therefore, it is impossible to generalise performance quality to unseen data[253].

Classical image segmentation methods are algorithms which are not based on supervised machine learning[257]. A range of these algorithms exist for 2D, 3D and spatiotemporal cell segmentation[257]. Such models typically rely on background-foreground separation techniques, such as Otsu thresholding[258]. Provided the foreground can be isolated, segmentation must then be undertaken by a separate region-based technique, such as seed-point extraction or the Watershed algorithm[259,260]. The success of these methods, in fluorescence microscopy, usually depends on cell geometry and image SNR[261]. Further, while parameter estimation is possible for some classical methods, it is not often built in[262,263]. 3D Simple Segmentation performs volume segmentation by binarising z-stacks subject to a given threshold, then performing 3D Watershed to separate objects[264]. 3D Spot Segmentation first identifies object seeds, obtained from local intensity maxima, then applies a local threshold around each seed and clusters voxels with values higher than the local threshold[264]. All-encompassing algorithms, which achieve true segmentation in a single application, probe for specific geometric structures in image data[265]. This

can make them unsuitable for segmentation of cells or organelles with complex morphologies, as is often seen in biological data[265]. Track analysis software, such as Trackmate, uses linking algorithms to detect objects across time series data and build contours which follow each object over time[266]. Trackmate offers several linking algorithms such as nearest-neighbour tracking, which links objects based on the shortest distance to their predicted position in the next frame[266].

Topological analysis techniques have shown promise in decomposing image intensity variations across gradient fields, which is a precursor for successful cell and organelle segmentation[228,267-269]. This chapter introduces an alternative segmentation algorithm built on the principles of persistent homology and intensity mode seeking, denoted Topological Boundary Line Estimation using Recurrence Of Neighbouring Emissions (TOBLERONE). Both 2D and 3D volumetric (3DTOBLERONE) implementations are introduced. Although structurally similar, spatiotemporal (t-stack) data is analysed separately to volumetric (z-stack) data to track dynamic processes such as splitting and merging of components. To permit this, a topological segmentation tool for temporal FM data (tempTOBLERONE) is introduced. Unlike supervised machine learning models, which are trained on labelled data and output labelled data, TOBLERONE is unsupervised, and therefore outputs labelled data without the need for training.

In this work, each variant of TOBLERONE is applied to simulated data and, where applicable, sensitivity and specificity is compared with existing segmentation algorithms under imposed image artefacts. We also quantify the impact of Gaussian blur and noise on algorithmic performance. Further, all algorithms are demonstrated on experimental confocal microscopy data. We demonstrate 2D TOBLERONE on HEK293 cells and map membrane heterogeneity in images of *Cryptococcus gattii*,

stained with polarity-sensitive membrane dye di-4-ANEPPDHQ. We apply 3DTOBLERONE in practice by segmenting scans of Jurkat T-cells and find that segmentation statistics regarding cell geometry agree with existing literature. Finally, we demonstrate tempTOBLERONE on dynamic, live-cell data of *Schizosaccharomyces pombe* undergoing nuclear division and GFP-GOWT1 mouse stem cell data (taken from the Cell Tracking Challenge[270]).

2.2 Materials and methods

HEK293 cell culture

HEK293 cells were cultured with Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) at 37°C in a 5% CO_2 incubator. Cells were passaged and seeded onto Ibidi µ-slide 8-well glass-bottomed chambers 24h prior to imaging. Cells were stained with either 5mM di-4-ANEPPDHQ from a 5mM ethanol stock solution, for cell plasma membranes, or 1X NucBlue, for DNA (as a proxy for nuclei), 30 minutes before imaging. All media and dye were sourced from Thermo Fisher Life Technologies, Paisley, UK.

Cryptococcus gattii cell culture

C. gattii cells were cultured with yeast peptone dextrose (YPD) broth at 25°C under rotation and stained with di-4-ANEPPDHQ as in HEK293 cell line. Samples were treated with either 100µM 2-hydroxyoleic acid (2OHOA) or 20µM 7-ketocholesterol 3 hours before imaging and staining. All media and dye were sourced from Life Technologies.

Jurkat T cell culture

Jurkat E6.1 T cells were cultured in R10 medium (Roswell Park Memorial Institute [RPMI] 1640 medium supplemented with 10% FBS, 1mM penicillin and streptomycin [PenStrep], 2mM L-glutamine and 1mM sodium pyruvate [all from Sigma-Aldrich, Madison, WI]) at 37°C in a 5% CO_2 incubator. 1mL of cells were resuspended in Nuclear Mask Deep Red solution (1 in 200 dilution from stock, 250X concentrate in dimethyl sulfoxide [DMSO]), for DNA, and WGA-AlexaFluor 488 (10 µg/mL concentration) for cell plasma membranes (both from Life Technologies). Cells were incubated for 15 minutes at 37°C and washed three times in phosphate buffer saline (PBS), then resuspended in 4% paraformaldehyde (PFA) for fixation, incubating for a further 15 minutes at 37°C, then washing again three times in PBS (PFA and PBS supplied by Life Technologies).

Schizosaccaramyces pombe cell culture

S. pombe was grown in yeast extract with supplements (YES) medium. Media preparation and genetic techniques follow protocols given in literature[271]. An *S. pombe* strain expressing BOP1-mCherry (ribosome biogenesis protein 1) was produced through genetic transformation of the wild type strain under homologous recombination. Plasmids containing the mCherry gene and 3'UTR fragment of BOP1 gene (SPAP32A8.03c) were constructed via standard molecular biology methods given in literature[271]. Endogenous BOP1 gene was tagged with mCherry and maintained all native regulatory elements. Transformation was performed using lithium acetate-based method described in literature[271].

Static confocal microscopy

Live HEK293 and *C. gattii* cells were imaged with a Zeiss LSM 780 laser-scanning confocal microscope at 37°C, while Jurkat T cells were imaged with a Zeiss LSM 900 confocal in confocal scan mode at 37°C. Excitation and emission channels were defined as in **Table 2.1** for each dye. 4× line averaging was used across all acquisitions. For *C. gattii* cells imaged with di-4-ANEPPDHQ, GP images were constructed by calculating GP values for each pixel, as in literature (see **Appendix**)[108].

Dye Name	**Excitation Wavelength**	**Emission Range**
di-4-ANEPPDHQ	488nm	500-580nm (green), 620-750nm (red)
NucBlue	405nm	420-500nm
Deep Red	638nm	650-750nm
AlexaFluor 488	488nm	500-600nm

Table 2.1: Excitation wavelength and fluorescence channels for each dye.

S. pombe dynamic imaging

S. pombe cells were placed in sealed growth chambers containing 2% agarose YES medium. Imaging was undertaken using a Zeiss Axiovert 200M microscope with a Plan Apochromat 100× 1.4 NA objective. The microscope comprised an UltraView RS-3 confocal system, with a CSU21 confocal optical scanner, 12-bit digital cooled Hamamatsu Orca-ER camera, and krypton-argon triple line laser illumination source. Image z-stacks with seven sections and frame depth of 0.5μm were acquired at intervals of 1 minute. Then, t-stacks were obtained by undertaking z-stack maximum projection using FIJI/ImageJ (v. 2.9.0)[272].

Simulating images and evaluating algorithmic performance for 2DTOBLERONE

To produce simulated 8-bit images which better represented real cell morphologies, all experimental data, including *C. gattii*, *S. pombe*, Jurkat T cells and HEK293 images, were put through CellPose v3.0.10 to extract cell masks[273]. In total, 2167 masks were identified and each cell mask was converted into a separate binary image. For each simulation, an empty 256 by 256 image matrix was initialised (with all pixel intensities set to 0). Then, for each image matrix, the following process was repeated until 5 failures:

1. Randomly select a cell mask image from the available masks.
2. Add the mask image to the image matrix at a randomly-selected location.
3. Check if the max pixel intensity of the image matrix is greater than 2 (i.e. direct overlap of two cells). If so, record a failure and reject mask placement, otherwise continue.

This entire process was repeated to give 100 simulated images. Then, each image was multiplied by a mask pixel intensity, with intensities ranging between 0.1 and 0.9 in increments of 0.1. Gaussian blur (with standard deviation σ_1 ranging between 0.2 and 2 in increments of 0.2) was applied in base R (v. 4.2.3) and Gaussian noise (with standard deviation σ_2 ranging between 3 and 30 in increments of 3) was applied using the blur function of the EBImage R package (v. 4.19.13)[274]. Gaussian distributions were chosen to reflect the properties of noise and blur which may arise from fluorescence microscopy[275]. This gave a total of 900 data quality conditions, which were considered for each image, giving 90000 simulations in total. 2DTOBLERONE was performed on each simulation with the mask pixel intensity as persistence threshold. Sensitivity and specificity (see **Appendix** for formulas) were calculated for each simulation within an 11 by 11 pixel grid around mask boundaries. The relative

difference between the number of components returned and the number of true cell masks in the image was recorded for each image.

Simulating images and evaluating algorithmic performance for 3DTOBLERONE

For each 8-bit simulation, an empty 256 by 256 by 50 voxel array was initialised (with all voxel intensities set to 0). Then, for each array, the following process was repeated until 3 failures:

1. Randomly select a 3D cell mask image from the available masks (as above).
2. Add the mask image to the array at a randomly-selected location.
3. Check if the max voxel intensity of the image matrix is greater than 2 (i.e. direct overlap of two cells). If so, record a failure and reject mask placement, otherwise continue.

This entire process was repeated to give 100 simulated z-stacks. Each stack was multiplied by a mask voxel intensity ranging between 0.4 and 0.9 in increments of 0.1. Gaussian blur (with standard deviation σ_1 ranging between 0.4 and 2 in increments of 0.4) was applied in base R (v. 4.2.3) and Gaussian noise (with standard deviation σ_2 ranging between 6 and 30 in increments of 6) was applied using the blur function of the EBImage R package (v. 4.19.13)[274]. This gave a total of 150 data quality conditions, which were considered for each z-stack, giving 15000 simulations in total. 3DTOBLERONE was performed on each simulation with the mask voxel intensity as persistence threshold. Sensitivity and specificity (see **Appendix** for formulas) were calculated for each simulation within an 11 by 11 by 11 voxel grid around mask boundaries. The relative difference between the number of components returned and the number of true cell masks in the image was recorded for each stack.

Simulating images and evaluating algorithmic performance for tempTOBLERONE

First, 2D images were generated as for 2DTOBLERONE above. Then, for each image, 19 additional time frames were considered, with each frame determined by comparing to the image in the previous frame. For each cell mask in the image, the following process was repeated until 5 failures:

1. Randomly select an angle θ between 0 and 2π and a step size s between 0 and 10 pixels.
2. Move the cell mask to the pixels at the end of the path of length s in the direction of angle θ (taking ceiling of values for precise integers).
3. Check if the max pixel intensity of the image is greater than 2 (overlap). If so, reject the move and record failure, otherwise proceed to next mask.

This entire process was repeated to give 100 simulated t-stacks. Mask intensities, Gaussian blur and Gaussian noise were simulated and imposed as for 3DTOBLERONE above, giving 15000 simulations in total. tempTOBLERONE was performed on each simulation with the mask pixel intensity as persistence threshold. Sensitivity and specificity (see **Appendix** for formulas) were calculated for each simulation within an 11 by 11 by 3 pixel grid around mask boundaries. The relative difference between the number of components returned and the number of true cell masks in the image was recorded for each stack.

Simulated images for failure cases

Two *C. gattii* cells were manually segmented from FM data and overlaid onto an image with no background. Three additional copies of this image were made. The first image was overlaid with Gaussian noise with standard deviation of 50. In the second image, the distance between cells was reduced such that their mask boundaries overlapped.

In the third image, the background signal was increased to reduce image contrast. This was achieved by first multiplying all pixel intensities by 0.5, then overlaying Gaussian noise with standard deviation of 10 and Gaussian blur with standard deviation of 0.5. Finally, all pixel intensities were mapped linearly from a 0-1 intensity range to a 0.25-0.75 intensity range. Images were 256 by 256 pixels in size. All image processing was undertaken in in FIJI/ImageJ (v. 2.9.0).

Persistent homology in image analysis

The raw output data of fluorescence microscopy provides an image or stack, represented as a numeric matrix or array. TOBLERONE takes such an array as input, alongside a persistence threshold τ, which determines the algorithm's sensitivity to overlapping components. A density field is interpreted directly as the density of photons emitted by fluorescent probes across the ROI, represented in the grayscale image as the intensity of each pixel. A filtration scheme is constructed by ordering the intensity of each pixel across each frame. In each connected component, the pixel or voxel of greatest intensity is denoted the root or mode and any pixel adjacent to the component may be attached provided its intensity is no less than τ different from the mode[227]. This implementation of persistent homology is analogous to ordering a sequence of binarised images by thresholding over all possible intensities (**Figure 2.1a-c**). This is achieved by mapping each active pixel onto a node and assigning an edge to each pair of adjacent nodes – this representation is referred to as the grid topology (**Figure 2.1d**). The primary assumption of TOBLERONE is that a connected component is defined by a set of pixels in which a path between any two pixels can be achieved by a finite series of lateral or diagonal movements across nodes within the grid topology. This is equivalent to connecting adjacent nodes within a 3 by 3 grid, for

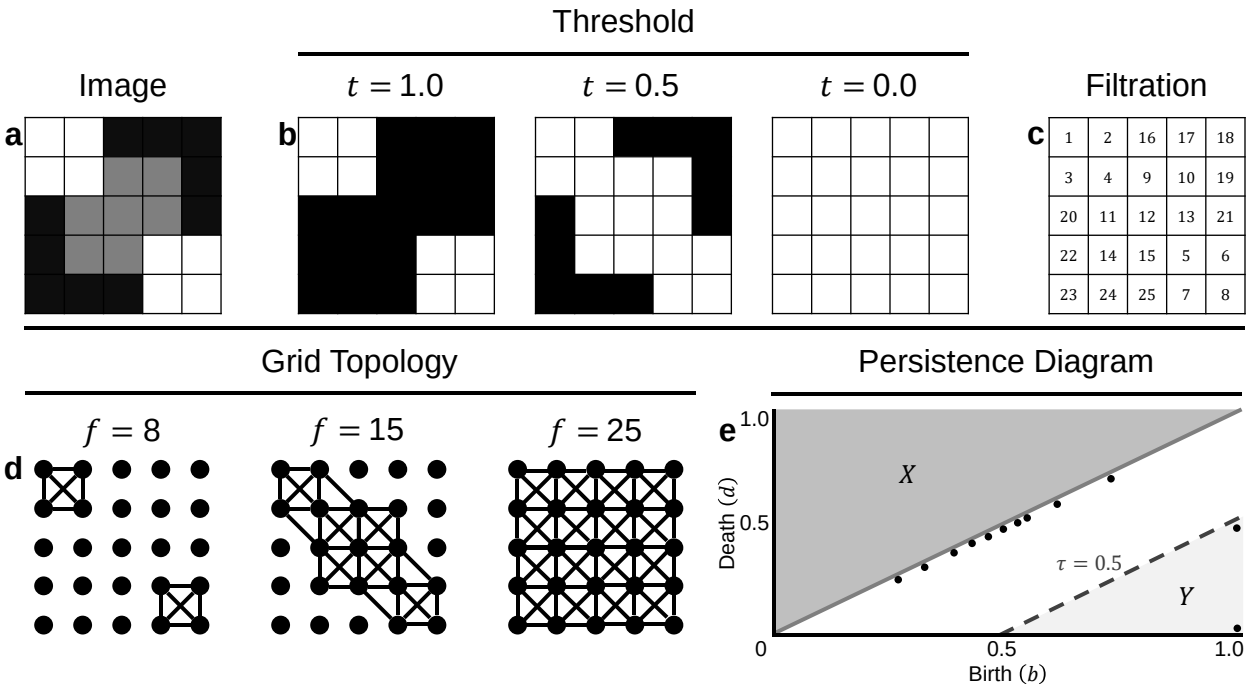


Figure 2.1: Schematic of TOBLERONE. **a** An example image with 25 pixels, represented by a matrix in which each entry contains a numeric value between 0 and 1, denoting fluorescence intensity. **b** Binarised image at thresholds of *t* = 1.0, 0.5 and 0.0, respectively. **c** The filtration matrix constructed by assigning each pixel a value corresponding to the order at which it was activated. Pixels which became active at the same threshold are numbered arbitrarily. **d** The network representation of the image at filtration values corresponding to the thresholds above. Here, *f* denotes the maximum filtration value permitted. As the filtration value increases, two connected components form and then merge into one component. **e** The persistence diagram constructed from the topological decomposition of the image. This plots the birth threshold, b, against the death threshold, d, for each component identified. The persistence of each component is represented by $\tau = b - d$. Two prominent points are found in region Y, given by $\tau \geq 0.5$ which correspond to the two bright objects in the original image.

2DTOBLERONE, and a 3 by 3 by 3 grid, for 3DTOBLERONE (**Figure 2.2**). The output of TOBLERONE is a list of distinct connected components, represented by lists of

pixels. See **Figure 7.1** (**Appendix**) for conceptual diagram of 2D and 3D TOBLERONE. See **Sections 7.2.1 and 7.2.2** for algorithm pseudocode.

Parameter selection

An appropriate persistence threshold can be estimated by initialising at $\tau = 0.5$ and perturbing the threshold until the returned number of connected components matches the anticipated number. Increasing the threshold results in greater merging and reduces the number of connected components found, while decreasing the threshold reduces the penalty on segmentation and increases the number of components returned (**Figure 2.3**). For data derived under the same imaging conditions, an estimate for the persistence threshold can be derived from a representative image and then applied to all images in the data set.

Persistent homology for 2D image + time data

Since 3D TOBLERONE identifies separate connected components in 3D space, all voxels connected to the initial root will be aggregated into one component. This would be inappropriate for live-cell video data, as the algorithm would be unable to track splitting (e.g. by mitotic division) or merging (e.g. by vesicle fusion) of objects which were otherwise distinct. Instead, tempTOBLERONE processes each frame with 2DTOBLERONE, then collates components across frames. Functionally, this determines each temporally-distinct connected component and the corresponding spatial roots for each frame. Initialising at the first frame, the algorithm iterates over each pixel in each component and compares with pixels occupying the same positions in the following frame (**Figure 2.4**). If the number of unique roots is unchanged, the

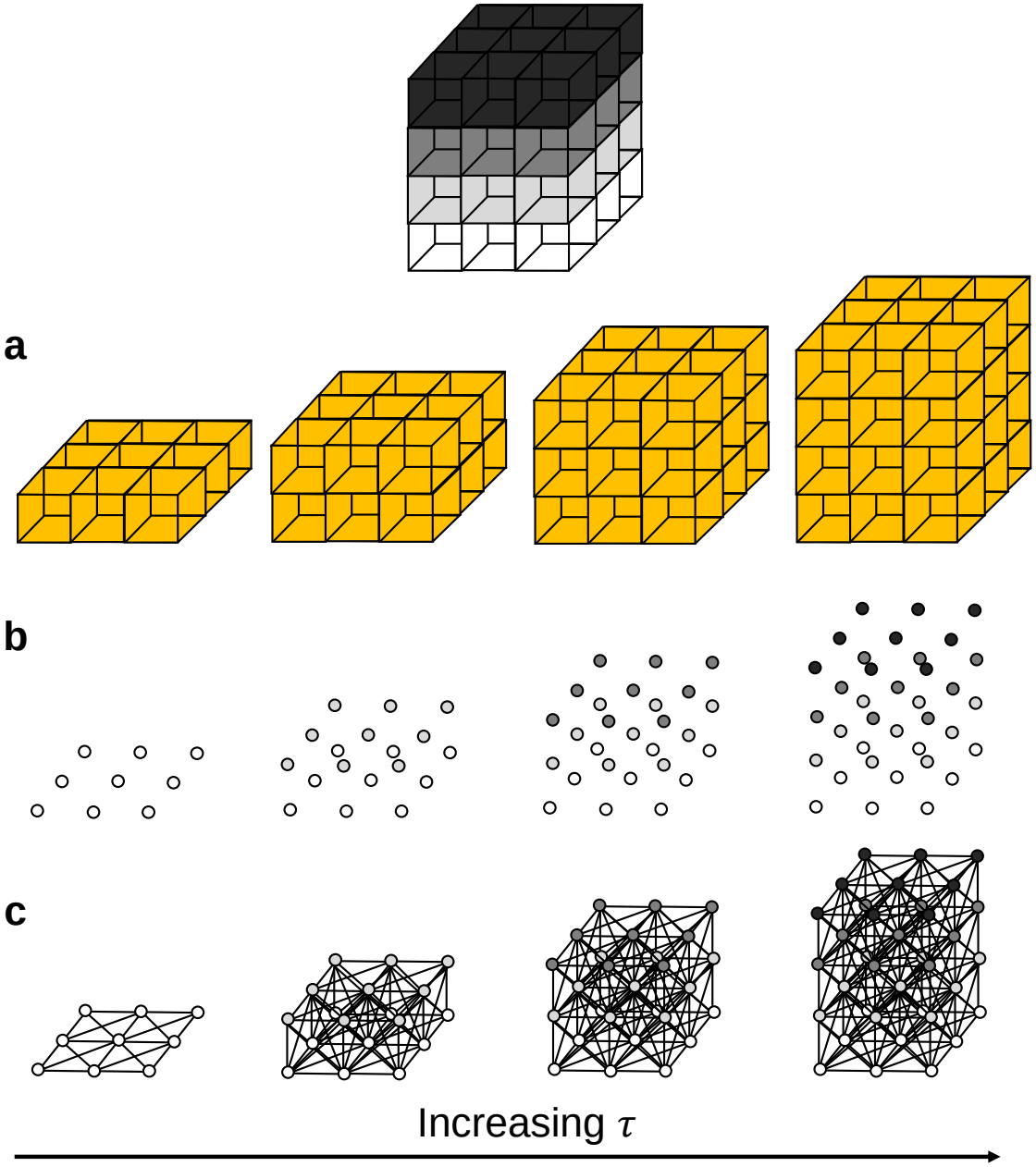


Figure 2.2: Topological representations of a z-stack. **a** Stacks may be binarised by thresholding so that specific voxels become activated. **b** Each voxel represents a point in 4D grayscale colour-space. **c** The network representation establishes connectivity between neighbouring active voxels. Increasing the persistence threshold permits connection of lower intensity voxels.

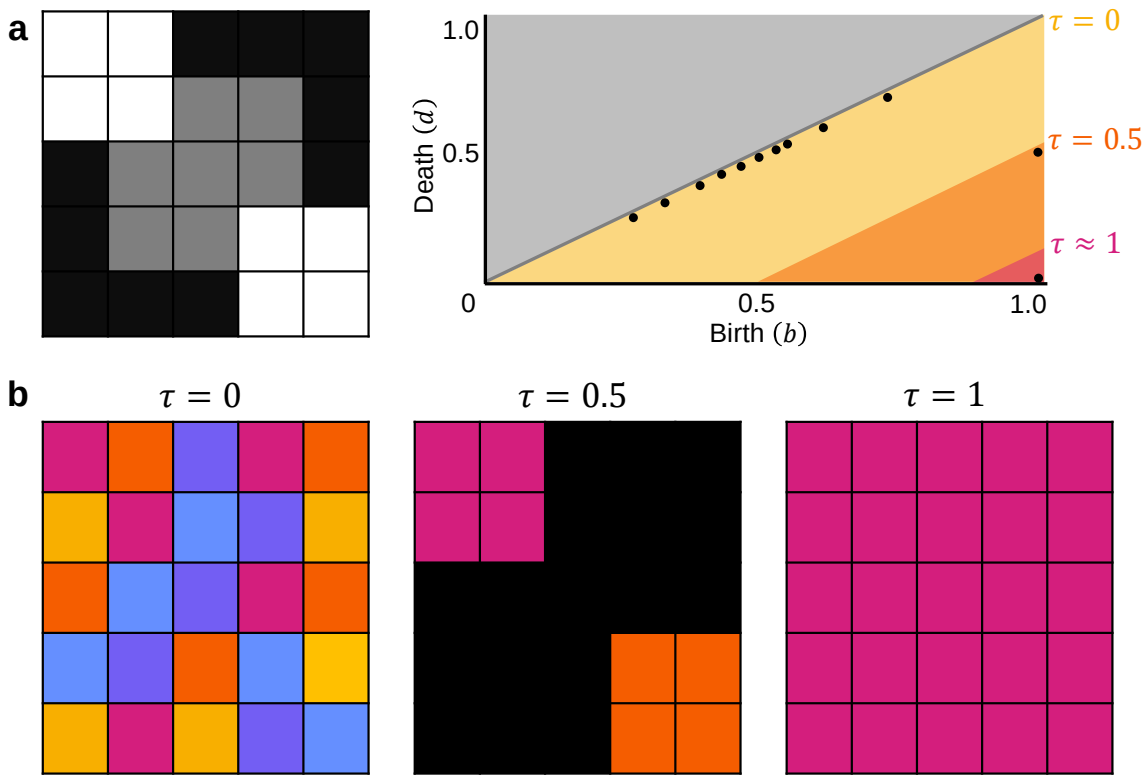


Figure 2.3: The impact of changing persistence on the segmentation identified by TOBLERONE. **a** The persistence defines the granularity of the segmentation. **b** As persistence increases, components undergo increased connectivity which reduces the number of components returned, but increases size.

spatial root of the spatial component is assigned as the new spatiotemporal root of a spatiotemporal component. If two or more unique roots are shared among the same pixels in the following frame, the component whose root has brightest intensity absorbs all other components and overwrites their roots (**Figure 2.5**). This allows for a single component to split into multiple components in a subsequent frame (**Figure 2.5a-b**). Simultaneously, each spatial root is recorded across all frames and if any root is found to be connected to two separate spatiotemporal components, then this implies that the components have merged. As such, all spatiotemporal components which share a single spatial root are aggregated, preserving the component with the brightest

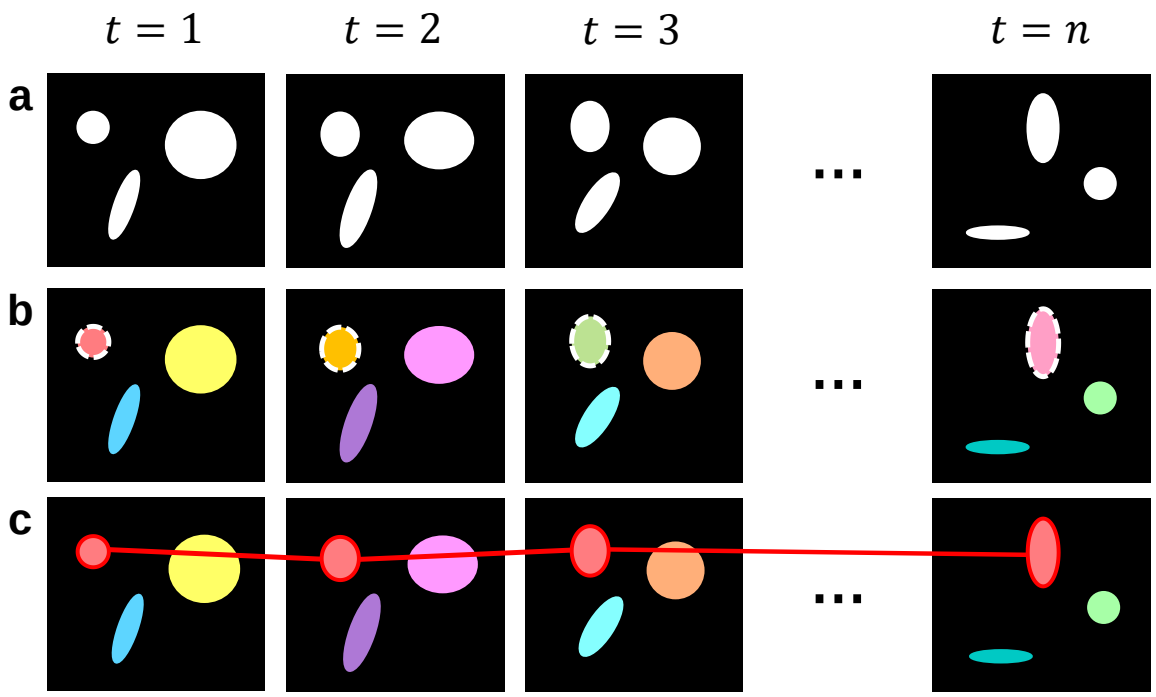


Figure 2.4: Topological decompositions of videos, or t-stacks. **a** Videos are comprised of a series of distinct frames. **b** 2D segmentation establishes connectivity within frames, but not between frames. **c** Temporal topological segmentation connects components spatially and temporally while tracking addition or deletion. An example spatiotemporal component (pink) is formed from a series of spatial components.

spatiotemporal root (**Figure 2.5c-d**). See **Figure 7.2** (**Appendix**) for conceptual diagram of tempTOBLERONE and **Section 7.2.3** for algorithm pseudocode.

Image and stack segmentation software

The TOBLERONE software package (v. 1.0.3) was written in the R programming language (v. 4.2.3) and employed in the integrated development environment RStudio, (2022.07.1+554). TOBLERONE is available for use under GNU General Public License (v. 3.0). Otsu thresholding and the Watershed algorithm were undertaken using built-in functions with the EBImage R package (v. 4.19.13)[274]. Otsu thresholding is parameter-free and so required no finetuning. Appropriate input parameters were

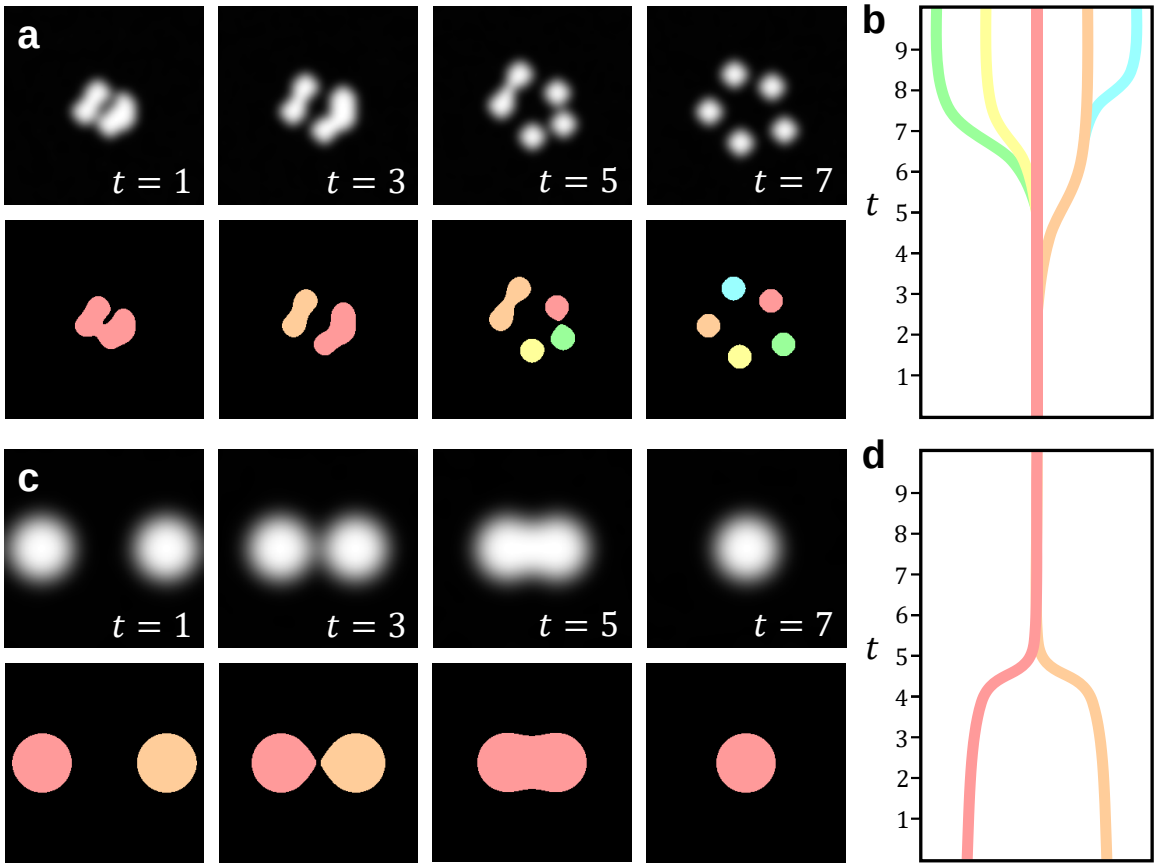


Figure 2.5: Tracking component segmentation over time with tempTOBLERONE. **a** Diagram of one component splitting into five across several time frames. **b** Schematic lineage tree of the spatiotemporal objects given in **a**. **c** Diagram of two components merging into one. **d** Schematic lineage tree of the spatiotemporal objects given in **d**. Exemplar stacks taken from Panconi et al[277].

determined iteratively for the watershed algorithm. 3D Simple Segmentation and 3D Spot Segmentation were undertaken in FIJI/ImageJ (v. 2.9.0) using the 3DSuite plugin (v. 4.0.93)[264]. Suitable input parameters were determined iteratively for both 3D classical algorithms. Nearest-neighbour tracking was applied with Trackmate (v. 7.11.1) in FIJI/ImageJ using pre-defined masks from tempTOBLERONE to determine track lines[266,276]. For *C. gattii* cells imaged with di-4-ANEPPDHQ, mean GP values were taken per-cell and calculated by averaging GP values from all pixels comprising the component boundary, as identified by TOBLERONE.

2.3 Results

2.3.1 2D segmentation of simulated images with TOBLERONE

100 simulated images were generated from cell masks (see **Materials and Methods** for all data simulation methods). Gaussian blur was simulated over each pixel with standard deviation $\sigma_1 = 1.2$, and Gaussian noise with standard deviation $\sigma_2 = 18$. All mask pixel intensities were fixed at 0.5. Each image was segmented under TOBLERONE, alongside two alternative segmentation approaches, Otsu's method and the Watershed algorithm, to compare performance[278,279]. For each algorithm, we calculated segmentation sensitivity and specificity (see **Appendix** for formulas). Here, sensitivity represents the probability of labelling an individual pixel as active, given that the pixel is active in the ground truth image, and specificity represents the probability of labelling an individual pixel as inactive, given that the pixel is inactive in the ground truth image[280]. To avoid skewing statistics, we considered pixels only within an 11 by 11 grid around each pixel comprising the component boundaries (that is, 5 pixels in each direction). Each image took no longer than 5 seconds to segment on a single processor. Under these image conditions, TOBLERONE achieved a mean sensitivity of 0.9412 and a mean specificity of 0.8762. TOBLERONE surpassed the sensitivity of Otsu thresholding (0.8726), but did not surpass the specificity (0.9306). Further, TOBLERONE showed lower sensitivity than the Watershed algorithm (0.9964) but surpassed its specificity (0.4322). In this case, the sensitivity of the watershed algorithm was greater on account of oversaturated segmentation. Algorithm performance on exemplar images is given in **Figure 2.6**.

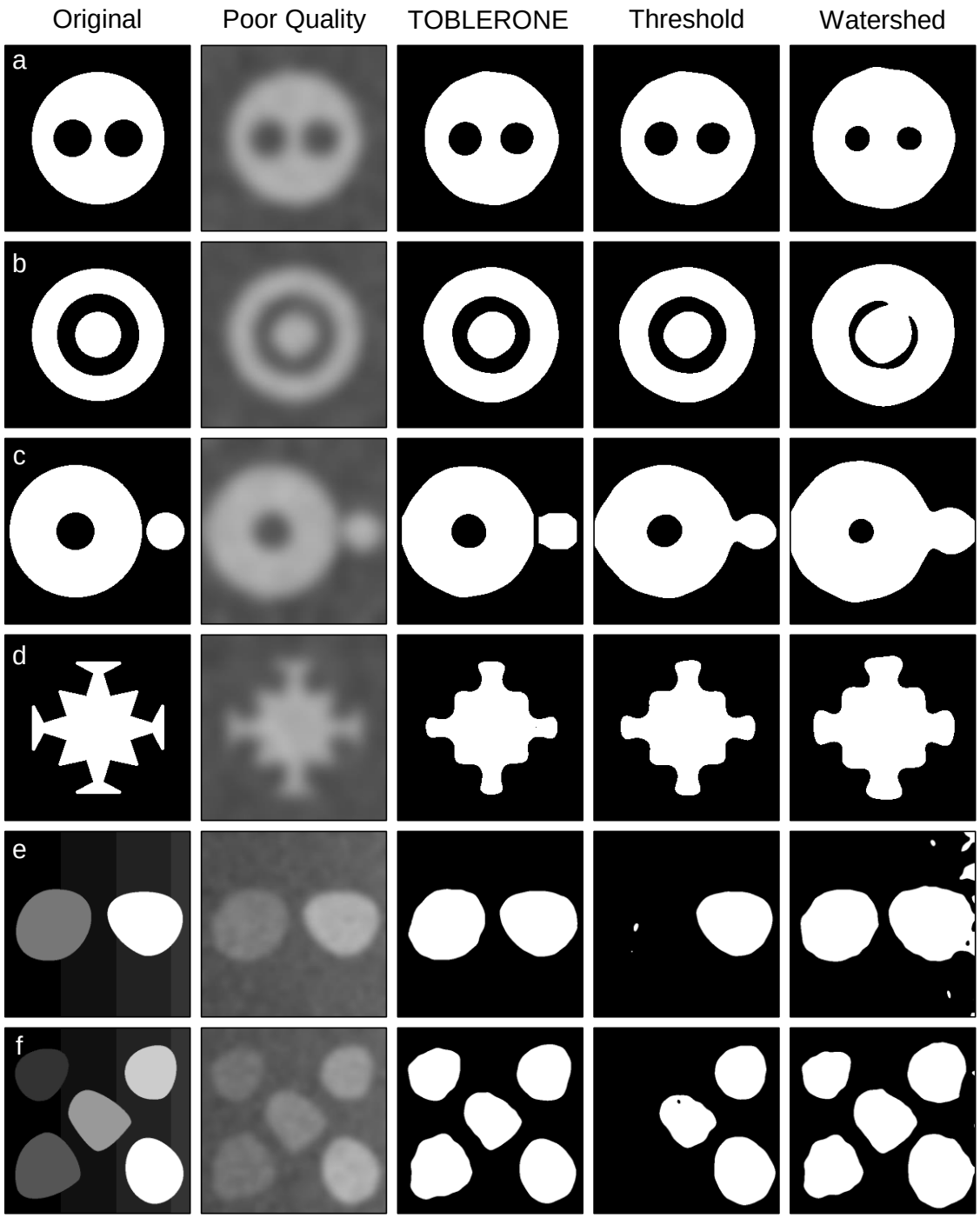


Figure 2.6: Segmentation of exemplar image data for visual representation (taken from Panconi et al[281]). **a-f** A series of toy images exhibiting a range of geometries and topologies. Gaussian noise and blur are simulated over each image with standard deviations $\sigma_1 = 1.2$ and $\sigma_2 = 18$, respectively. TOBLERONE, Otsu thresholding and the Watershed algorithm are performed to recover the original segmentation.

To further explore the impact of image degradation on TOBLERONE's performance, each of the 100 simulated images was overlaid with varying degrees of Gaussian noise and blur (see **Materials and Methods** for quantification). Each image was segmented by TOBLERONE, with sensitivity and specificity calculated as above. Results are given in **Figure 2.7** and suggest that TOBLERONE yields a small, quantifiable decrement in both sensitivity and specificity as either σ_1 or σ_2 are increased. However, segmentation sensitivity shows a greater response to changes in noise, while specificity seems more sensitive to changes in blur.

An overview of common failure cases is given in **Figure 2.8**. This includes toy images with increase shot noise (**Figure 2.8b**), overlapping cell boundaries (**Figure 2.8c**), and low image contrast (**Figure 2.8d**). Noise can yield unwanted activation or deactivation of nodes in the grid topology. This can both disrupt connectivity within components and increase the likelihood of detecting pixels in the background. Increased cell proximity, which is more likely to occur as cell density increases, may cause adjacent cells to be aggregated into one connected component. This may then require additional post-processing (e.g. by Watershed) to get an accurate cell count or boundary line profile. As the image contrast decreases, the difference in intensity between background and cell pixels also decreases. If pixels in the background exceed the intensity of pixels in the cell boundary, they may be connected to the component. Under ideal conditions, the lowest intensity pixel across all cells will be brighter than the highest intensity pixel in the background.

2.3.2 2D segmentation of fluorescence microscopy data with TOBLERONE

TOBLERONE was used to segment two cell types: the R265 strain of *C. gattii*, which are typically spherical, and human embryonic kidney (HEK293) cells, which exhibit a

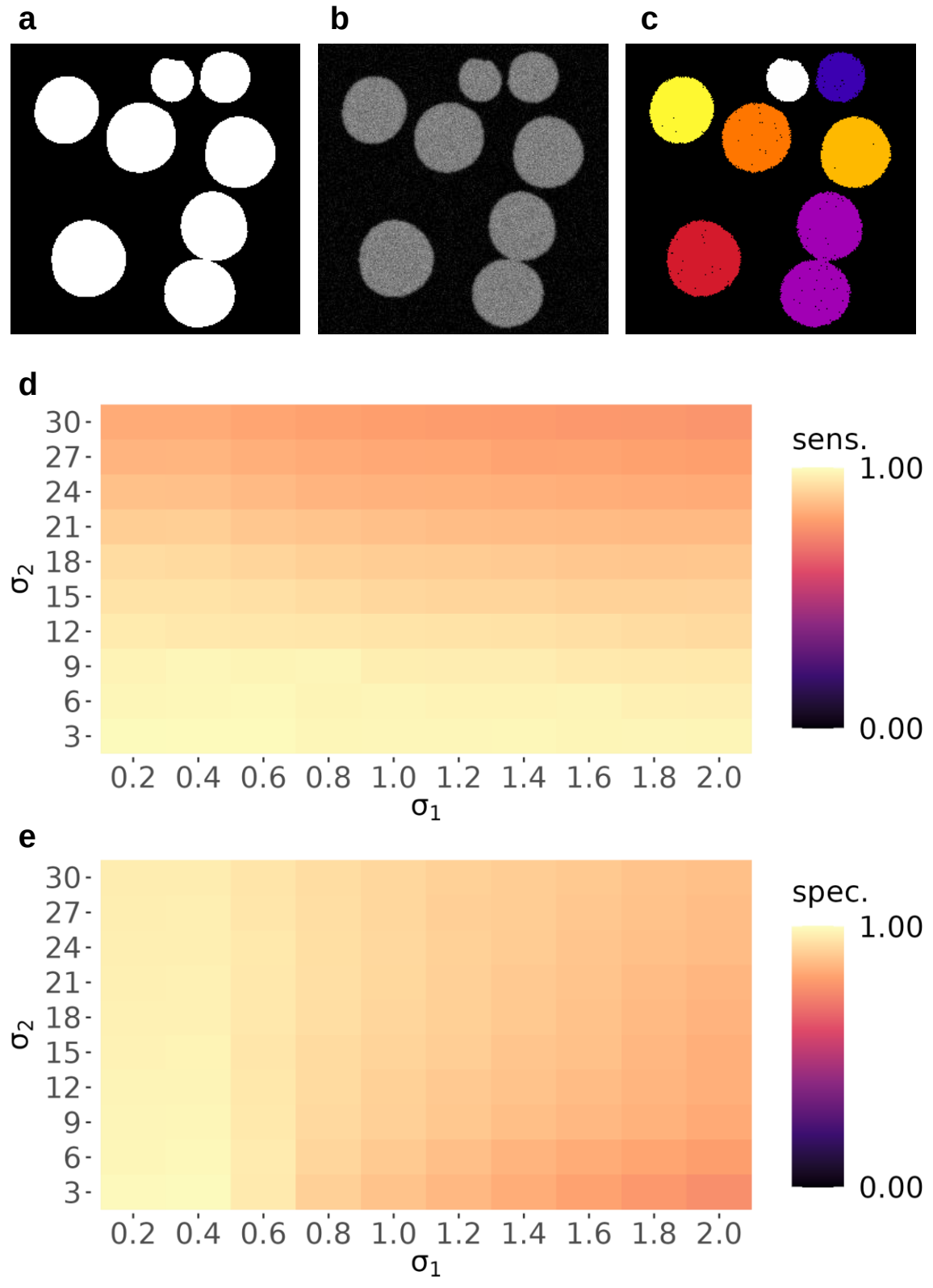


Figure 2.7: Impact of Gaussian blur and noise on sensitivity and specificity of TOBLERONE. **a** An example simulated image constructed from cell masks. **b**

Gaussian noise and blur are simulated over each image (for this case, with standard deviations $\sigma_1 = 1.0$ and $\sigma_2 = 15$). **c** Cell masks are recovered by TOBLERONE and compared to the ground truth to calculate sensitivity and specificity. **d-e** Heatmaps of the mean sensitivity and specificity (as defined above) of TOBLERONE over all simulated images.

more complex geometry, incorporating finger-like protrusions (see **Materials and Methods** for all cell culture and imaging methods). Cryptococcal cells were cultured in YPD media and HEK293 cells in DMEM media. Both were held in coverslip-bottomed microscope dishes, stained with di-4-ANEPPDHQ (for cell plasma membranes) or Nucblue (for DNA in HEK293 cells) for 20 minutes at 37°C and imaged live via a Zeiss LSM 780 laser-scanning confocal microscope. Segmentation was performed with TOBLERONE, Otsu thresholding and the Watershed algorithm. Example cell and organelle masks from HEK293 data are displayed in **Figure 2.9**. For this example, TOBLERONE seemed to successfully identify the non-convex morphology of the HEK293 cell, while also showing discrimination against background. Results suggest that both whole cells (**Figure 2.9a-d**) and organelles (**Figure 2.9e-h**) can be segmented via TOBLERONE.

Further, GP images (**Figure 2.10a-c**) were calculated for each *C. gattii* data set. Line profiles were extracted from the boundaries of the cell masks to approximate the plasma membrane (**Figure 2.10d-f**) and the average GP was determined across all pixels in each line (**Figure 2.10g**). This was repeated for control (untreated) cells, as well as those treated with 2OHOA or 7-ketocholesterol (**Figure 2.10h**). Between the control and both treatment conditions, we determine a statistically significant difference in the average GP value at the 1% significance level, suggesting that both 2OHOA and 7-ketocholesterol reduce membrane order in *C. gattii*.

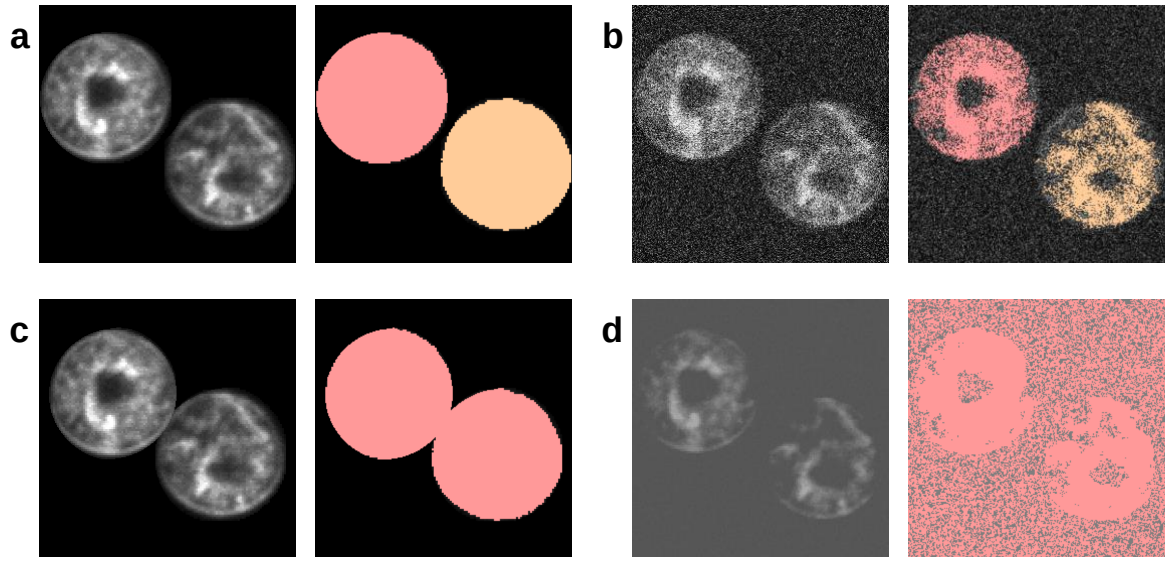


Figure 2.8: Common failure cases for TOBLERONE. In each panel, a toy image is given (left) alongside the output of TOBLERONE (right) with persistence $\tau = 0.9$. Segmented components are given in red and (where applicable) orange. **a** Two *C. gattii* cells, segmented manually and overlaid onto an image with no background. **b** The image in **a**, overlaid with Gaussian noise with standard deviation of 50. **c** The image in **a**, with cells now in closer proximity of each other such that their boundaries overlap. **d** The image in **a**, with lower image contrast and higher background signal. All images are 256 by 256 pixels.

2.3.3 3D segmentation of simulated z-stacks with 3DTOBLERONE

As with planar TOBLERONE, a set of 100 z-stacks was simulated from cell masks (see **Materials and Methods** for all data simulation methods). Spherical Gaussian blur was simulated over each voxel with standard deviation $\sigma_1 = 1.2$, and Gaussian noise with standard deviation $\sigma_2 = 18$. All mask voxel intensities were fixed at 0.5. Each stack was segmented using 3D TOBLERONE, as well as 3D Simple Segmentation and 3D Spot Segmentation[264]. An exemplar segmentation for each algorithm is given in **Figure 2.11**. Sensitivity and specificity were calculated as above (see **Appendix**), considering only voxels within an 11 by 11 by 11 grid (5 voxels in

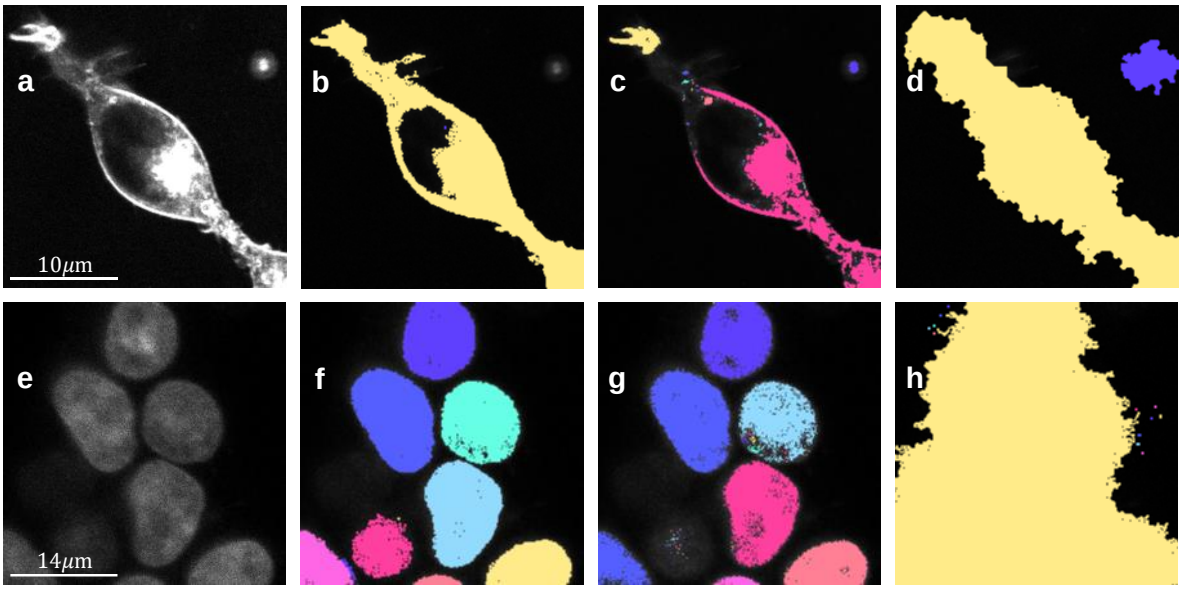


Figure 2.9: Results of segmentation algorithms on experimental data. **a** Image of a HEK293 cell stained with di-4-ANEPPDHQ. **b-d** Masks of the HEK293 cell as identified by TOBLERONE, Otsu thresholding and the Watershed algorithm, respectively. **e** An image of several HEK293 cell nuclei stained with Nucblue. **f-h** Masks of each nucleus as identified by TOBLERONE, Otsu thresholding and the Watershed method, respectively.

each direction) of each active voxel comprising the ground-truth object boundaries. Each simulation included a 256 by 256 pixel ROI with a depth of up to 50 frames. For all stacks, runtime was less than 2 minutes on a single processor.

3DTOBLERONE achieved a mean sensitivity of 0.9362 and a mean specificity of 0.9549. The sensitivity of 3DTOBLERONE surpassed the sensitivity of both 3D Simple Segmentation and 3D Spot Segmentation (0.2863 and 0.3248 respectively). The specificity of 3D Simple Segmentation (0.9998) and 3D Spot Segmentation (0.9977) exceeded the specificity of 3DTOBLERONE. To quantify the impact of data quality on 3DTOBLERONE's performance, each of the 100 simulated images was overlaid with varying degrees of Gaussian noise and blur (see **Materials and Methods** for

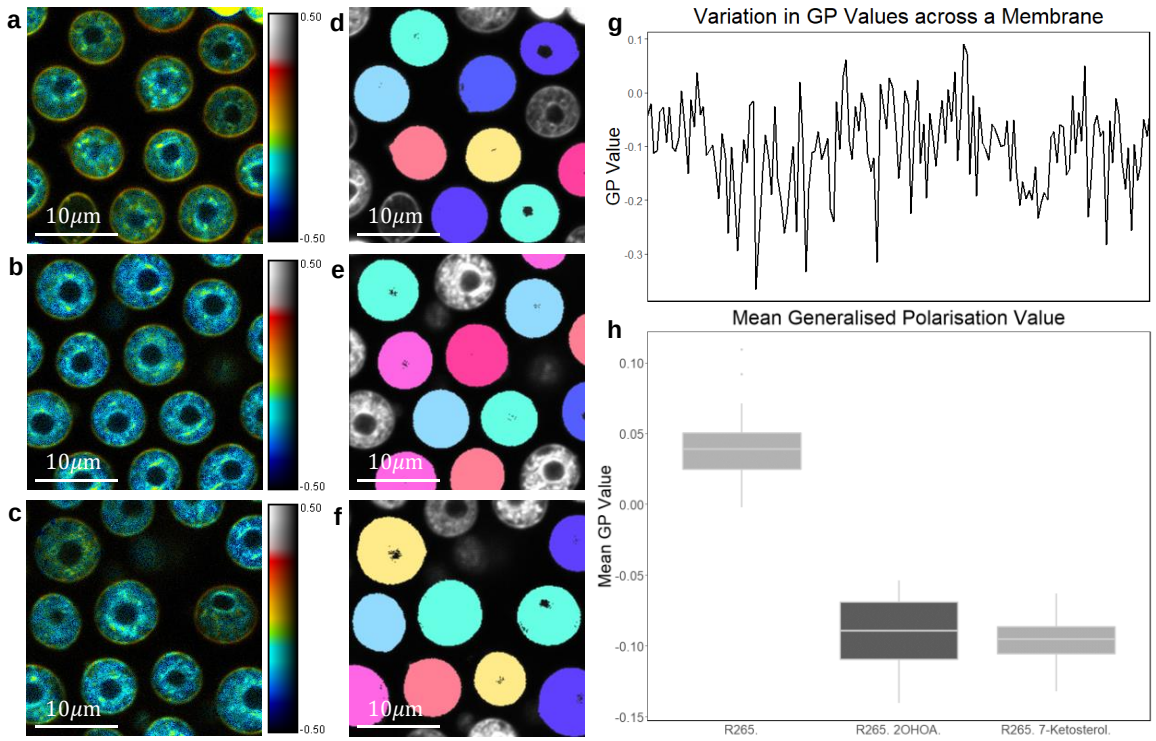


Figure 2.10: TOBLERONE as a tool for identifying variations in membrane lipid order. **a-c** GP images of *C. gattii* cells from the control group, those treated with 2OHOA and those treated with 7-ketocholesterol, respectively. Cells were stained with di-4-ANEPPDHQ. Psuedocolour applied to reflect the difference in GP values across the membrane. **d-f** Masks of the same cells, as found by TOBLERONE, overlaid onto the grayscale images. Incomplete cells lying on the image periphery have been manually excluded. **g** The GP line profile extracted from the boundary of a *C. gattii* cell. **h** Mean GP value across all pixels comprising the membrane. Mean values were taken per cell. Treatment conditions are as follows: untreated R265 cells (R265), R265 cells treated with 2OHOA (R265 2OHOA) or R265 cells treated with 7-ketocholesterol (R265 7-Ketosterol).

quantification). **Figure 2.12** highlights the change in mean sensitivity and specificity across all data for these data quality conditions. Results suggest that the sensitivity and specificity of 3DTOBLERONE generally deteriorate as σ_1 or σ_2 increase. As with 2DTOBLERONE, segmentation sensitivity shows a greater response to changes in noise, while specificity seems to be more sensitive to changes in blur.

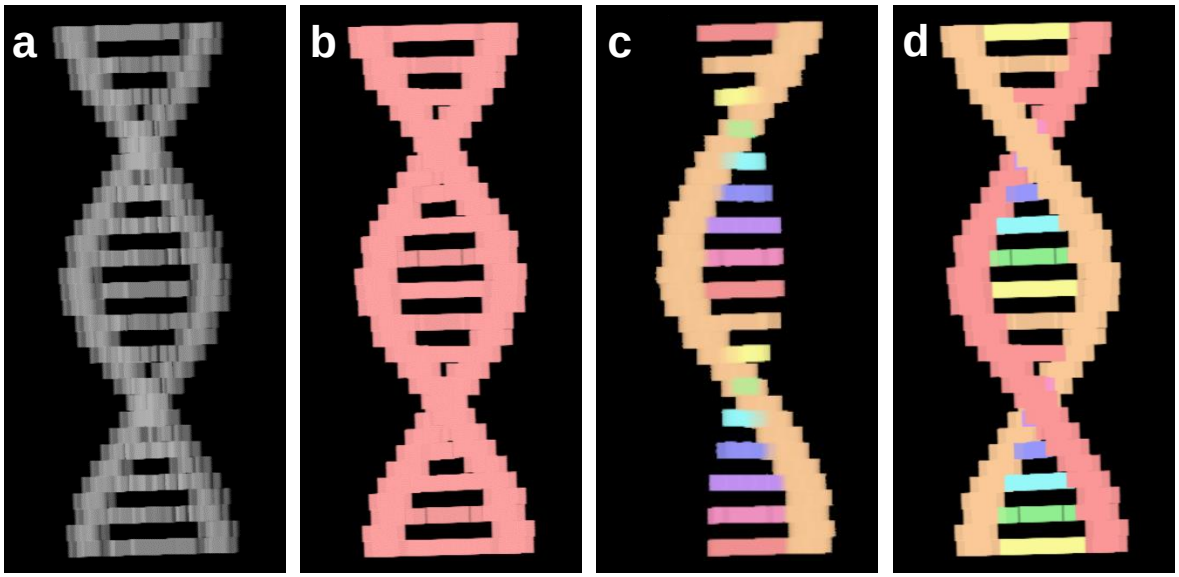


Figure 2.11: Exemplar cases of 3D segmentation. **a** A simulated double helix structure. Branches between the two main backbone strands have a lower voxel intensity than the strands themselves. **b** Results of 3D Simple Segmentation on helix data. The entire structure is returned as one object. **c** Results of 3D Spot Segmentation on helix data. A significant portion of the object is no longer detected. **d** Results of 3D TOBLERONE on helix data. The two main backbone strands and each branch between are detected as separate components. Exemplar stacks were adapted from Panconi et al[277].

2.3.4 3DTOBLERONE for volumetric cell segmentation

Non-activated Jurkat T-cells were cultured and labelled with both Nuclear Mask DeepRed and WGA-AlexaFluor 488 before fixation with 4% PFA. A z-stack was recorded for each colour channel using a Zeiss LSM 900 confocal microscope in confocal scan mode. This cell line typically exhibits a spherical morphology, with cells floating separately in suspension (**Figure 2.13a**), making them a good candidate for testing image segmentation algorithms. Compared to 3D Simple Segmentation (**Figure 2.13b**) and 3D Spot Segmentation (**Figure 2.13c**), 3DTOBLERONE shows appropriate discrimination against background, but may be sensitive to membrane perturbations or blur at the cell periphery (**Figure 2.13d**). In this instance, we have

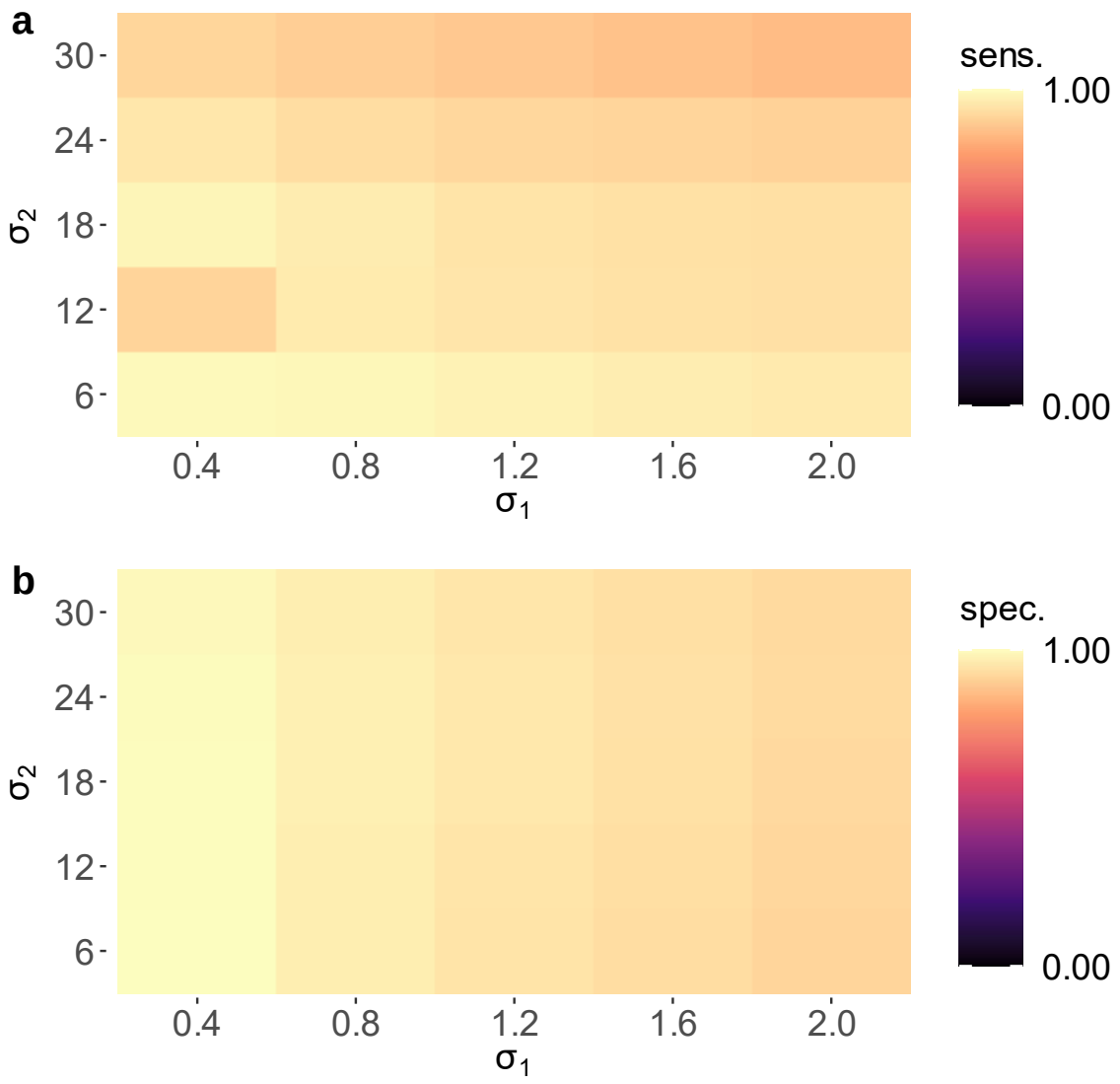


Figure 2.12: Results of 3DTOBLERONE segmentation on simulated data sets. Changes in mean sensitivity (**a**) and specificity (**b**) for all data sets are recorded against the standard deviations of the Gaussian blur (σ_1) and Gaussian noise (σ_2) used.

extracted the distribution of cell volumes (**Figure 2.13e**) for all 31 cells identified. Results suggest that the average volume of a Jurkat T-cell is 1961.6μm^3, corresponding to an average diameter of ~15.5μm (assuming circularity), which is in accordance with existing literature[282-284].

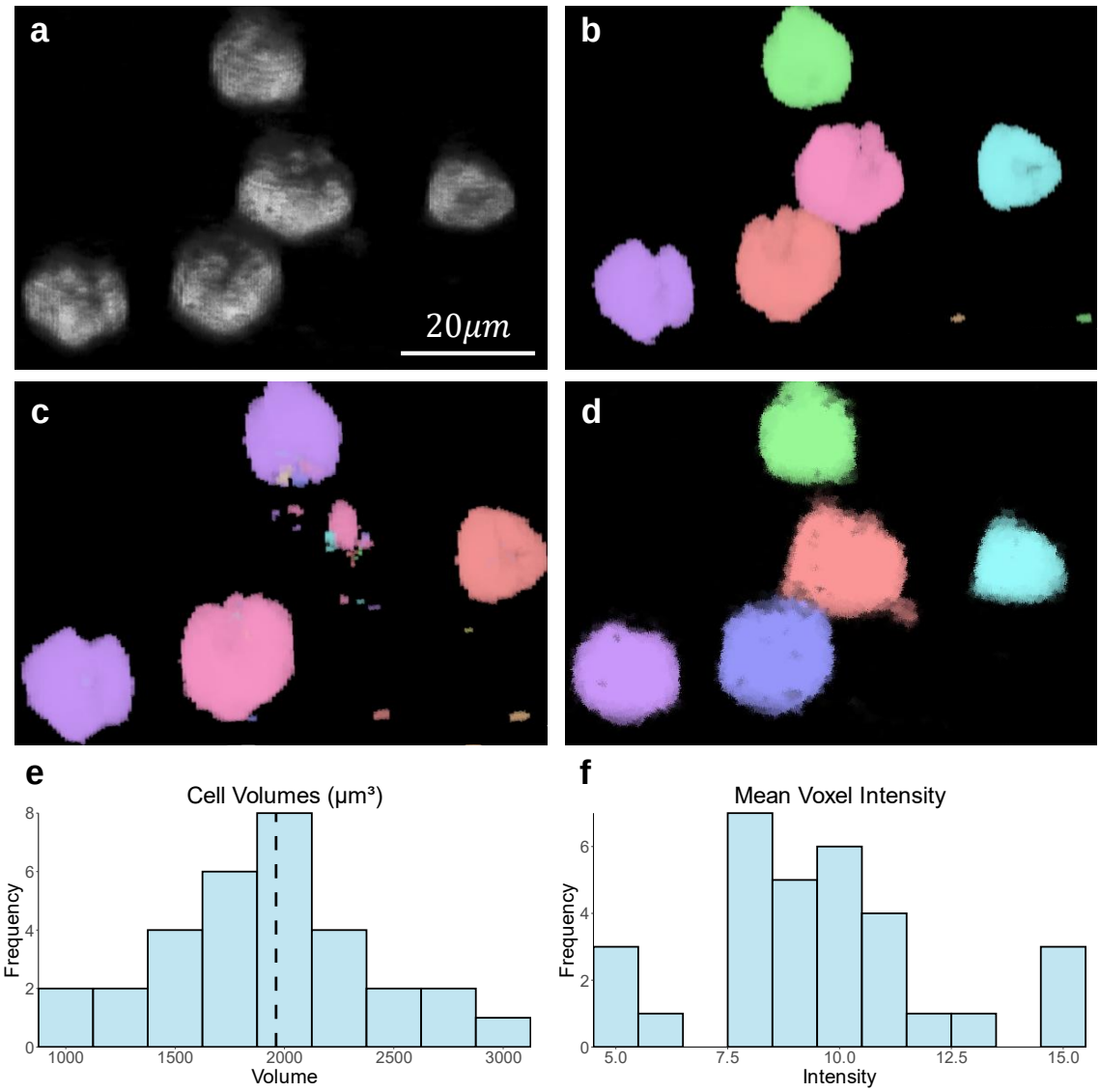


Figure 2.13: Segmentation algorithms on experimental data. **a** 3D visualisation of fluorescing Jurkat T-cells. **b** Results of 3D Simple Segmentation on cell data. **c** Results of 3D Spot Segmentation on cell data. **d** Results of 3D TOBLERONE on cell data. **e** Histogram of cell volumes identified by 3D TOBLERONE, the mean volume of 1961.6μm^3 is signified by a dashed line. **f** Histogram of mean voxel intensity identified by 3D TOBLERONE.

2.3.5 Dynamic t-stack segmentation with tempTOBLERONE

Further to the analyses on volumetric simulations, we generated 100 t-stacks, with cell masks moving between frames to simulate cell motility (see **Materials and Methods** for all data simulation methods). Here, we quantify the sensitivity and specificity of

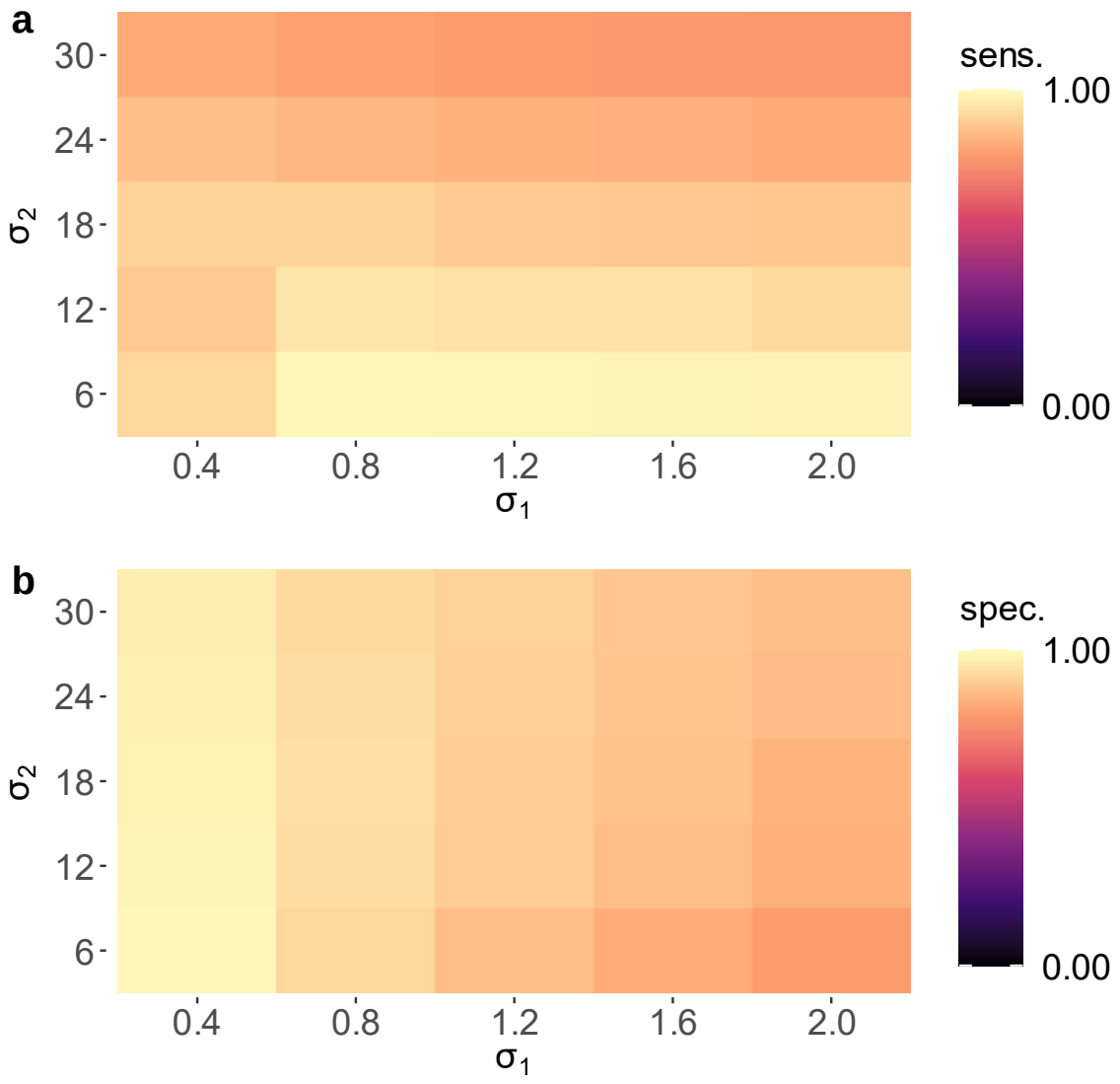


Figure 2.14: Results of tempTOBLERONE segmentation on simulated data sets. Changes in the mean sensitivity (**a**) and specificity (**b**) for all data sets are recorded against the standard deviations of the Gaussian blur (σ_1) and Gaussian noise (σ_2) used.

tempTOBLERONE under varying noise and blur (see **Materials and Methods** for quantification). As before, we determine the sensitivity, specificity and number of connected components returned for each data set. For t-stacks, an 11 by 11 by 3 pixel grid was explored around each active ground-truth pixel. This ensured that each frame

would only be compared to the previous and subsequent frames. All simulated data sets, each including a 256 by 256 pixel ROI with up to 22 frames, took no longer than 2 minutes to segment on a single processor. The mean sensitivity and specificity scores of the segmentations returned by tempTOBLERONE for all data quality conditions are given in **Figure 2.14**. As above, tempTOBLERONE experiences a drop in both sensitivity and specificity as either σ_1 or σ_2 are increased. Across all data quality conditions, tempTOBLERONE achieved an average sensitivity of 0.8606 and average specificity of 0.9749.

2.3.6 TIA for cell tracking and trajectory mapping

S. pombe, expressing BOP1-mCherry, was cultivated and genetically modified under standard protocols, with the BOP1 gene tagged with mCherry through homologous recombination[271]. Cells were imaged in sealed growth chambers containing 2% agarose YES medium, captured using a Zeiss Axiovert 200M microscope equipped with an UltraView RS-3 confocal system, with image z-stacks obtained at 1-minute intervals. Then, z-stack maximum projection images were processed using FIJI/ImageJ. Images were separated into distinct channels to distinguish nuclei from cell membranes and analysis was conducted on the underlying nuclear dye channel (taken at 600-710nm wavelength light in accordance with mCherry emission spectrum) to isolate nuclear envelopes from cell plasma membranes (**Figure 2.15a-b**) and promote clearer segmentation. Changes in connectivity, here brought about by nuclear division, are tracked over time (**Figure 2.15c**). In particular, an additional component was generated between the 12th and 13th minute of acquisition, suggesting that progression from telophase to interphase in the nuclear division of *S. pombe* can occur in under a minute, in accordance with existing literature[285]. Furthermore, we apply

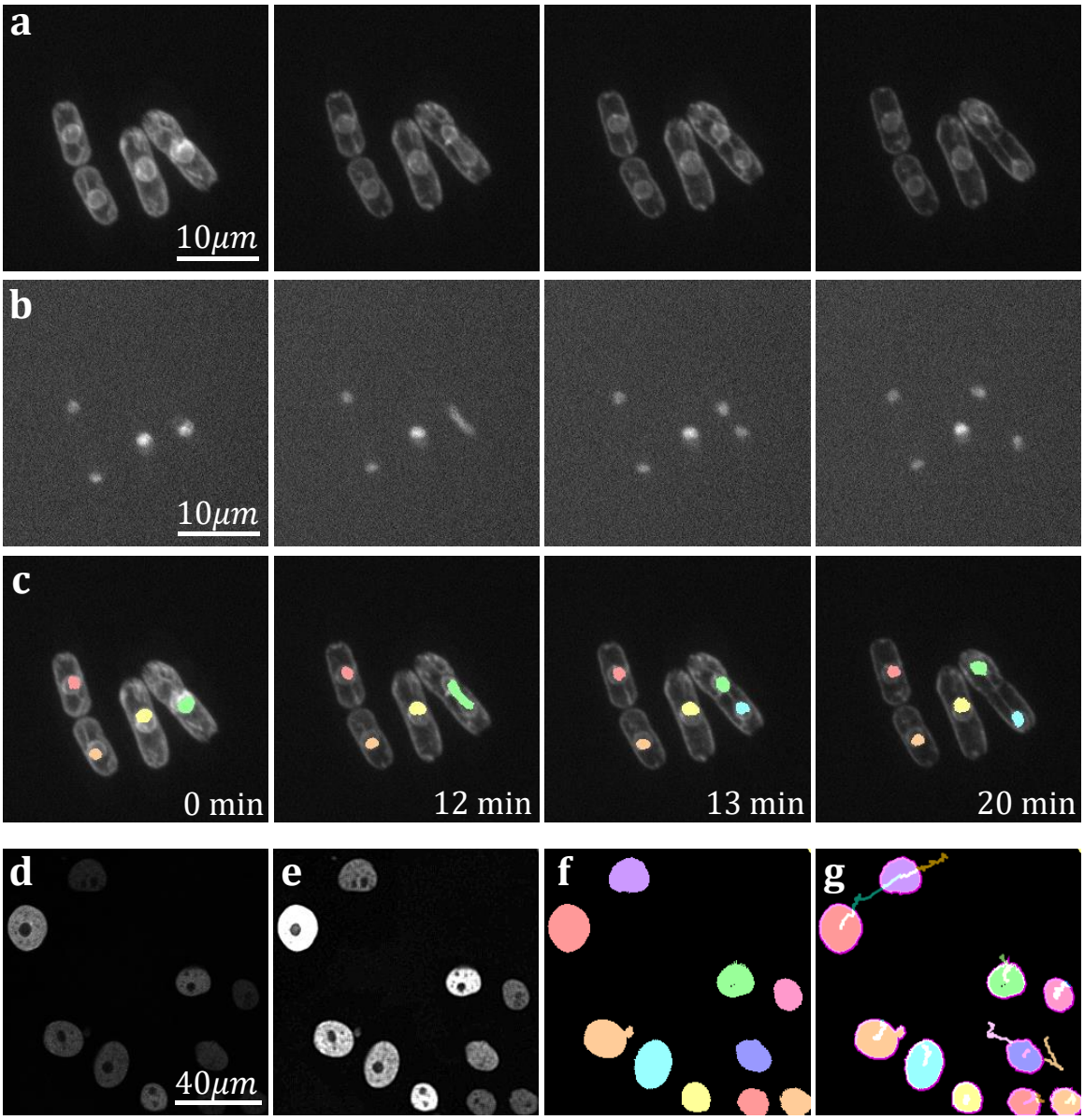


Figure 2.15: Applications of tempTOBLERONE to cell data. Time series data of *S. pombe* cells (**a**) and nuclei (**b**) undergoing nuclear division. **c** Spatiotemporal segmentation of nuclei undergoing division. tempTOBLERONE was performed on stacks in **b**, with masks overlaid onto images in **a**. A nuclear division is recorded at 13min. **d** Snapshot of GFP-GOWT1 mouse stem cell data. **e** Pre-processing: brightness of each pixel is doubled to increase image contrast. **f** Segmentation results from temporal TOBLERONE. **g** Track lines derived from applying Trackmate to segmentation.

tempTOBLERONE to GFP-GOWT1 mouse stem cell data (**Figure 2.15d**), acquired under a Leica TCS SP5 with a Plan-Apochromat 63x/1.4 (oil) objective lens, as taken from the Cell Tracking Challenge[270,286]. Following pre-processing (**Figure 2.15e**), results show segmentation with clear background separation (**Figure 2.15f**). By inputting cell masks directly into track analysis software, such as Trackmate, we can identify cell trajectories (**Figure 2.15g**)[266,276]. This suggests that topological video analysis may be a viable avenue for tracking cell motility and mitotic processes.

2.4 Discussion

This chapter introduces TOBLERONE, an image segmentation algorithm which makes use of persistent homology to construct connected components from pixel intensity variations. This allowed for generic image segmentation and object quantification in 2D spatial, 3D volumetric and 2D + time data. Each algorithm was performed on a series of simulated data sets to quantitatively compare segmentation sensitivity and specificity against pre-existing classical methods. Then, TOBLERONE was applied to conventional FM data sets, comprising a range of cell and organelle geometries. This included *C. gattii*, *S. Pombe*, Jurkat T cells and GFP-GOWT1 mouse stem cells, as well as HEK293 cells and their nuclei.

2.4.1 Summary of results

2D TOBLERONE achieved a mean sensitivity of 0.9412 and a mean specificity of 0.8762. Notably, Otsu thresholding showed a sensitivity of 0.8726 and specificity of 0.9306, while the Watershed algorithm achieved a sensitivity of 0.9964 and specificity of 0.4322. Therefore, 2D TOBLERONE may be beneficial to users who prioritise segmentation sensitivity without largely compromising specificity. 3DTOBLERONE achieved a mean sensitivity of 0.9362 and a mean specificity of 0.9549, which

surpassed the sensitivity of 3D Simple Segmentation and 3D Spot Segmentation (0.2863 and 0.3248) but did not surpass the specificity (0.9998 and 0.9977). However, this may arise from a lack of discrimination against the background. Notably, the difference between sensitivity and specificity is lower for 3DTOBLERONE than the other algorithms considered here, which may suggest that 3DTOBLERONE provides a better trade-off between the segmentation statistics. Across all data sets, tempTOBLERONE achieved an average sensitivity of 0.8606 and average specificity of 0.9749. For all variants, this implies that most pixels belonging to an object were correctly identified and most pixels belonging to the background were correctly ignored, respectively. Quantification of algorithmic performance under imposed Gaussian blur (with standard deviation σ_1) and Gaussian noise (with standard deviation σ_2) suggests that all TOBLERONE variants yield a small decrement in both sensitivity and specificity as either σ_1 or σ_2 are increased. However, sensitivity experiences a greater decrement with changes in noise, while specificity seems to be more sensitive to changes in blur. This may be because noisy pixels within the cell body may not be detected by the algorithm, while increased Gaussian blur raises the intensity of background pixels at the cell periphery (and makes them more likely to be aggregated into the cell component).

In experimental HEK293 data, TOBLERONE showed invariance to cell geometry and suitable background-foreground separation. Performance remains consistent across both cell plasma membranes and nuclei, suggesting that planar TIA may be a viable method for both cell and organelle segmentation. By extracting boundary line profiles from cell plasma membranes of *C. gattii*, we produce quantitative analysis of membrane order under varied cell culture conditions. In particular, confocal imaging with the polarity-sensitive dye di-4-ANEPPDHQ reveals a decrease in GP value for *C.*

gattii cells supplemented with either 2OHOA or 7-ketocholesterol, compared to control. Applications to volumetric experimental data of Jurkat T cells suggests an average diameter of approximately ~15.5µm, in line with existing literature[282-284]. Concurrently, in *S. pombe* t-stack data, we observe progression in nuclear division from telophase to interphase in under a minute, in line with existing reports[285]. These statistics act as a proof-of-principle for further use of TOBLERONE in FM data segmentation and highlight the qualitative benefits of TIA for cell segmentation.

2.4.2 Considerations for the use of TOBLERONE in cell segmentation

TOBLERONE permits image segmentation without parameterisation of geometric object properties and, in the case of supervised machine learning, the requirement of training data sets. The persistence parameter quantifies the difference between the minimum and maximum intensities of pixels within any component. However, if the background intensity is greater than the minimum object intensity, TOBLERONE is liable to overestimate object size. In this case, a lower persistence threshold may be required, but this may lead to over-segmentation – in which single objects are segmented into several fragments. Further, since each connected component is considered separately, a new entry must be generated in the data structure for every component identified. When appending this new entry to the list, the R programming language will copy the entire list to a new location in memory to accommodate the new element[287]. This can slow data processing, leading to longer execution times and increased memory usage[287]. Deterioration of signal within the cell by noise may also lead to sparser segmentation. TOBLERONE may aggregate multiple cells into one component in image data which presents many cells in close proximity. To avoid this drawback, there needs to be at least a one-pixel gap of “background” (i.e. pixels with

intensity more than τ different from brightest cell pixel intensity) between all pixels in one cell and all pixels in any other cell.

Video segmentation may require separate algorithms for the purposes of binarisation, segmentation and tracking, which themselves depends on user preferences and parameterisation. Traditional tracking algorithms, such as nearest-neighbour approach, suffice for the purposes of establishing temporal connectivity, but may not handle object appearance and disappearance[266]. Linear assignment problem tracking links objects across frames by solving a cost matrix that minimises the overall distance between detected positions[288]. This permits object appearance and disappearance, but may not explicitly track lineage without adaptation[288]. tempTOBLERONE traces the behaviour of dynamic components through changes in connectivity and accounts for the appearance, disappearance, and lineage of components. However, it should be noted that tempTOBLERONE is unsuitable for t-stacks in which each cell's position does not overlap with its position in the following frame. This depends largely on the temporal resolution of the microscope, the imaging setup (drift), and cell motility. In particular, if the total displacement of a cell across any two consecutive frames exceeds the cell length, tempTOBLERONE will not correctly track the cell.

2.4.3 Concluding remarks

In summary, this work presents a novel application of TIA for segmenting fluorescence microscopy data which is, in accordance with literature, the first of its kind. TOBLERONE presents a complementary approach to image analysis, which circumvents the constraints of geometric or machine-learning based image segmentation. As discussed, images from conventional microscopy cannot resolve below the diffraction limit of light, which means we cannot fully distinguish nanoscale

lipid compartmentalisation or molecular interactions. However, SMLM, can be used to determine GP values at individual molecule localisations with nanoscale precision. As above, topological data analysis tools may prove beneficial in segmentation and quantification of this alternate coordinate-based data type. However, in order to validate these analytic techniques, ample simulated data is required. In the next chapter, we introduce a computational framework for simulating protein aggregation dynamics which can recapitulate static spatial and marked point patterns from experimental data.

Chapter 3: *In silico* protein aggregation dynamics

As discussed in **Chapter 2**, a TDA framework can be used for feature extraction in conventional FM images. The output of SMLM may also take the form of a marked data type, with individual molecules localised across a continuous range of spatial coordinates. Therefore, topological feature extraction may also be applicable to SMLM data. However, to determine whether topological methodologies generalise to SMLM data, it is necessary to validate against ground truth data with a range of point pattern geometries. This data could arise from a mathematical model which approximates experimental data. Furthermore, the dynamic reorganisation of some transmembrane proteins into nanoscale clusters is hypothesised to regulate downstream signalling. As such, this model could serve a dual purpose: to simulate SMLM data which represents the geometric properties of experimental data, and to predict protein aggregate behaviour in dynamic systems. The following chapter concerns the development of a computational framework (termed ASMODEUS) for modelling protein aggregation dynamics on the plasma membrane. ASMODEUS is validated on simulated data presenting a range of cluster properties. We demonstrate how the model could be used to recapitulate the varied point pattern geometries arising from SMLM. Further, we highlight how ASMODEUS can be used to model dynamic receptor aggregation in both simulated data and in experimentally-derived DNA-PAINT localisation data for the model protein complex TCR.

Contributions

Contributing authors

L. Panconi[1,2,3], O. P. L. Dalby[4,5], C. Zaza[4], D.M. Owen[1,3,6], S. Simoncelli[4,5] & J. Griffié[6,7]

Affiliations

[1]Institute of Immunology and Immunotherapy, University of Birmingham, Birmingham, UK.

[2]College of Engineering and Physical Sciences, University of Birmingham, Birmingham, UK.

[3]Centre of Membrane Proteins and Receptors (COMPARE), University of Birmingham, Birmingham, UK.

[4]London Centre for Nanotechnology, University College London, London, UK.

[5]Department of Chemistry, University College London, London, UK.

[6]School of Mathematics, College of Engineering and Physical Sciences, University of Birmingham, Birmingham, UK.

[7]Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden.

Author contributions:

L. P. and J. G. conceptualised the project and methodology. L. P. wrote ASMODEUS simulation code, performed simulations and data analysis, and wrote the following chapter. O. P. L. D., C. Z. and S. S. undertook Jurkat T cell culture, CD3ζ DNA-PAINT data acquisition and qPAINT post-processing (**Figures 3.9** and **3.10**). J. G., S. S. and D. M. Owen provided project supervision.

3.1 Introduction

Membrane protein dynamics are highly organised and play a role in regulating cellular signalling processes[130,195,289]. A range of biophysical processes, with interacting and competing effects, are hypothesised to govern protein aggregation dynamics (PAD) at

the plasma membrane[29,31,197,242]. These biophysical phenomena culminate in the restriction of free molecular diffusion and typically promote non-random re-organisation of proteins into spatial clusters[64]. It is hypothesised that these clusters convert analogue extracellular signals into digital intracellular signals, in a process known as digitisation, and that this conversion increases signalling fidelity[16]. As such, the formation of such nanodomains is a key membrane property affecting cellular behaviour and signalling pathway functionality[290-292].

For example, nanoscale TCR clustering is largely responsible for initiating T cell early activation and the development of the immune response[293]. When contact occurs between a T cell and an antigen presenting cell, a close-contact zone forms, TCRs are engaged, and the local organisation of the receptors and their associated membrane proteins shifts. For instance, the glycoprotein CD3ζ organises into nanoclusters during assembly of the TCR–CD3 complex[293,294], and the formation of this complex is the primary determinant of T cell activation and the immune response[295]. There is ongoing research into whether therapeutic mediators which induce or disrupt TCR aggregation could be used in immunotherapy or the treatment of autoimmune disorders[13,296,297].

Accurate characterisation of nanoscale clustering is essential for interpreting the biological systems which regulate cell behaviour. Recently, analyses tailored to protein maps derived from super-resolution modalities have seen widespread advancement and implementation. As such, there is a requirement for a simulator that can model protein aggregation dynamics, with nanoscale spatial resolution, corresponding to the data type of this class of microscope. This model could present a dual purpose: to generate static distributions which recapitulate SMLM data, and to model emergent protein aggregate behaviour in dynamic systems. However, the presence of multiple

interacting biophysical processes yields increasing computational complexity, and a practical simulator would recapitulate experimentally-derived protein distributions without requiring a detailed knowledge of the myriad of biophysical influences.

Research into the mechanistic properties of biophysical systems has found increasing success in the development of computational models of cellular signalling and metabolic control[298]. Advanced simulation techniques (such as SuReSim[299], FluoSim[300], SMeagol[301], TestSTORM[302], ThunderSTORM[179], LocMoFit[167] and SMIS[303]) generate localisations by simulating the photophysical properties of fluorescent molecules. Traditionally, these models assume the spatial distribution of points and then recapitulate them from an analogous probability distribution, typically Gaussian, with clusters and structures placed at defined intervals[212,214,243]. However, this imposes strict geometric properties on simulation, whereas molecular distributions observed through bio-imaging could express highly varied shapes and densities[304,305]. Ripley's K function, which measures point density within concentric circles, can be used to compare cluster properties across a range of spatial scales and estimate ensemble cluster parameters[306]. State-of-the-art simulators have used the linearised, localised Ripley's function, L_R (a scalar value assigned to each localisation), to measure cluster affinity and influence PAD in ABMs[214,307]. In this work, we expand upon this model, incorporating the full range of the global Ripley's K-function to guide system evolution.

Here, we introduce Agent-based Spatiotemporal Molecular Distributions Evolving Under Simulation (ASMODEUS)**,** an agent-based model (ABM) designed for simulating PAD. ASMODEUS is invariant of input data geometry and relies only on a measure of local density. This permits emulation of experimentally-derived protein maps without imposing restrictions on data structure. By learning topological features

directly from data, we circumvent simplifying assumptions and generate model simulations which better capture the distribution, density and dynamics of protein aggregates. Within this framework, we include a method for introducing nucleation sites, which aggregate a fixed number of nearby agents and can represent cross-linking or forcible aggregation of proteins (induced, for instance, by therapeutic strategies). We apply the model to investigating the role of the CD3ζ distribution, validated against experimental data acquired from DNA-PAINT microscopy in activated and non-activated T cells. We deduce the receptor cross-linking conditions required to maximise the probability of sustained CD3ζ cluster formation as a proxy for T cell activation. Further, we demonstrate a proof-of-concept for ABMs of multiple interacting molecular species – which, to the best of our knowledge, is the first of its kind.

3.2 Materials and methods

Jurkat T cell culture

Jurkat E6.1 T cells were cultured in R10 medium (Roswell Park Memorial Institute [RPMI] 1640 medium supplemented with 10% FBS, 1mM penicillin and streptomycin [PenStrep], 2mM L-glutamine and 1mM sodium pyruvate [all from Sigma-Aldrich, Madison, WI]) at 37°C in a 5% CO_2 incubator.

Bilayer preparation

0.4mM liposome solution with a lipid molar ratio of 97.4% DOPC (Avanti Polar Lipids, 850375C), 2% DGS-NTA(Ni) (Avanti Polar Lipids, 790404C), 0.1% Biotinyl-Cap-PE (Avanti Polar Lipids, 870273C) and 0.5% PEG5,000-PE (Merck, Kenilworth, NJ, USA, 880220P-200MG) was produced by vesicle extrusion through a 100nm pore-size

polycarbonate filter. Glass coverslips were cleaned via sonication with 2% Hellmanex detergent for 30 minutes at 37kHz, 70% power, before rinsing with deionised water. Coverslips were then washed with 70% ethanol and dried under N_2. Dry coverslips were plasma cleaned immediately prior to use (Diener Zepto plasma cleaner, 40kHz generator, 90s at 70W power). Ibidi sticky chamber slides (Ibidi Sticky-Slide VI 0.4, 80608) were attached to the cleaned coverslip and 0.4mM liposome solution was added to each well for 20 minutes, then washed three times with lipid buffer (0.1% Bovine Serum Albumin (BSA), 2mM $MgCl_2$ and 1mM $CaCl_2$ in PBS). 0.1mM of $NiCl_2$ in 2% BSA in PBS was added for 20 minutes to recharge NTA groups and block surface prior to protein addition. Surfaces were subsequently washed three times with lipid buffer. Disruption of the lipid bilayer was avoided by maintaining 50µL of lipid buffer in wells. For functionalisation of the bilayer with biotinylated proteins, 12.5µg/mL of streptavidin (Cambridge Bioscience RayBiotech, 228-11469-2) in lipid buffer was added for 20 minutes and then washed three times with lipid buffer. The streptavidin coated bilayer was functionalised with addition of 3.4µg/mL biotinylated aCD3 (Biolegend, 317320), 3.4µg/mL biotinylated aCD28 (Biolegend, 302904) and 200ng/mL His-tagged ICAM-1 (Thermo Fisher Scientific, A42524) in lipid buffer for activated conditions, or 200ng/mL His-tagged ICAM-1 alone in lipid buffer for non-activated conditions. After 20 minutes, the bilayer was washed three times with lipid buffer to remove unbound proteins.

DNA-Fab conjugation

A malemide-PEG2-succinimidyl ester coupling reaction was used for DNA labelling of Fab Fragment Donkey Anti-Rabbit IgG (Jackson ImmunoResearch, 711-007-003). Thiolated-DNA 5'-Thiol-ACACACACACACACACACA-3' (Eurofins Genomics,

Ebersberg, Germany) (13µL, 1mM) was reduced for 2h by incubating with DTT solution (30µL, 250mM) (Thermo Fisher Scientific, Waltham, MA, USA, A39255) at room temperature on a shaker in darkness. Maleimide-PEG2-succinimidyl ester crosslinker solution (1µL, 23.5mM) (Sigma-Aldrich, St. Louis, MO, USA) was incubated with anti-rabbit Fab fragment (50µL, 26µM) for 90 min at 4°C on a shaker in darkness. Excess DTT and crosslinker were removed under spin filtration by Microspin Illustra G-25 columns (GE Health-care, Chicago, IL, USA) and Zeba spin desalting columns (7K MWCO, Thermo Fisher Scientific), respectively. Resultant products were mixed and incubated overnight at 4°C on a shaker in darkness. Residual unbound DNA was removed by Amicon spin filtration (Merck) and DNA-Fab concentration was measured using a NanoDrop One spectrophotometer (Thermo Fisher Scientific). Quantification of DNA-Fab coupling ratio through spectrophotometric analysis revealed a ratio of ~1.

T cell activation on bilayer

Wells containing functionalised bilayers were washed with sterile 1X PBS then warmed at 37°C for 15 minutes prior to adding Jurkat T cells. Cells were added to the bilayer at a density of 10^6 cells/chamber and left to incubate for 15 minutes at 37°C. Cell fixation was achieved using 4% PFA in PBS warmed to 37°C, then leaving cells for 30 minutes at room temperature before washing with 60mM Glycine in PBS. Cells were permeabilized with Triton X-100 (VWR, 28817.295) at 0.1% for 5 minutes at room temperature and washed with 60mM Glycine in PBS before immunostaining. Cells were then blocked with 5% BSA in PBS for 1h at room temperature.

CD3ζ immunostaining

Cell staining was performed with a primary antibody against CD3ζ (Abcam, ab40804), achieved with 5% BSA in PBS at a concentration of 1.7µg/mL for 1h at 37°C, then washing with PBS. Secondary staining was performed with the DNA conjugated Rabbit Fab in 5% BSA in PBS at a concentration of 4µg/mL, this was left for 1h at room temperature, and washed with PBS. For drift correction, 90nm gold nanoparticles (Cytodiagnostics, G-90-100) were added to each well for 10 minutes then washed with PBS. Imager DNA 5'-TGTGTGT-Cy3B-3' (Eurofins Genomics) was added at 0.5nM in a solution of 500mM NaCl, 1mM EDTA in 1X PBS, immediately prior to imaging.

DNA-PAINT imaging and image reconstruction

Imaging of Jurkat E6.1 T cells was undertaken on a custom-built TIRF microscope, based on a Nikon Eclipse Ti-2 microscope (Nikon Instruments, Tokyo, Japan) equipped with a 100× 1.49 NA oil immersion TIRF objective (Apo TIRF) and a Perfect Focus System. A 560nm excitation laser (MPB Communications, 1 W) was used with a polarised and quarter lambda waveplate to ensure circular polarisation. The excitation beam was passed through a clean-up filter (FF01-390/482/563/640-25, Semrock, Rochester, NY, USA) and coupled into the objective via a beam splitter (Di03-R405/488/561/635-t1-25×36, Semrock). Fluorescence light was spectrally filtered by an emission filter (FF01-446/523/600/677-25, Semrock), imaged with an sCMOS camera (ORCA-Flash4.0 V3 Digital, Hamamatsu, Hamamatsu City, Japan) and subject to 2 × 2 binning, yielding a final pixel size of 130nm in the focal plane. 10000 frames were taken per acquisition with 100ms integration time and a laser power density of 1.6kW/cm^2 over the sample. Post-processing of raw fluorescence data was performed in Picasso (v. 0.6.4) and MATLAB (R2022a)[308]. Drift correction

was undertaken via redundant cross-correlation using 90nm gold nanoparticles (Cytodiagnostics, G-90-100) as fiducials.

CD3ζ post-processing and cluster analysis

After filtering and drift-correction, super-resolved Jurkat E6.1 T cell images were analysed with DBSCAN cluster analysis under the Picasso Render module. DBSCAN parameters were set as *minPts* = 5 and ϵ = 13nm. The ϵ parameter was taken to be the localisation precision of the DNA-PAINT images, determined via nearest-neighbour based analysis, and commensurate with author suggestions. The *minPts* parameter was selected for 5 localisations, in accordance with binding kinetics of the imager–docking pair, the DNA imager concentration, the number of recorded frames, and statistical considerations for the distribution of single-molecule localisations.

qPAINT analysis

A custom-written MATLAB code was used to analyse the fluorescence time series of each detected cluster and estimate the per-cluster number of CD3ζ molecules via qPAINT analysis, as described in literature[158,213]. Frame numbers of each localisation within the same cluster, as determined by DBSCAN, were used to recover the dark state times per cluster – that is, the continuous amount of time in which single molecule localisations were absent. All per-cluster dark times were pooled to obtain a normalised cumulative histogram, which was then fitted to the exponential function: $1 - \exp(t/\tau_d)$ to estimate the per-cluster dark time, τ_d. The qPAINT index of each cluster (qi) was taken as the inverse of the corresponding dark time. The qPAINT indices of clusters with a maximum point distance of 50nm were plotted as a cumulative histogram and fitted via a multi-peak Gaussian. Peaks were found with a frequency of

0.0075Hz. This corresponds to the qPAINT index of a cluster of localisations representing one binding site, that is, CD3ζ monomers (qi1). The number of proteins within a cluster of localisations was defined by the ratio between qi1 and the corresponding qPAINT index (qi1/qi). Protein positions were plotted by recovering a likely distribution of CD3ζ receptors in each cluster via k-means clustering, where k is equivalent to the ratio between qi1 and qi.

Point cloud simulation

Spatial point patterns were simulated by generating a fixed number of circular clusters, n_t, with pre-determined cluster radius, r_t, and number of points per cluster, p_t, and then overlaying outliers subject to a given background-to-cluster ratio, b_t. Central cluster coordinates were randomly selected such that all clusters would be contained within the 3µm × 3µm ROI. Then, p_t points were uniformly generated around each centre at distances of $r_t X^2$, where $X \sim Unif(0,1)$. The choice of parameter values used in point pattern simulations (unless otherwise specified) is given in **Table 3.1**.

Parameter	**Function**	**Values**
n_t	Number of clusters	10
r_t	Cluster radius.	10nm, 30nm, 50nm
p_t	Number of points per cluster.	5, 15, 30
b_t	Background to cluster ratio.	0, 0.33, 0.66

Table 3.1: Parameters used to simulate target distributions.

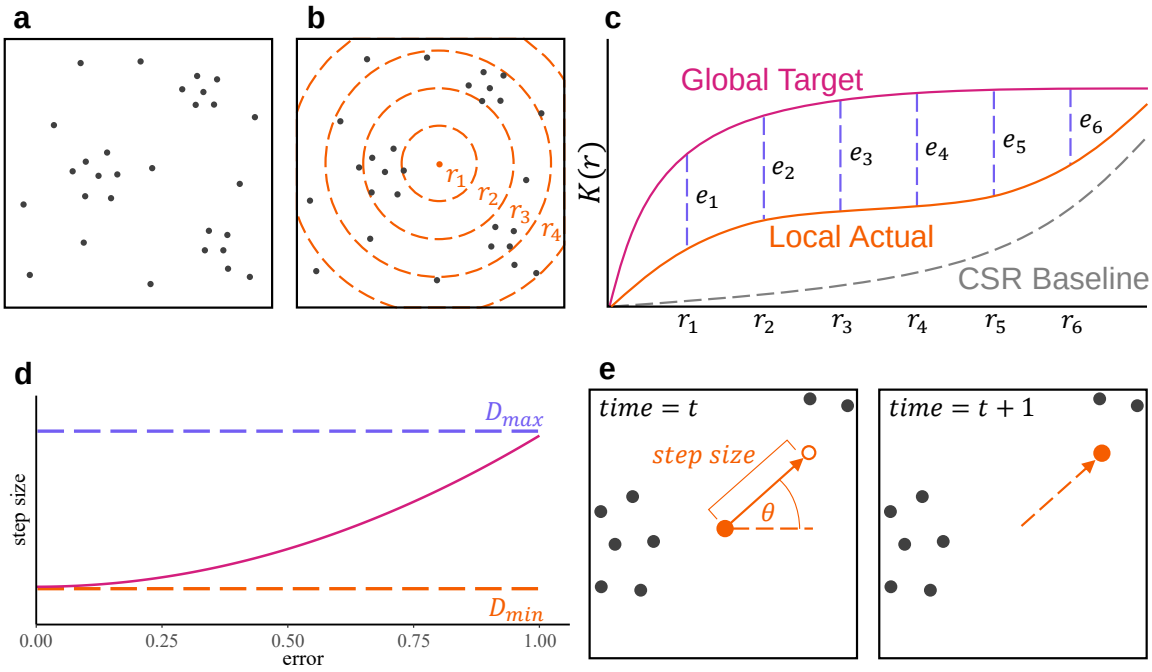


Figure 3.1: Simulating PAD with ASMODEUS. **a** Example target point cloud. **b** Calculating density for each point. **c** Formulation of Ripley's K function and calculation of ratiometric error. Error is defined between 0 and 1 for each discrete radial value, with the maximum value of 1 assigned below the CSR baseline. **d** Step size is bound between minimum and maximum step size parameters (D_{min} and D_{max}) and increases quadratically with average error. **e** Zoomed region centred around the orange point marked in **b**. From one frame to the next, the point is offset in the direction of angle θ, where $\theta \sim Unif(0, 2\pi)$ is chosen randomly, with step size calculated from **d**. This process is repeated for each point at each time frame.

Simulating PAD through agent-based modelling

The ABM is defined under the assumption that proteins are represented by infinitesimal points, which move laterally across the plasma membrane, approximated by a 2D surface. In order to carry out the simulation, a target point pattern must first be provided – this may be taken from protein localisation data, as output by SMLM. Simulation parameters must also be given: these include a maximum permissible step size, D_{max}, a minimum permissible step size, D_{min}, an ROI size, and a time frame

over which to run the simulation (for guidance on parameter selection, see below). Protein motility is stochastic, with each agent moving in a random direction subject to a deterministic step size across each time frame. For each point in the cloud, the error between the localised K function and global target K function (**Figure 3.1a-c**) is determined, and then used to calculate the step size (for error and step size formulas, see **Appendix**). Each point is transposed in a random direction along a path of length equal to its step size (**Figure 3.1d-e**). If any point is transposed out of the bounds of the ROI, it will be reflected back at the angle of incidence with distance equal to the length of the path not yet traversed. This process is iterated over all time frames. See **Figure 7.3** (**Appendix**) for conceptual diagram of ASMODEUS and **Section 7.2.4** for algorithm pseudocode.

Parameter selection

The molecular counts, ROI size and K function are automatically determined from the input point pattern. The minimum and maximum molecular velocities are proportional to the expected diffusion coefficient of the protein embedded in the plasma membrane, which can be derived experimentally or mathematically[309]. Average diffusivity can be quantified from imaging methods such as single particle tracking or fluorescence correlation spectroscopy[108]. The expected step size Δx taken in time Δt can be estimated through the equation $\Delta x = 2(D\Delta t)^{1/2}$, where D is the diffusion coefficient[310].

Generalised PAD simulations

The ROI size was fixed at 2µm × 2µm, in accordance with a standard ROI size for super-resolution microscopy. A diffusion coefficient of $D \approx 0.1\mu m^2/s$ was selected –

this represents the diffusion coefficient of a T cell receptor (0.082μm^2/s, rounded to the nearest 10^{-1}), taken as a model molecular species[79,296]. Note that this framework could be applied to any protein with known diffusivity coefficient and TCR serves only as an exemplar here. This equated to a maximum step size of $\Delta x \approx$ 63nm per each 10ms time frame, and so $D_{max} =$ 63nm was selected. The average error across all points was recorded at each frame. The convergence time was defined as the first time frame in which the variance in the error of all future frames fell below 0.05. When applicable, DBSCAN cluster analysis was performed to extract cluster descriptors, for comparison with the original target parameters[198]. Changes in the point pattern distribution and global K function were tracked for all simulations (**Figure 3.2**). First, 100 simulations were run over 100000 frames and the maximum time taken to achieve convergence was measured at 7529 frames. As such, the maximum frame number for all future simulations was taken at double this value, rounded to the nearest 1000, to increase the probability of convergence. Therefore, unless otherwise specified, each simulation was run over 15000 frames, with each frame representing 10ms of real time, yielding a total simulated time of 2.5 minutes. All simulations converged within the allocated time.

Evaluating algorithmic performance

The error (see **Appendix** for formulas) serves as a measure of difference between the target and true Ripley's function, but does not necessarily determine whether the algorithm has recapitulated the cluster properties present in the target data. As such, we also record the difference in three cluster properties (cluster radius, number of clusters, and points per cluster) between target and simulated data at each discrete

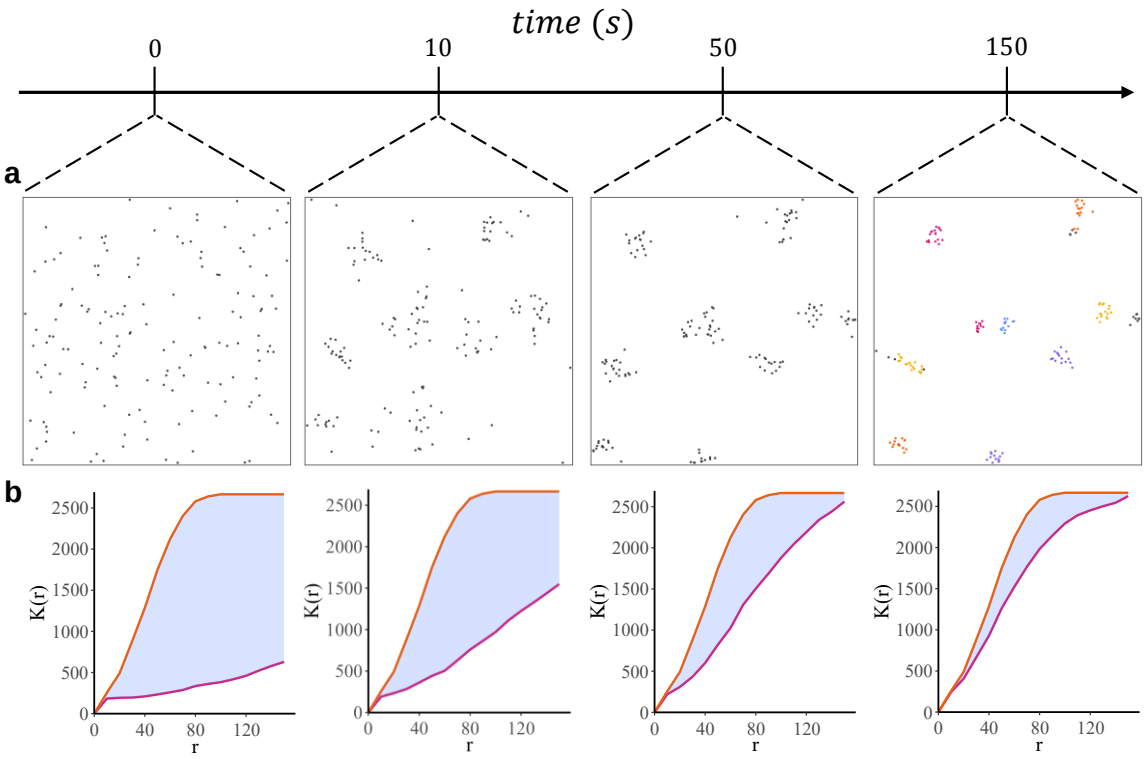


Figure 3.2: Example simulation. **a** Initial distribution starting at CSR gradually converges to a clustered distribution over 2.5 minutes. **b** Corresponding Ripley's K function of simulation (magenta) gradually converges to target K function (orange). Error (blue) decreases with time.

time frame. These properties are averaged over all simulations, then quantified as r_{dif}, n_{dif} and p_{dif}, respectively.

Modelling receptor nucleation

Multiple target distributions may be provided simultaneously – here, the error of each point is taken to be the minimum of the errors between all target functions. Given a cross-link radius r_c and number of proteins to recruit n_c, the following process is repeated for each nucleation site: select the densest point (that is, the point with most neighbours in a radius r_c) which has not yet been visited by the algorithm, this is chosen to be the central cross-linked protein. Then, determine the $n_c - 1$ closest

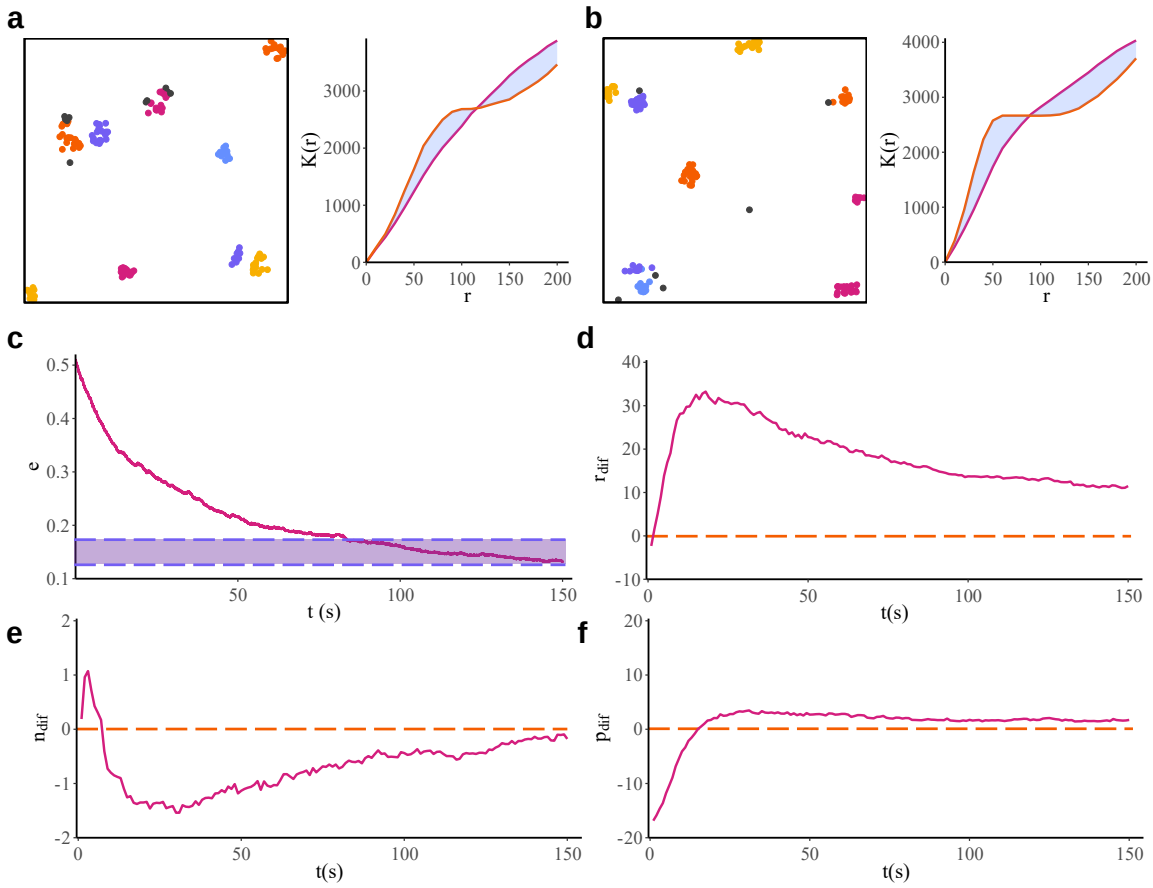


Figure 3.3: Convergence of cluster properties. **a-b** Simulations at convergence with pseudo-coloured clusters. Global Ripley's K function (magenta) approaches the target function (orange) in both cases. **c** Average ratiometric error between target and actual Ripley's K over time. Convergence is defined at the first point where there is less than 5% variability in the error (purple). Convergence of **d** cluster radius, **e** number of clusters and **f** points per cluster over time. Ground truth target is given in orange for all plots.

neighbours and move each to within an $r_c$nm radius of the central protein. Record each point used in the above step and exclude them from the creation of any further nucleation sites.

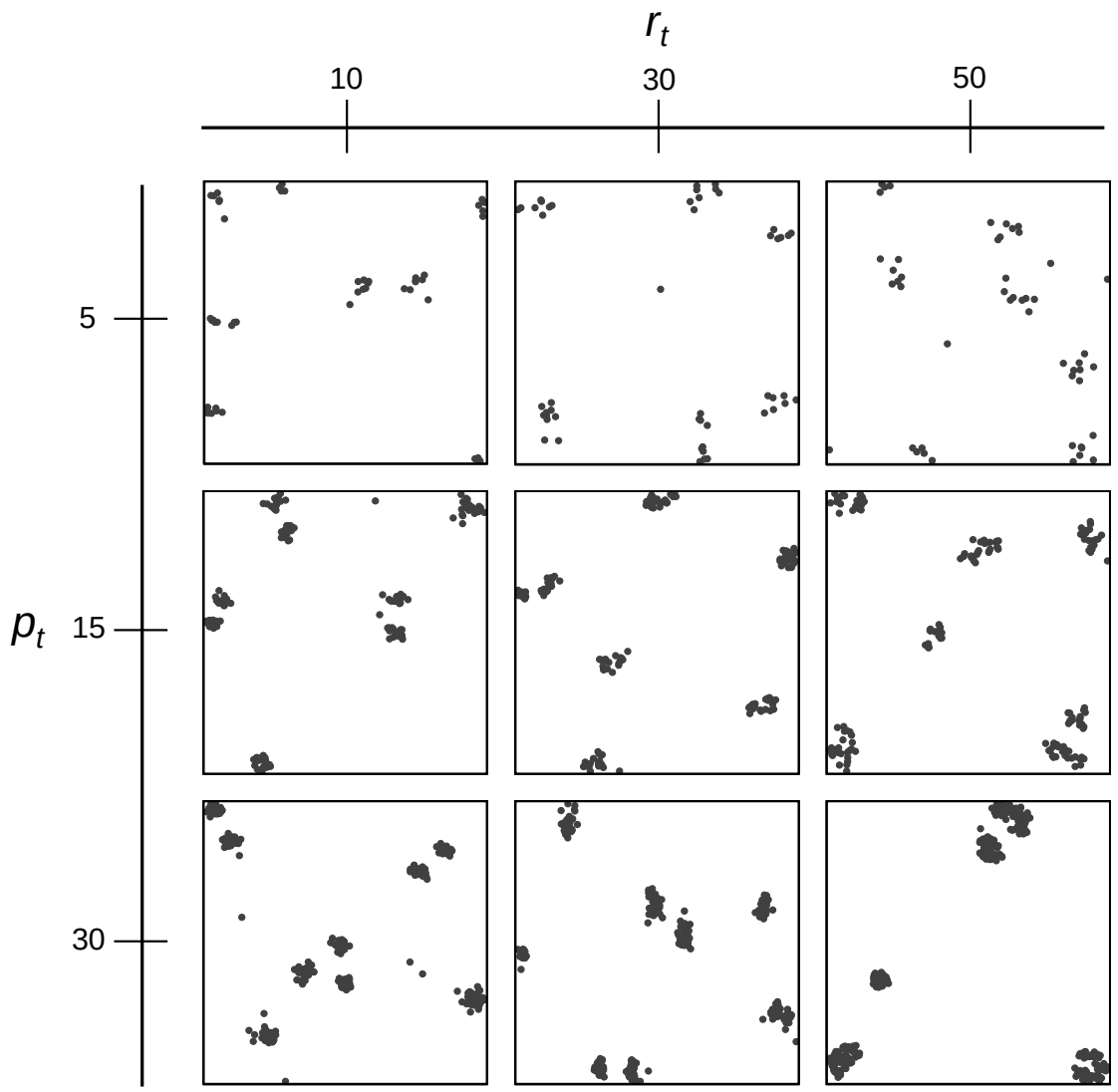


Figure 3.4: Simulation outputs under varying target cluster radii (r_t) and points per cluster (p_t). For each simulation, 10 clusters were simulated with a background to cluster ratio of 0.

Simulating multiple populations

Each simulation can incorporate an arbitrary number of distinct molecular species within the same ROI. Species may be simulated separately, by giving each population a separate target and ignoring species outside their own, or co-clustered with other populations, by giving a subset of the populations the same target and only considering the density of species within that subset. Interactions between populations

are recorded and tracked across consecutive frames. Here, we consider a three-protein, four-population model comprising activators, inhibitors and agents, which may be active or inactive (inhibited), depending on interactions with the first two molecular species. In particular, agents become active when within 5nm of an activator and inhibited when within 5nm of an inhibitor. All agents are initialised as inactive and the percentage of active agents is recorded over all time frames. In this context, activators may co-cluster with agents, or inhibitors may co-cluster with agents, but both do not occur simultaneously. By outputting the final point pattern of any simulation into another, any of the above regimes can be concatenated.

3.3 Results

3.3.1 Agent-based simulations conserve cluster properties

To validate model efficiency, a series of simulated point clouds, with varying cluster properties, were input to ASMODEUS. For each simulation, the change in error was recorded over time (**Figure 3.3c**). Further, the number or clusters, average cluster radius, and points per cluster, as defined by input parameters, were tracked (**Figure 3.3d-f**). All ABM simulations were initiated at CSR, with an error of 1.0 at time point 0 (see **Appendix**). The average error over all simulations reduced to a value of 0.163 at convergence, representing a 6-fold reduction, and all simulations reached convergence within the model runtime of 2.5 minutes. Furthermore, both the number of clusters and points per cluster fell to within 10% of their input target values (relative to peak deviation) at convergence, while the slope of the cluster radius depicts closer radii at greater runtime. Example simulations show that ASMODEUS can achieve a range of cluster radii and densities (**Figure 3.4**). However, caution should be taken for

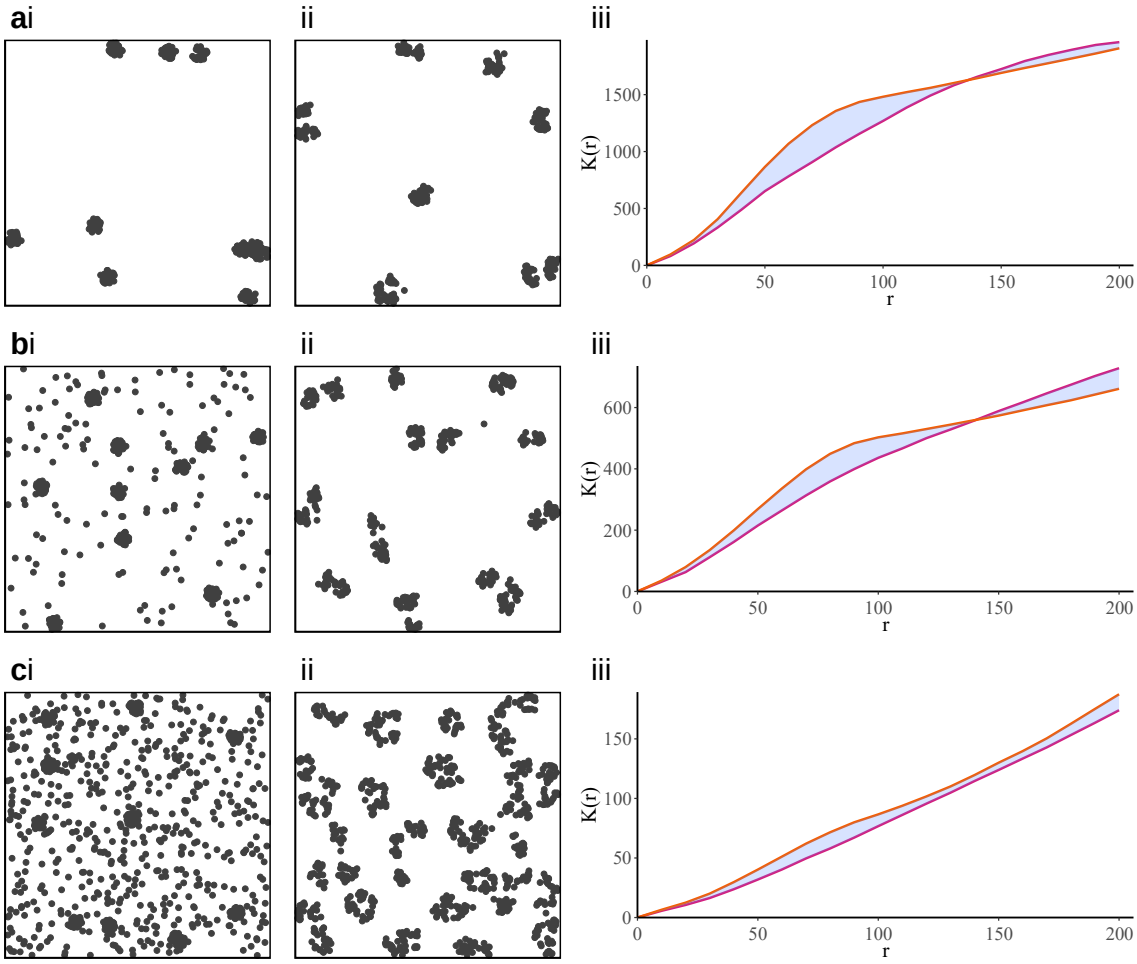


Figure 3.5: The impact of increasing background to cluster ratio on simulation output. For each simulation, the following panels are shown: (i) the target distribution, (ii) static plot of the simulation at convergence and (iii) difference between target (orange) and global (magenta) K function. Corresponding background to cluster ratios for **a-c** are 0.0, 0.33 and 0.66, respectively. In this case, the simulator favours local cluster density over distribution of outliers.

point patterns of particularly high background (**Figure 3.5**), as the model favours local point density.

It should be noted that variance or inconsistency among cluster properties in the target distribution, particularly regarding cluster radius and density, can produce undesirable simulation outputs which deviate from the target cluster properties. **Figure 3.6** depicts

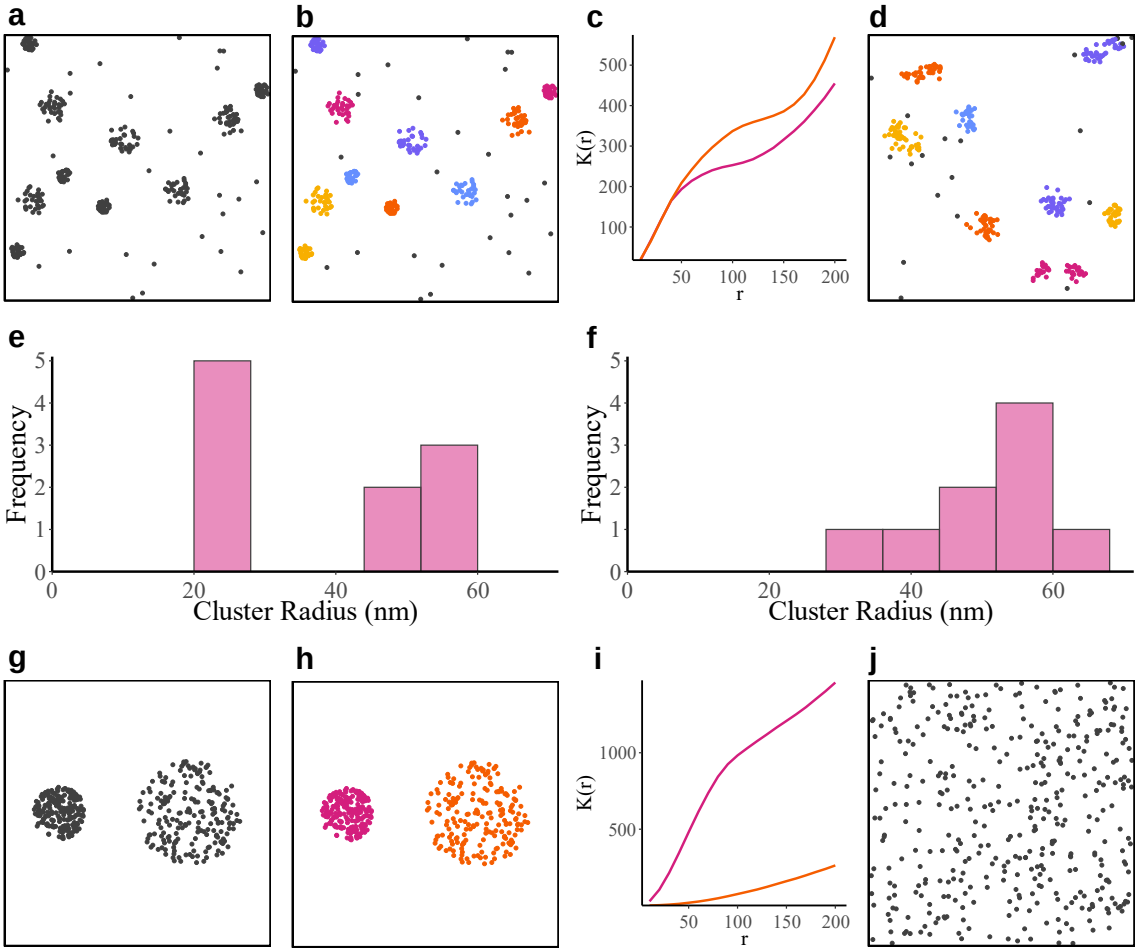


Figure 3.6: Exemplar failure cases for ASMODEUS. **a** Mixed cluster sizes. A distribution of 5 clusters of radius 25nm overlaid with 5 clusters of radius 50nm. All clusters contain 30 points. Background-to-cluster ratio is set to 0.1. **b** Clusters identified from the point cloud in **a** using DBSCAN cluster analysis with parameters $\epsilon = 50$nm and $minPts = 25$. **c** Global Ripley's K function (magenta) versus the target function (orange). **d** The final point cloud output by the simulation. **e** Histogram of cluster radii (half-max of all intra-cluster distances) from all clusters given in **b**. Cluster radii appear to aggregate around two local maxima at 25nm and 50nm, respectively. **f** Histogram of cluster radii from all clusters given in **d**. The distribution of radii is skewed towards larger cluster sizes and does not seem to show clear partitioning. All binsizes: 8nm. **g** Mixed cluster sizes with high density. A cluster with radius 50nm adjacent to a cluster with radius 100nm. Both clusters contain 200 points. Background-to-cluster ratio is set to 0.1. **h** Clusters identified from the point cloud in **g** using DBSCAN cluster analysis with parameters $\epsilon = 100$nm and $minPts = 100$. **i** Global

Ripley's K function (magenta) versus the target function (orange). **j** The final point cloud output by the simulation. ASMODEUS is unable to aggregate 200 points into a relatively small region of space (circles of ~50nm and ~100nm) within the simulation runtime, and so the distribution appears CSR.

some exemplar failure cases, the first of which is for a target distribution with mixed cluster sizes (**Figure 3.6a-b**) – half of which are 25nm, and the remainder 50nm. In spite of the global Ripley's K function appearing to converge to the target K function (**Figure 3.6c)**, the output distribution presents cluster radii which do not appear to be present in the target distribution. A histogram of recorded cluster radii (taken here as half the maximum distance between all points in the same cluster) show that the target distribution presents clusters with radii which seem to aggregate around two local maxima, corresponding to 25nm and 50nm respectively (**Figure 3.6e**). However, the histogram of simulated cluster radii (**Figure 3.6f**) seems to suggest a skew towards larger cluster sizes and does not show the same partitioning. The second example (**Figure 3.6g-j**) presents relatively dense clusters (200 points per cluster in two clusters of radius 50nm and 100nm). Here, ASMODEUS is unable to aggregate enough points within the expected cluster radii by the end of simulation runtime, and so the output distribution remains CSR.

3.3.2 Modelling receptor nucleation

Single-population simulations may be subjected to multiple target functions simultaneously, which permits convergence towards one of several possible distributions. Distributions may be perturbed by a sudden forced aggregation of points into nucleation sites (see **Materials and Methods**). For PAD simulations, the case of simultaneously targeting one CSR and one clustered target (**Figure 3.7a-b**) is of

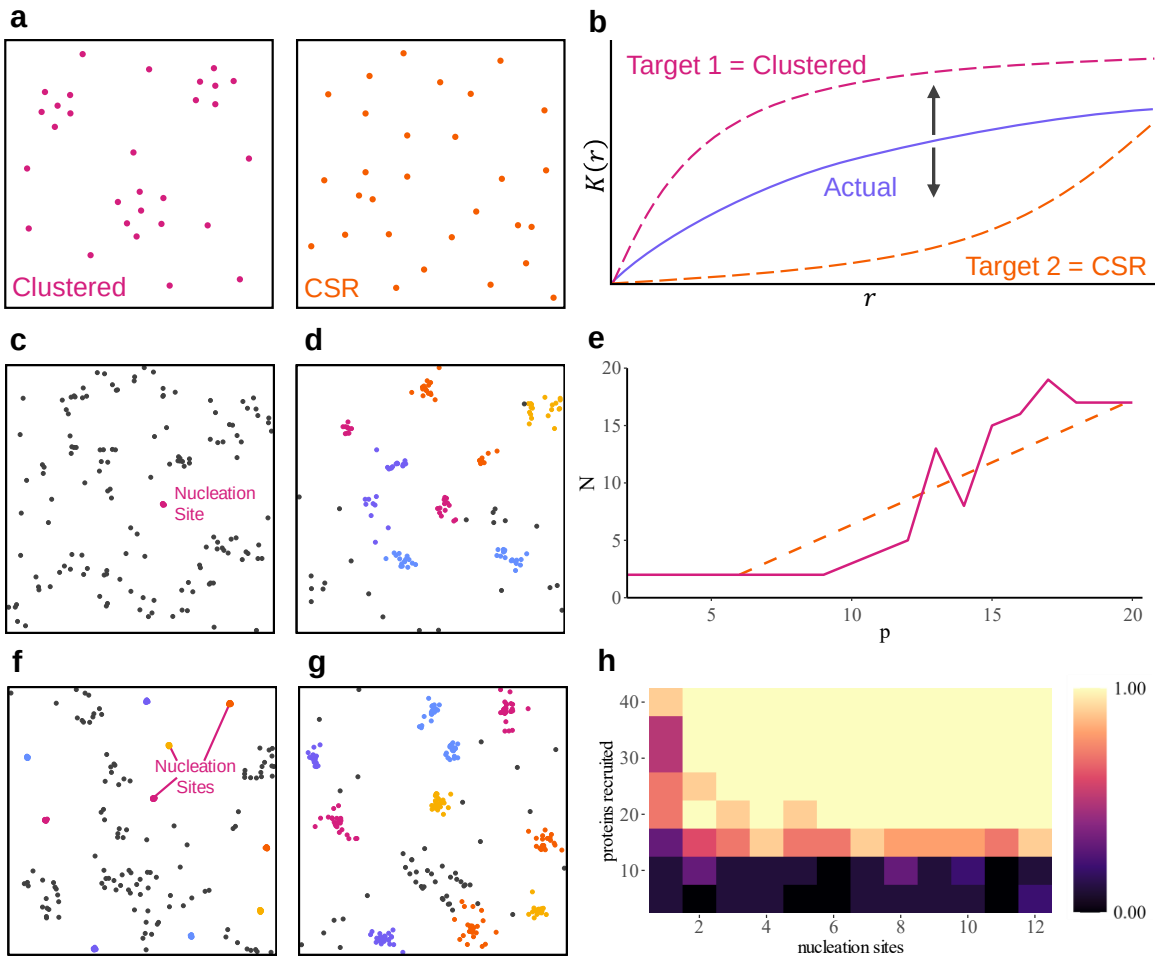


Figure 3.7: Modelling receptor nucleation. **a** ASMODEUS can map a distribution to multiple targets simultaneously. We consider a single population with two possible target distributions, one clustered (magenta) and one CSR (orange). **b** Multiple target functions are provided, and the target with lowest error is selected. **c** Formation of cross-linked cluster (nucleation site) in simulated data. **d** Global clustering induced as a result of cross-linking. **e** Cluster maximiser for the whole system (*N*) versus the number of points per cluster in the target distribution (*p*). Line of best fit is given in orange. **f** Formation of multiple nucleation sites in simulated data. **g** Global clustering induced as a result of nucleation. **h** Estimated probability of converging to clusters versus number of nucleation sites and number of recruited proteins. Simulated target distribution is defined by 10 clusters each containing 20 points.

particular interest, as a shift in receptor organisation to form nanoclusters can bring about signal digitisation[148] and dictate several important biological processes, such as formation of the immunological synapse[210]. We tested the impact of nucleation on system convergence by first inducing a single cross-linked cluster of proteins to each simulation (**Figure 3.7c-d**). After 1 second of simulation time, we randomly designate one agent to be a cross-linker and forcibly aggregate proteins into a cluster. Target distributions were generated such that each cluster contained a given number of points, *p*. We define the cluster maximiser, *N*, as the number of cross-linked proteins required to maximise the probability of the simulation converging to a clustered distribution, as opposed to CSR. We repeat 10 simulations for each condition and estimate the probability as the proportion of these simulations which converged to clusters.

Results suggest an approximately linear relationship between *p* and *N* (**Figure 3.7e**), suggesting that one cross-linker may be sufficient to induce clustering within the ROI and that the likelihood of global clustering increases as the number of proteins recruited increases. By incorporating several nucleation sites (**Figure 3.7f-g**), we can further increase the probability of convergence. Here, we track convergence against both the number of nucleation sites and proteins recruited on a target distribution defined by 10 clusters of 20 points (**Figure 3.7h**). We repeat these experiments for target distributions with varied cluster parameters (**Figure 3.8**), and find that the probability of convergence increases as protein recruitment increases, even at a relatively low number of nucleation sites.

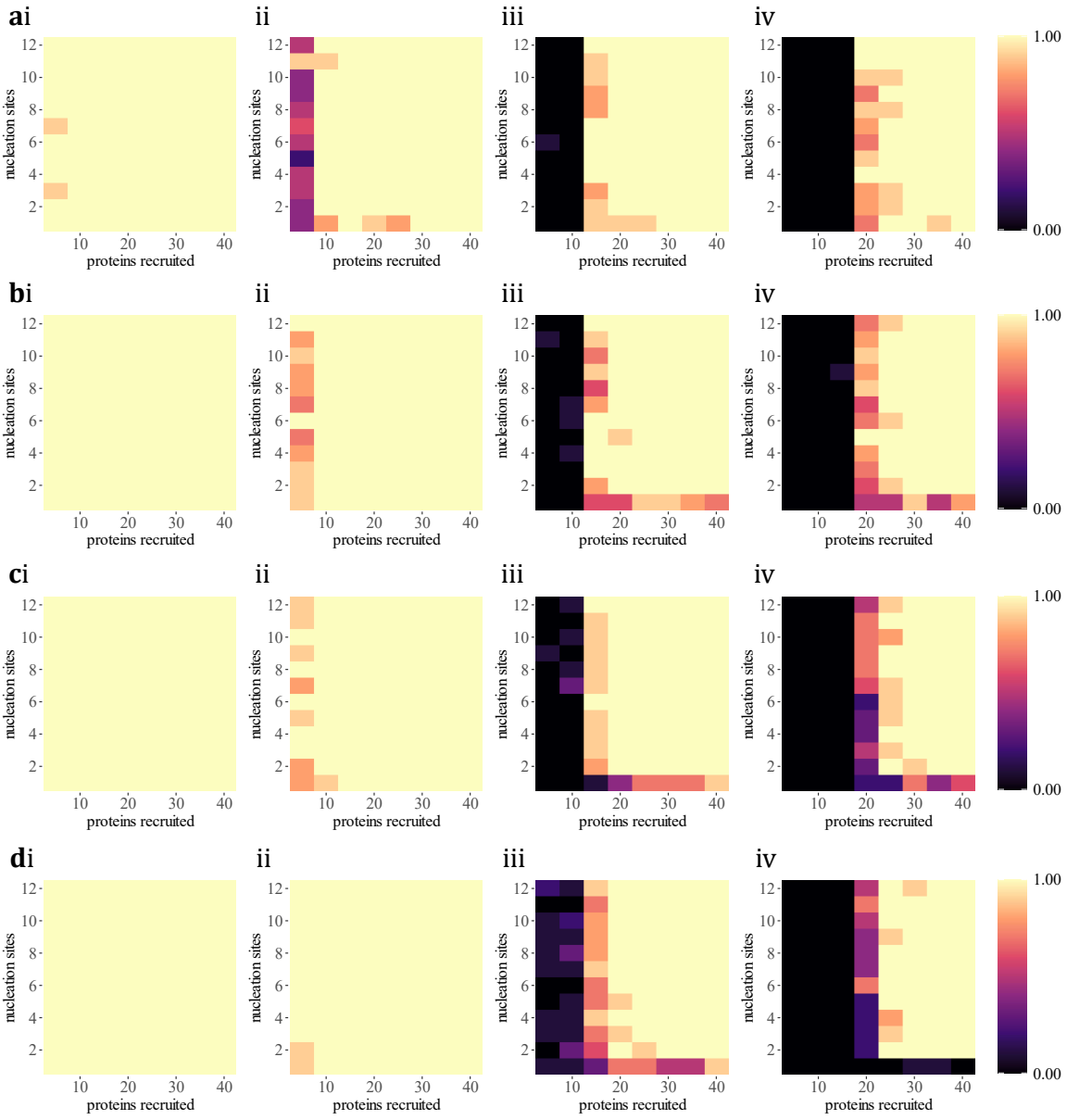


Figure 3.8: Estimated probabilities of converging to clusters under target distributions with **a** 4 clusters, **b** 6 clusters, **c** 8 clusters and **d** 10 clusters containing (i) 5, (ii) 10, (iii) 20 or (iv) 30 points per cluster.

3.3.3 Simulating CD3ζ distributions in pre- and post-activation T cells

To model nucleation of CD3ζ, we applied our methodology to point pattern data acquired from DNA-PAINT microscopy. Jurkat E6.1 T cells were fixed over functionalised lipid bilayers and immunostained with a primary antibody against CD3ζ.

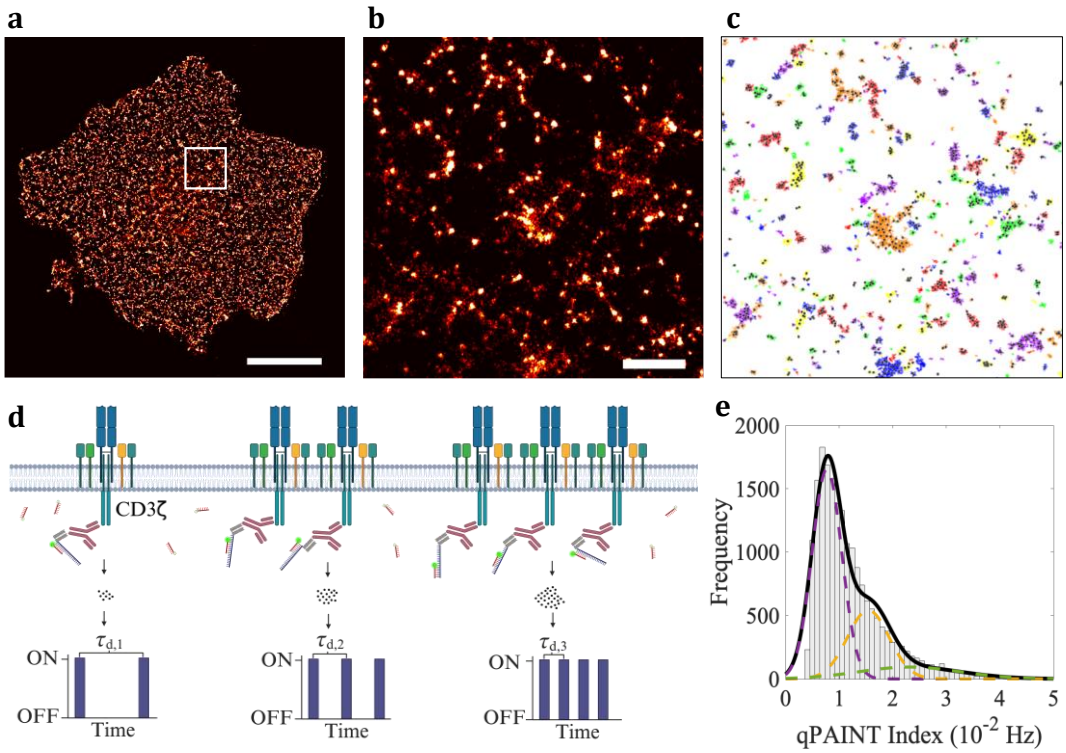


Figure 3.9: Overview of DNA-PAINT imaging and qPAINT analysis. **a** DNA-PAINT representative image of CD3ζ molecules in an activated Jurkat T cell. The white box indicates a zoom-in region to highlight subsequent cluster analysis and protein map conversion. Scale bar: 5µm **b** Zoom view of the selected ROI in previous panel. Scale bar: 500nm. **c** DBSCAN cluster analysis and protein map output for selected ROI. Each colour represents a different cluster selected by DBSCAN. Black dots on top of each cluster represent the position of CD3ζ monomers using k-means clustering, where k is equal to q_{i1}/q_i. **d** Overview of qPAINT analysis pipeline. T-cell receptors are stained with a primary antibody against CD3ζ and then with a DNA conjugated rabbit Fab as a secondary staining. The Cy3B-labeled imager strands bond transiently from solution to the complementary docking strands attached to the Fab. Lower panel shows schematic representation of the number of localisations and intensity traces for example clusters of 1, 2 and 3 CD3ζ molecules. These traces show characteristic fluorescence on- and off-times (τ_d) depending on the amount of docking strands in each cluster. The frequency of the imagers binding to their docking strand scales linearly with the number of docking strands, and this is the principle of qPAINT. When the number of monomers in the cluster increases, the dark time decreases. **e** Histogram of internal calibration qPAINT indexes per cluster pooled from all data

samples fit with a multi-Gaussian function. This was achieved by selecting very small clusters in the biological data set. This behaviour is consistent with the presence of 1, 2, or 3 DNA docking strand units within a cluster of localisations. Dark line represents the 3-Gaussian fit exhibiting peaks at multiples of qPAINT index of 0.0075 Hz. Each dash line represents a Gaussian centered at 1, 2 and 3 times the qPAINT index (see **Materials and Methods**).

Secondary staining was performed immediately after with DNA conjugated Rabbit Fab and imaged on a custom-built microscope (**Figure 3.9**). Image reconstruction, post-processing and quantitative analyses were undertaken in Picasso and MATLAB (see **Materials and Methods**). ROIs for non-activated (**Figure 3.10a**) and activated (**Figure 3.10b**) conditions were segmented and the target K functions were averaged across each condition (**Figure 3.10c-d**). Simulations were run over a range of nucleation parameters with 10 repeats for each condition, yielding 9600 simulations in total (**Figure 3.10e-f**). **Figure 3.10i** suggests that nucleation parameters could be optimised to increase the likelihood of inducing global clustering. Generally, the greater the number of nucleation sites and proteins recruited, the higher the probability of simulating T cell activation. As in the simulated data, the activation outcome depends most significantly on the number of proteins recruited, with at least 20 proteins required to maximise the probability in most conditions.

3.3.4 Modelling interacting molecular species

Building on the single-population model, we developed a system for simultaneously modelling several interacting molecular species (see **Materials and Methods**). The distributions of each underlying population may be aggregated separately (**Figure 3.11a**) or co-clustered (**Figure 3.11b**). A four-population model was proposed, including activators, inhibitors and agents (which may be active or inactive). Agents

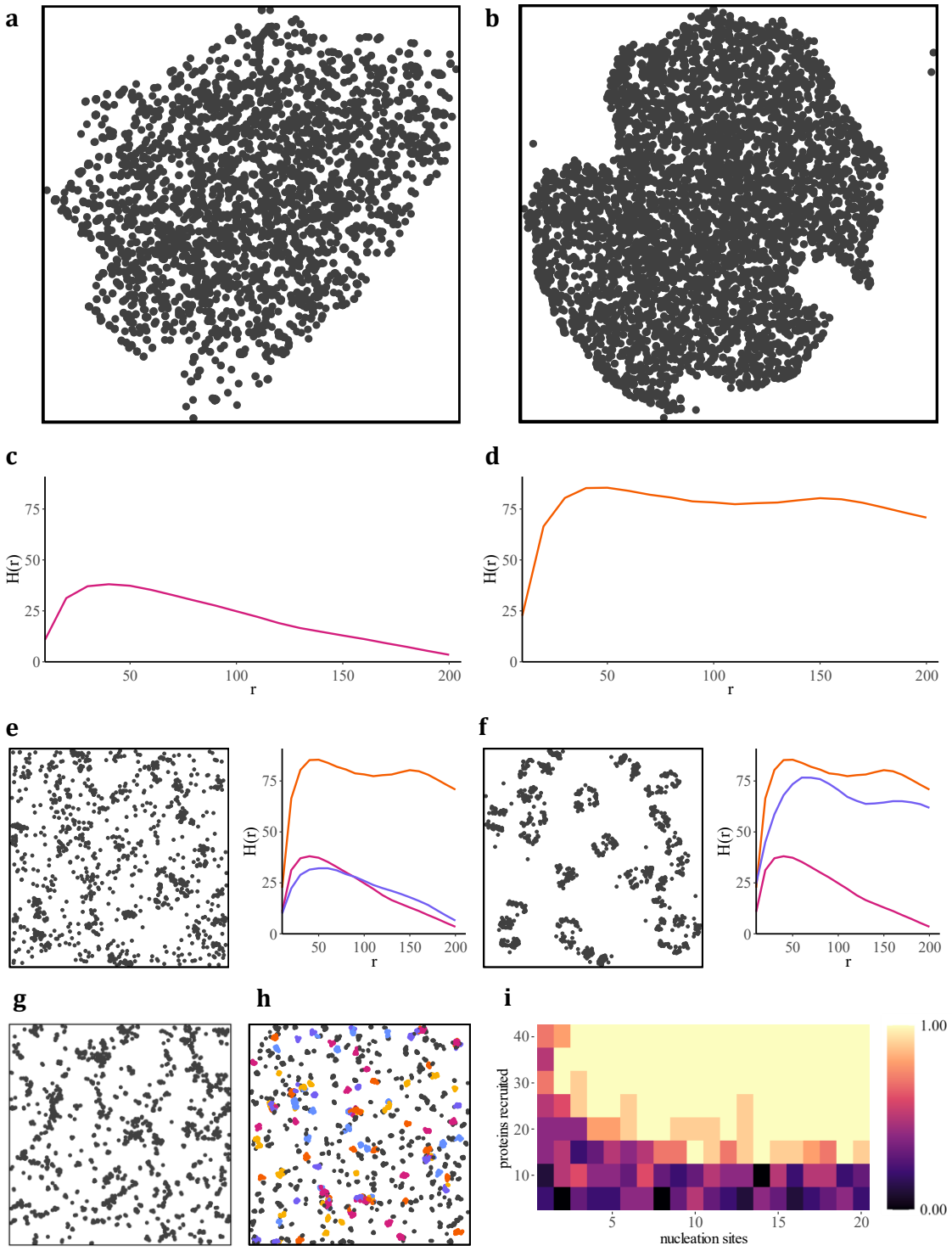


Figure 3.10: Simulated T cell activation via induced nucleation. Protein maps of CD3ζ distributions were acquired from **a** non-activated and **b** activated Jurkat T cells (dimensions: 30μm by 30μm). Ripley's H functions taken from CD3ζ distributions in **c** non-activated and **d** activated conditions. **e** A point pattern was first fitted to a non-activated target. The simulated H function (purple) can fit to either the non-activated

(magenta) or activated (orange) target functions. Nucleation was induced after 10s. **f** Resulting distribution in final simulation frame. Depending on the degree of nucleation, the point pattern will shift to the activated target and induce pronounced global clustering. The simulated H function now aligns with the activated target. A 3µm by 3µm ROI shows CD3ζ distribution on Jurkat T-cell plasma membrane **g** before activation and **h** after activation. Denser clusters are highlighted in the post-activation ROI. **i** Probability of inducing whole system aggregation versus number of nucleation sites and number of recruited proteins.

become activated or inhibited when within a pre-assigned range of activators or inhibitors respectively – here, the range is taken to be 5nm (**Figure 3.11c**). Populations may be simulated under a co-clustered regime (**Figure 3.11d**), a CSR regime (**Figure 3.11e**) or an isolated cluster regime (**Figure 3.11f**). We record the change in proportion of active agents across the simulation (**Figure 3.11g**).

3.3.5 Simulating varied dynamics across multiple populations

To validate the impact of co-clustering on signal digitisation, we emulate an activator-inhibitor-agent model, in which activators are first clustered independently of inhibitors and agents, which are themselves co-clustered (**Figure 3.12a**). Halfway through the simulation, the regime was switched, and activators were co-clustered with agents while inhibitors were left isolated (**Figure 3.12b**). Target distributions were kept consistent throughout analysis (see **Materials and Methods**) the percentage of activated agents was tracked across all simulation frames. Only the step size parameter, D_{max}, was varied between simulations (**Figure 3.12c-d**). Results suggest that simulations with lower step sizes consistently show a greater proportion of activated agents prior to the regime switch, rather than after, when compared to simulations with larger step sizes (**Figure 3.12e**). In addition, a shift in the percentage

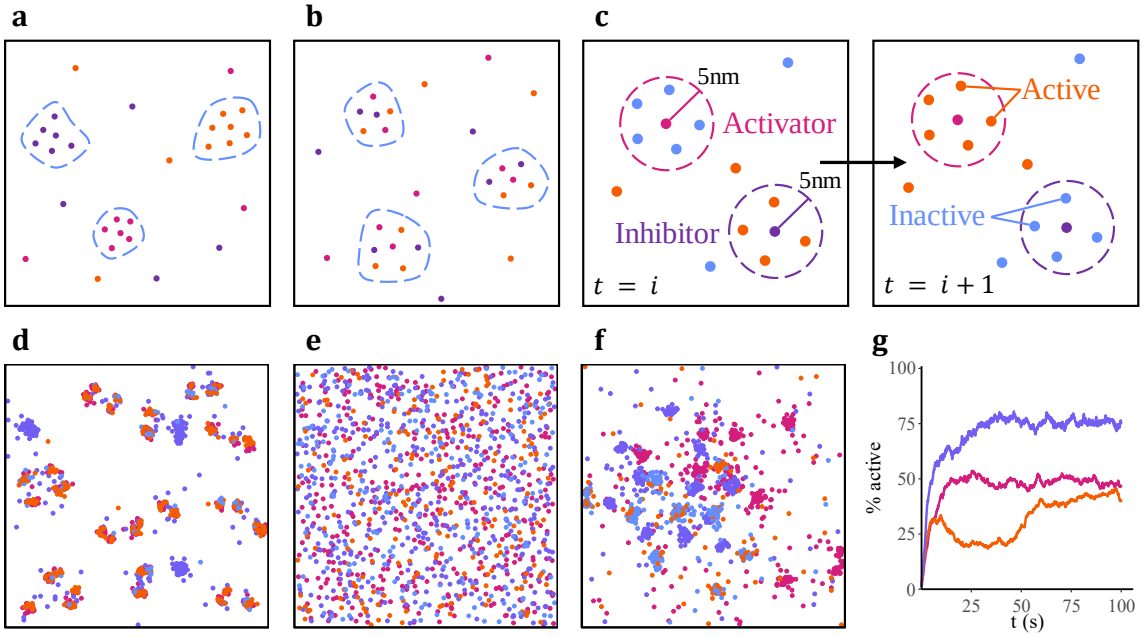


Figure 3.11: Multiple population modelling. Each population can be simulated **a** separately or **b** co-clustered with other populations. **c** Interactions between populations can be recorded and tracked across consecutive frames. We consider a four-population model, including activators, inhibitors and agents (active or inactive). Agents are activated or inhibited within 5nm of activators or inhibitors respectively. Example cluster variations for multiple species: **d** a co-clustered regime **e** a CSR regime and **f** an isolated cluster regime. **g** The percentage of activated agents in the co-clustered regime (purple) consistently exceeds the percentage activated in the CSR (magenta) and isolated cluster case (orange), even with the same population sizes.

of activated agents can be observed at the halfway point, when the regime switch occurs.

3.4 Discussion

In this work, we introduced ASMODEUS, an agent-based model for simulating dynamics of transmembrane proteins and generating toy protein maps, which could be used to simulate SMLM data. We applied this framework to re-deriving simulated

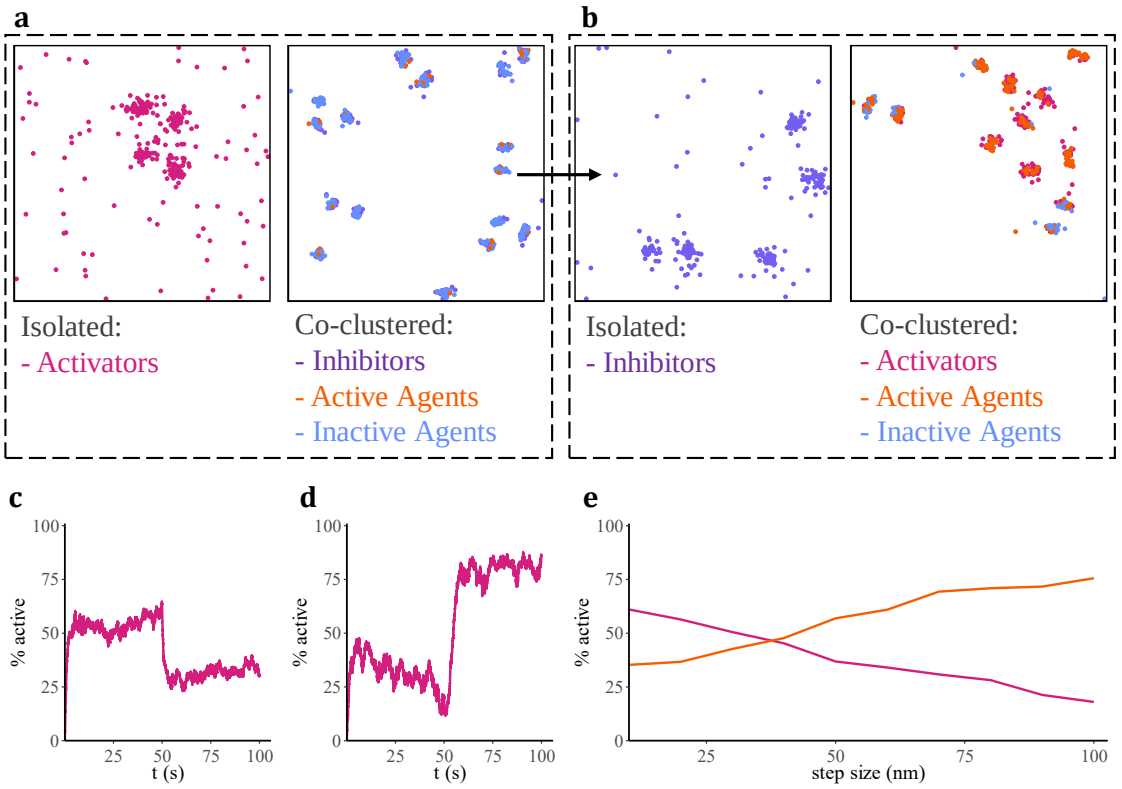


Figure 3.12: Proof-of-concept for modelling PAD with multiple molecular species. **a** Activators are first simulated independently of inhibitors and agents, which are co-clustered. **b** Halfway through the simulation, the regime is switched, and activators are co-clustered with agents, while inhibitors are considered separately. The percentages of activated proteins are recorded over time. **c** Example case with step size of 20, in which activation is more prevalent prior to switch. **d** Example case with high step size of 60, in which activation is more prevalent after switch. **e** Average percentage activated before (magenta) and after (orange) switch across a range of step sizes.

distributions with known ground truth clusters and extended this model to incorporate several interacting molecular species. Furthermore, we presented a proof-of-concept for the use of ASMODEUS in simulating the receptor oligomerisation or cross-linking required to induce a phase shift in protein distribution.

3.4.1 Simulating static protein maps with ASMODEUS

ASMODEUS was tested on a series of simulated point clouds with varying cluster properties. All simulations reached convergence within the model runtime of 2.5 minutes and, on average, the error showed a 6-fold reduction at convergence. The number of clusters and points per cluster fell to within 10% of their target values. The slope of the cluster radius shows slower convergence, which may improve at greater runtime. This suggests that an agent-based modelling approach, using the Ripley's K function as target for reinforcement learning, may suffice for simulating cluster affinity in point patterns and can induce aggregation with pre-defined cluster properties. However, ASMODEUS may not be suitable for point patterns of particularly high background density, as the model favours local point density over global point density and may unexpectedly increase the total number of clusters. For target distributions with mixed cluster properties, ASMODEUS may output unpredictable cluster geometries, despite converging to the K function, or fail to converge altogether. This suggests that ASMODEUS may perform best on target data with consistent cluster properties.

Since system dynamics depend on pairwise interactions between agents, the computational complexity of running a simulation increases quadratically as the number of agents increases linearly[311]. That said, the software presented here has been computationally optimised and is integrated in a compiled programming language where possible. As a reference, a simulation of 500 points running over simulated 150 seconds takes no more than 5 real-time minutes to complete on a single processor.

3.4.2 Modelling signal digitisation and molecular interactions with ASMODEUS

If a simulation is provided with two possible targets – one of which represents a clustered distribution and the other CSR – then the PAD model may converge to either distribution and enter into a steady state. If the dynamic distribution has converged towards the CSR target, in which receptor density is approximately homogeneous, then we have shown here that a small perturbation to the density distribution, by spontaneous clustering of agents, may force the distribution out of its steady state. Depending on the number of nucleation sites or proteins per nucleation site, the PAD model may undergo a phase shift and converge towards the clustered target, in which receptor density is heterogeneous. By repeating simulations, we are able to estimate the probability of convergence to the clustered target after nucleation. Therefore, ASMODEUS may provide a framework for estimating the optimal nucleation parameters required to induce receptor clustering in dynamic simulations. In applications to TCR-CD3 distributions from DNA-PAINT, we find that receptor nucleation selectively elicits a phase shift – this depends on the number of nucleation sites and proteins recruited, but generally requires each nucleation site to recruit 20 proteins. This could inform choices for therapeutic mediators which induce or disrupt TCR aggregation in immunotherapy or the treatment of autoimmune disorders – however, this remains to be validated *in vitro*.

In multiple population models, we determine that larger step sizes promote increased activation after inducing a change to receptor co-clustering. In addition, we observe a sudden shift in the percentage of activated agents after the clustering regime switch occurs. Primarily, this acts as a proof-of-concept for ABMs in modelling dynamic, interacting protein species – which, to the best of our knowledge, is the first of its kind. Results may suggest that changes to molecular organisation between members of

interacting species can bring about a phase shift in system-wide activation. With a small change in organisation rules, governed by the degree of membrane order, a digital response can be induced in simulations. This approach may help estimate and quantify the effect of interacting species distributions, which may be used as a predictor of digitisation under variable clustering regimes.

3.4.3 Concluding remarks

Traditional mathematical models are derived from systems of continuous differential equations, which aim to capture biophysical interactions and biochemical kinetics through geometric constraints and parameterisation. Despite the capabilities of modern microscopy techniques, these parameters and hyperparameters are difficult to determine in the context of transmembrane protein dynamics[300]. Point pattern simulation typically requires knowledge of an expected point distribution, derived from constraints on pattern geometry. Agent-based approaches are modular, easy to aggregate and designed to predict dynamic system behaviour under a set of postulated mechanisms[298]. As such, ABMs are well-suited to capturing the biophysical processes that bring about macro-scale responses, such as nucleation, digitisation, or activation[312,313]. In this work, we have presented an agent-based modelling approach for simulating the dynamics of transmembrane proteins and generating realistic toy SMLM data. The model can integrate several interacting molecular species and introduce perturbations to serve as proxies for natural and induced biophysical disruptions. This platform could be used to determine the nucleation conditions required to promote clustering of specific transmembrane proteins, and the simulations derived may provide a system for evaluating and comparing emergent phenomena in PAD ABMs. Notably, we have used our method to estimate the biophysical parameters

of receptor nucleation required to maximise the probability of TCR oligomerisation. As such, ASMODEUS could be used to simulate static protein maps for validating analytic techniques, to model the impact of perturbations on PAD systems, or to offer insight into the design of therapeutics which modify or induce receptor clustering[314-316]. In the following chapter, we use ASMODEUS as an SMLM data simulator to generate spatial point distributions, which are overlaid with marked point patterns. This provides ground truth data for testing and validating topological data analysis techniques for partitioning coordinate-based MPP data.

Chapter 4: Marked point pattern analysis

The focus of this chapter is on the development of two algorithms for identifying and extracting topological features from marked point pattern data. The first algorithm, P-Check, determines whether there is statistically significant evidence of non-random colocalisation in discrete MPPs. The second algorithm, JOSEPH, partitions continuous marked point pattern data into clusters defined by both a given spatial scale and a quantifiable distribution of marked values. By performing this analysis on MPPs from SMLM with polarity-sensitive probes, these tools may determine if and where lipid domains exist in ROIs from membrane data.

Contributions

Contributing authors

L. Panconi[1,2,3], J. Euchner[3,4,5], M. Makarova[3,6,7], D-P. Herten[3,4,5], D.M. Owen[1,3,8] & D.J. Nieves[1,3]

Affiliations

[1]Institute of Immunology and Immunotherapy, University of Birmingham, Birmingham, UK.

[2]College of Engineering and Physical Sciences, University of Birmingham, Birmingham, UK.

[3]Centre of Membrane Proteins and Receptors (COMPARE), University of Birmingham, Birmingham, UK.

[4]Institute of Cardiovascular Sciences, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK.

[5]School of Chemistry, College of Engineering and Physical Sciences, University of Birmingham, Birmingham, UK.

[6]School of Biosciences, College of Life and Environmental Science, University of Birmingham, Birmingham, UK.

[7]Institute of Metabolism and Systems Research, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK.

[8]School of Mathematics, College of Engineering and Physical Sciences, University of Birmingham, Birmingham, UK. &

Author contributions

L. P., D. M. O. and D. J. N. conceptualised the project and methodology. L. P. wrote simulation and PLASMA analysis code, performed simulations and data analysis. J. E. built the optical set-up, acquired di-4-ANEPPDHQ PAINT data, and performed ratiometric analyses (**Figures 4.6** and **4.10**). M. M. prepared GUVs for di-4-ANEPPDHQ PAINT. D-P. H. contributed ideas and concepts. D. J. N. developed the di-4-ANEPPDHQ approach and optimised generation of membrane patches (**Figures 4.6** and **4.10**).

4.1 Introduction

Partitioning of the plasma membrane into lipid-ordered and lipid-disordered phase regions is hypothesised to play an essential role in regulating cell signalling processes[317]. This phase separation may be characterised by nanoscale regions of the membrane, denoted here as lipid domains or nanodomains, which are locally homogeneous in lipid packing[317]. However, while microscale ordered lipid domains can be readily observed in synthetic bilayers, there is still ongoing debate about the

existence and nature of nanodomains in mammalian cell membranes[318]. Super-resolution methods can localise fluorescent molecules at the nanoscale and may be adapted to work with environmentally-sensitive membrane probes[319-321]. By using PAINT microscopy with polarity-sensitive probe di-4-ANEPPDHQ, we resolve marked point patterns of nanoscale GP distributions in live cells and synthetic membranes. However, the precise geometry of lipid nanodomains is not yet known and may vary across data sets.

As such, we have developed a TDA software package, PLASMA (Point Label Analysis of Super-resolved Marked Attributes), designed for identifying underlying domains in marked point pattern data. The first goal of our analysis is to determine whether underlying domains, defined by non-random subsets of data points with approximately equal marks, are present within the data. For this, we have developed an algorithm, P-Check, to determine if a point pattern expresses spatial heterogeneity in binary marked points, which may suggest the presence of domains. Although P-Check is suitable for detecting nanodomains, it does not offer any information as to their location within the ROI. To expand upon this, we developed a data segmentation approach, termed Justification of Separation by Employed Persistent Homology (JOSEPH). In JOSEPH, persistent homology is used to construct clusters in which points colocalise spatially and whose marks lie within a quantifiable, pre-determined range. In this context, P-Check is first used to determine the presence of membrane lipid nanodomains in artificial membranes and live cells, and then JOSEPH is used to partition the set of localisations and pinpoint nanodomains within the ROI. By leveraging this analysis over multi-modal SMLM with environmentally-sensitive fluorophores, we achieve nanometre-resolution mapping of membrane properties and measure changes in response to external perturbation with methyl-β-cyclodextrin

(MβCD). This methodology may present a useful tool for processing MPPs and identifying nanodomains in SMLM data.

4.2 Materials and methods

Rat mammary fibroblast cell culture and sample preparation

Rat mammary fibroblast (RAMA27) cells were gifted by Prof. David G. Fernig (University of Liverpool, UK). Cell culture was performed as in literature[322]. RAMA27 cells were cultured in phenol red free DMEM supplemented with 10% PBS, 50 ng/mL insulin and 50 ng/mL hydrocortisone (all from Life Technologies) at 37°C in 10% CO_2. 24h prior to imaging, cells were trypsinized, passaged and seeded onto Ibidi µ-slide 8-well glass-bottomed chambers pre-coated with 1 µg/mL fibronectin (Sigma-Aldrich), at 20,000 cells per well. Cells were left overnight in culture medium to adhere and spread. Samples were incubated for 30 minutes with 80 nM di-4-ANEPPDHQ in phenol red free DMEM to permit binding of single di-4-ANEPPDHQ molecules. Eight cells were imaged in total.

RAMA27 treatment with methyl-β-cyclodextrin

Seeded cells were exposed to 15 mM methyl-β-cyclodextrin (Sigma Aldrich, cat.no: C4555) in culture medium without FBS for 30 minutes. Cells were then washed 3 times with 1x PBS to remove MβCD and prepared for imaging.

Giant unilamellar vesicle (GUV) preparation

GUVs were prepared by electroformation, as described in literature[323]. Lipid film was formed by deposing chloroform solution of either DOPC (Avanti Polar Lipids) or a mixture of DPPC (Avanti Polar Lipids) with 30% cholesterol (Sigma-Aldrich) onto

indium tin oxide (ITO)-coated glass slides, such that 100µg of lipid was deposed in total. Lipid films were then dried under an air hood for 4h. GUVs were formed in a 200 mM sucrose solution at 50°C using 11 Hz, 1V of alternating electric current for 2h.

Artificial membrane patch preparation

A 1 in 200 dilution of GUVs was made in 1X PBS and added to an Ibidi µ-slide 8-well glass-bottomed chamber. GUVs settled in solution and burst in contact with the glass bottom under osmotic pressure of the 1X PBS, forming a patch of lipid bilayer. Samples were incubated for 30 minutes with 20 nM di-4-ANEPPDHQ in 1X PBS buffer to permit binding of single di-4-ANEPPDHQ molecules.

Ratiometric PAINT imaging with di-4-ANEPPDHQ

Imaging of artificial membranes and RAMA27 cells was undertaken on a custom-built fluorescence microscope (RAMM system, ASI) using a 100× 1.49 NA oil immersion objective (Apo TIRF; Nikon), using a beam shaper for homogeneous illumination (piShaper, AdlOptica). A 488nm laser excitation under highly inclined and laminated optical sheet (HILO) illumination and z-focus lock (CRISP, ASI). Fluorescence emission from binding of di-4-ANEPPDHQ to membranes was spectrally split by a LP640 dichroic mirror (FF640-FDi02-t3, Semrock) and imaged with two sCMOS cameras (Prime 95B, Photometrics) over a 130µm × 130µm field of view. Hardware was controlled using µManager 2.0 with additional custom microcontroller boards. 20000 frames were taken per acquisition with 50ms integration time and a laser power density of 50W/cm^2 over the sample. Emission fluorescence of < 640nm was filtered using a custom bandpass filter (552/96). All image stacks were retained for further analyses.

Post-processing and ratiometric analysis for di-4-ANEPPDHQ data

Image stacks were combined and single molecule localisations were fitted via a Gaussian PSF model. All post-processing prior to GP calculation was undertaken in SMAP/MATLAB[131]. Localisations within 50nm and 1 off-frame were grouped and then filtered by precision (0-50nm), PSF size (100-300nm), PSF asymmetry (0-0.2) and frame (400-Inf). The projective transformation matrix was calculated between channels. Image stacks were re-analysed using a Gaussian PSF fitting to assign emission intensities across both channels. Localisations were then filtered by photon count in both channels (50-5000 per ungrouped localisation and 100-5000 photons per grouped localisation). Localisations from the second channel were then transformed onto the first channel and corrected for drift. Finally, the GP value for each grouped localisation was recorded (see **Appendix** for formula).

Marked point pattern simulation

2000 spatial point patterns were simulated via agent-based modelling, under the ASMODEUS R package v. 1.0.0 (see **Chapter 3**), and ToMATo cluster analysis was performed, under the RSMLM R package (v. 1.0.0)[199,221]. For each ROI, the convex hulls of all clusters were calculated and designated as domains. The ROI was discretised into square bins, with each bin comprising either a domain or part of the background. $\lceil np \rceil$ points were randomly allocated to a domain bin and their positions were distributed uniformly across that bin (for parameter choices, see **Table 4.1**). All remaining $n - \lceil np \rceil$ points were uniformly distributed across the background. For CSR distributions, all n points were uniformly distributed, irrespective of domain placement. Each point was then assigned a mark value determined from one of two Normal

distributions, depending on whether they fell into a domain or the background[324]. Normal distribution parameters (described in **Table 4.1**) were selected randomly via Latin Hypercube Sampling[325]. 1000 additional CSR data sets were then generated with marked values selected as above. The ROI used in each simulation was 3μm × 3μm in size and overlaid with convex domains. Marked values were assigned so that the average value of all points inside any domain was positive, and the average outside all domains was negative.

Parameter	**Meaning**	**Min**	**Max**
R	Target domain radius (assuming circularity).	20	250
μ_1	Mean mark value in simulated domains.	0.01	1
μ_2	Mean mark value outside simulated domains.	-1	-0.01
σ_1	Standard deviation of marks in simulated domains.	0.01	0.5
σ_2	Standard deviation of marks outside simulated domains.	0.01	0.5
p_d	Proportion of points assigned to domains.	0.2	0.8
n	Number of points in marked point pattern.	100	5000

Table 4.1: Input parameters for simulated data sets and the range of values they were randomly generated from.

P-Check

We take as input a discrete marked point pattern – that is, a point pattern with marks assigned from a finite number of categories. For each point, construct a neighbourhood comprised of all other points within the chosen search radius (r), record the total number of neighbours (within the radius r) and the number of

neighbours which have the same mark as the point itself (**Figure 4.1a-c**). Summing these values over all points, we determine the weighted proportion, P_k, which serves as an estimate for the probability that a randomly-selected point and any randomly-selected neighbour share the same label (see **Appendix** for formula). The value of P_k is maximised if points belonging to the same category form distinct, spatial clusters – specifically, if there is no mixing between points of different categories, and clusters of each category are spatially separated. By convention, we denote the weighted proportion of the original marked point pattern as P_0. A permutation test is then performed over N trials by randomly shuffling the categorical marks of all points and recalculating the proportion P_k for each trial $k = 1,2, \dots, N$ (**Figure 4.1d**). If P_0 falls within the top $\alpha\%$ of all P_k values, then the significance level can be estimated as $\alpha\%$ (**Figure 4.1e)**. This suggests statistically significant evidence, at the $\alpha\%$ level, of non-random colocalisation of identically-marked points, which may correspond to specific domains[326]. See **Figure 7.4** (**Appendix**) for conceptual diagram of P-Check and **Section 7.2.5** for algorithm pseudocode. A search radius may be automatically estimated from the data (see below).

Justification of Separation by Employed Persistent Homology (JOSEPH)

Take a marked point pattern as input. For each point, a neighbourhood comprised of all other points within the chosen search radius, r, is constructed (**Figure 4.2a-b**). Each point is designated a similarity value defined as the difference between the point's own mark and the average mark of its neighbourhood. This value is normalised so that points which are most like their neighbourhood are given a similarity value

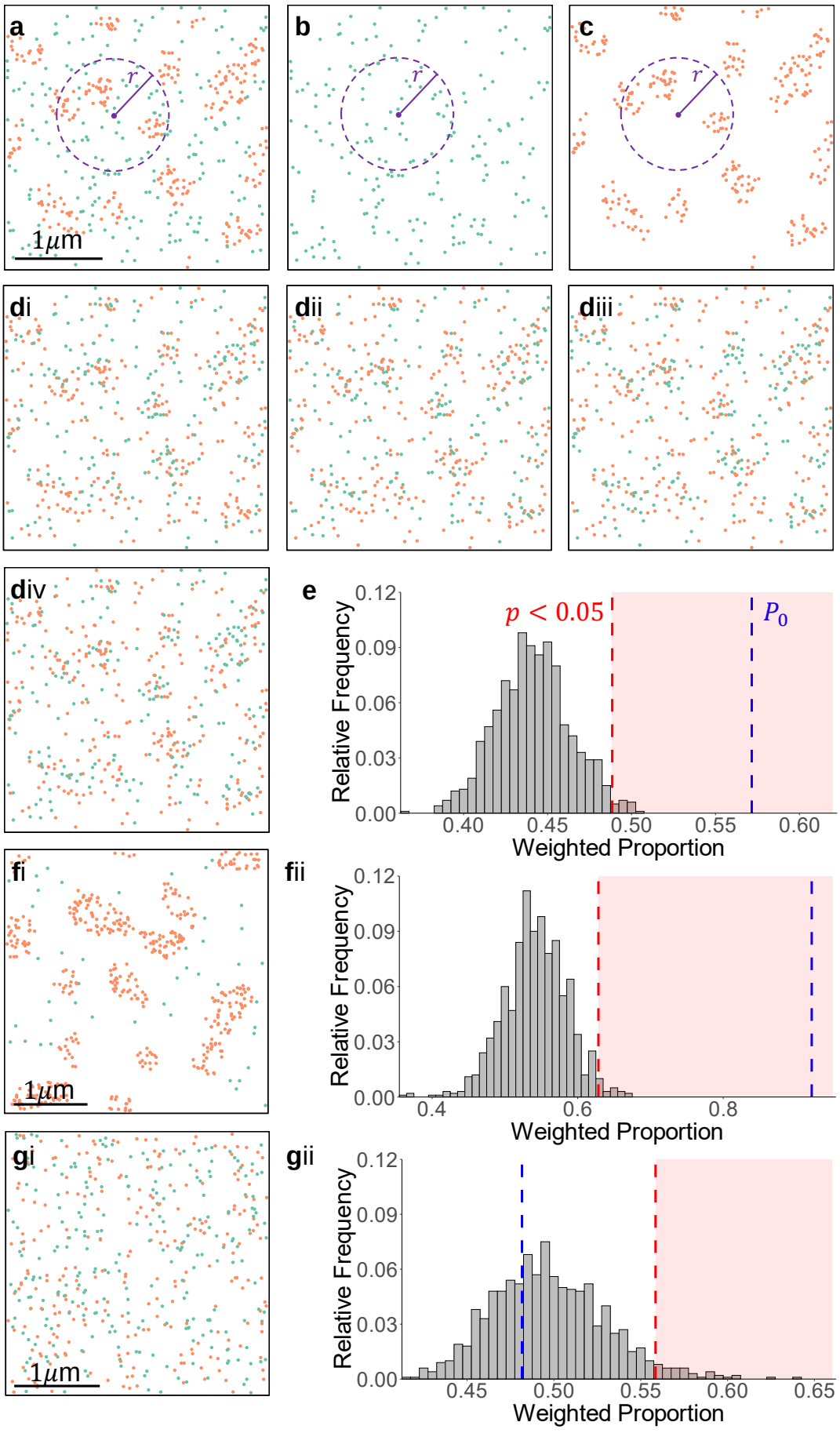

a
b
c
r
1μm
di
dii
diii
div
e
$p < 0.05$
P_0
Relative Frequency
0.12
0.09
0.06
0.03
0.00
0.40
0.45
0.50
0.55
0.60
Weighted Proportion
fi
fii
1μm
0.4
0.6
0.8
gi
gii
0.45
0.50
0.55
0.60
0.65

Figure 4.1: Schematic of P-Check. **a** A discrete point pattern can be segmented into subset point patterns for each category (green and orange; **b-c**). Frequencies can be calculated from both categories to determine the total number of neighbours and neighbours of shared category. These values are summed over all points and the ratio gives the weighted proportion. **d** A permutation test is performed, which randomises the mark of each point. Panels i-iv show example permutations. **e** The weighted proportion P_k can be calculated for each permutation k and compared P_0. If P_0 falls within the top $\alpha\%$ of all permutations, then the significance level can be estimated as $\alpha\%$. Binsize: 0.005. **f** Example data set with low overlap and pronounced clustering. Binsize: 0.005. **g** Example data set with high overlap over a CSR distribution. Binsize: 0.01. All ROI sizes: 1μm by 1μm. The box marked in red signifies the top 5% of weighted proportions and represents the region at which the value of P_0 (blue dashed line) is statistically significant at the 5% level.

ranging from 0-1, representing highest similarity and lowest similarity, respectively (**Figure 4.2c**). A filtration is constructed by ordering points from highest to lowest similarity, which can be used in persistent homology. A deviance threshold, denoted τ, defines the maximum acceptable difference between the marks of the root (starting point) and any other point in the cluster. Each point is connected to all neighbouring points whose similarity values are higher than its own and whose roots' marks are at most τ higher in value. In scenarios where two distinct clusters overlap, the cluster whose root has lowest similarity is absorbed by the opposing cluster. The convex hull of the identified clusters can be used as an estimate of the domain. See **Figure 7.5** (**Appendix**) for conceptual diagram of JOSEPH and **Section 7.2.6** for algorithm pseudocode. Estimates for both parameters may be automatically calculated from the data (see below).

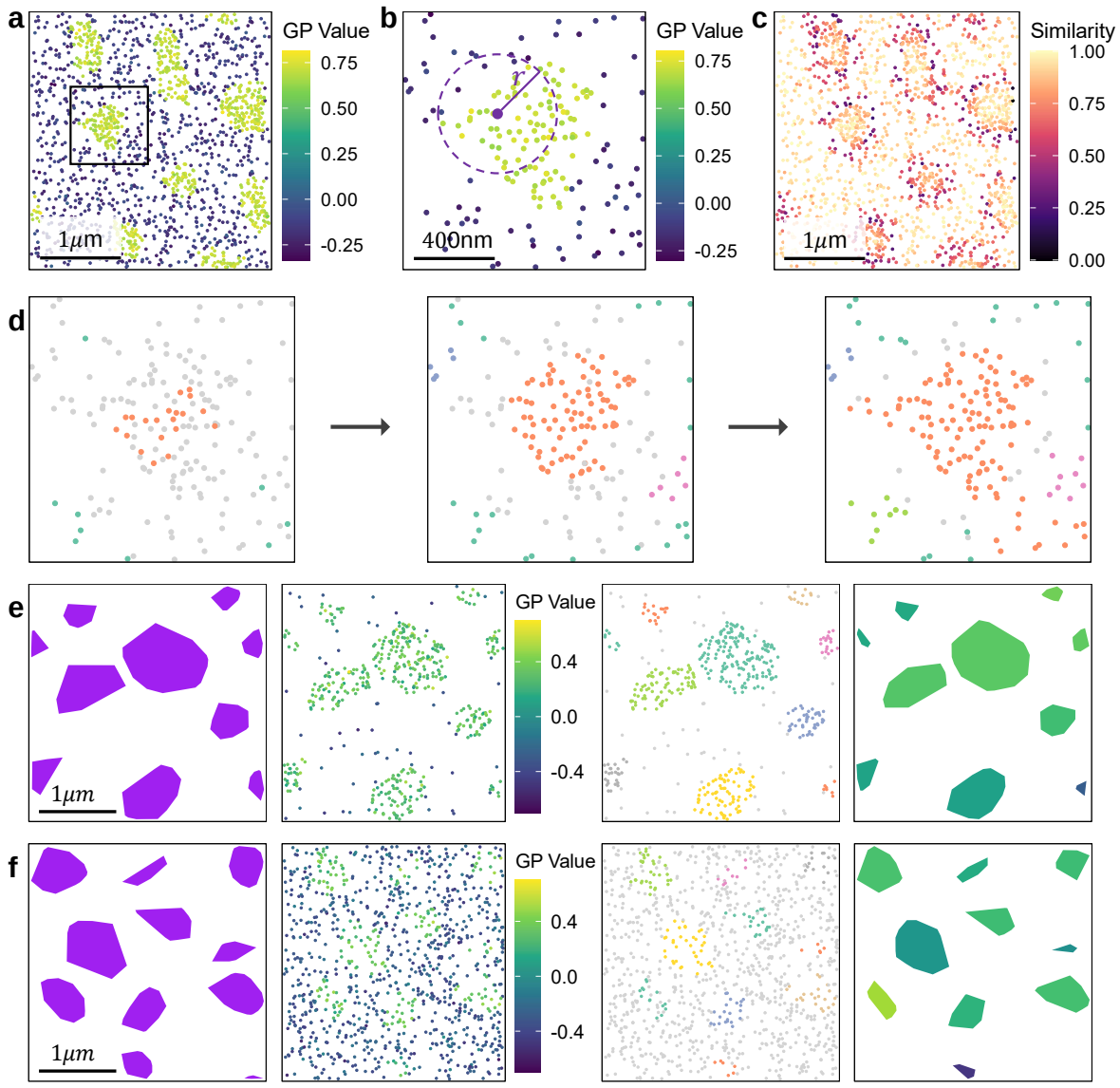


Figure 4.2: Workflow of JOSEPH. **a** Simulated marked point pattern with a highlighted region. **b** The highlighted region of the point pattern, each point's mark is compared to its neighbours. **c** Each point is assigned a similarity score. Points on the boundaries of clusters typically display the lowest similarity scores. **d** Clusters are constructed by iteratively attaching points to neighbours of higher similarity until the deviance threshold is met. **e** JOSEPH performance on a simulated clustered distribution. Ground truth underlying domains are given in purple. Points are randomly added onto the ROI with a 90% chance of being added inside a domain. Marks are pulled from a Normal distribution with mean 0.25 for points inside domains and -0.25 for points outside domains (standard deviation is 0.1 in both cases). Clusters are recovered by JOSEPH and convex hulls of these clusters are coloured by the average mark within

the cluster. **f** JOSEPH performance on a simulated CSR distribution. Simulation protocol is the same as in **e**, except a CSR distribution is overlaid across the whole ROI.

Calculation of intersection over union

Algorithmic performance was quantified by intersection over union (IoU), compared between ground truth domain hulls and hulls of clusters returned by JOSEPH. IoU scores were calculated between pairs of hulls by taking both the intersection and union of the areas of all hulls, then dividing the area of the intersection by the area of the union (see **Appendix** for formula). Intersection and union areas were calculated using the sf (Simple Features) R package (v1.0-16)[327].

Parameter estimation: search radius

The analytic techniques presented here require two key parameters: the search radius and the deviance. The search radius refers to the maximum distance permitted between any two neighbouring points of the same cluster. The cluster radius serves as a proxy for the search radius and can be estimated from the maximising value of the Ripley's H-function, as this represents the spatially-averaged distance at which the number of neighbours is maximised (**Figure 4.3a-b**)[192,328]. If the spatial distribution of the point pattern is known to be CSR, it should first be discretised and then partitioned into separate categories. Discretisation could be achieved by established methods of pattern classification[329]. In this instance, it can be shown that GP values taken from the ordered and disordered phases will approximate two distinct Gaussian distributions (see **Section 4.3.5**). As such, a Gaussian decision boundary, as described in the literature, is used to discretise the point pattern[329].

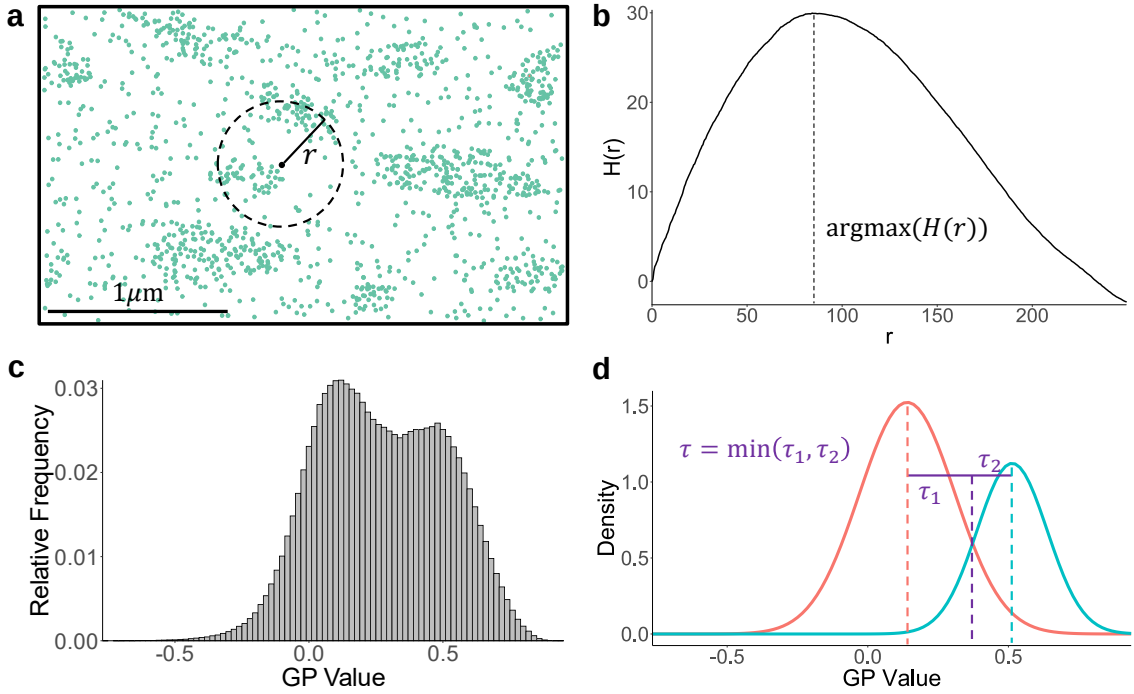


Figure 4.3: Techniques for parameter estimation. **a-b** The Ripley's H function depicts the spatially-averaged density of points at varying radii. The maximising value yields the most likely cluster radius. **c** Histogram of the distribution of all GP values acquired from artificial membrane data. Binsize: 0.02. **d** Results of fitting two Gaussian distributions to the histogram. The deviance τ is taken to be the minimum distance from either peak to the point of intersection.

Parameter estimation: deviance

As above, we assume that the distribution of marks is represented by the sum of two Normal curves, given as $f_1 + f_2 = cN(\mu_1, \sigma_1) + (1 - c)N(\mu_2, \sigma_2)$, and use Gaussian mixture modelling with maximum likelihood estimation to approximate the five unknowns[330]. Assuming the search radius does not exceed true domain size, the expected mark of any root in an ordered or disordered cluster will be approximately equal to the mean of the corresponding distribution. This is because a neighbourhood around this point will have an average mark that approximates the mean of the Gaussian, and therefore minimises the absolute difference. The deviance quantifies

the maximum permissible difference between the root mark (expected to be the mean) and any other point in the cluster. We therefore aim to maximise the deviance without absorbing outliers or points from other domains into the clusters. As such, we calculate the deviance as the minimum difference between the peaks μ_1, μ_2 and the point of intersection between f_1 and f_2 (**Figure 4.3c-d**). If the mark distributions are known to not be Gaussian, then a different sum of probability distributions can be fit. A distribution may be estimated visually by plotting a histogram of all marks. Once an estimate is known, a Kolmogorov-Smirnov test can be used to determine goodness-of-fit[331].

Marked point pattern analysis software

All simulations and analyses with P-Check and JOSEPH were conducted in the R programming language (v. 4.2.3). The PLASMA software package (v. 1.0.0) was written in R (v. 4.2.3) and employed in the integrated development environment RStudio, (2022.07.1+554). PLASMA is available for use under GNU General Public License (v. 3.0).

4.3 Results

4.3.1 Di-4-ANEPPDHQ-based ratiometric PAINT for generating marked point data

In order to generate marked point pattern data from SMLM, each fluorescent molecule must encode both its spatial position and some property of its environment within its fluorescent emission. di-4-ANEPPDHQ yields a large spectral blue shift in its emission, reflecting the decreased polarity and reduced hydration of the membrane, even for single emitters, in line with literature on bulk measurments[119]. To discriminate between

the spectral shifts of single di-4-ANEPPDHQ probes, two channels were used to detect emission, with channel 1 corresponding to wavelengths of 505-600nm and channel 2 corresponding to wavelengths of > 640nm. At di-4-ANEPPDHQ concentrations between 20-80nM, an average of 1825±688 molecules were localised per μm^2 with negligible background binding. For both channels, 50-5000 photons were collected per localisation and photon yields were used to calculate the GP value at the emitter position. This encoded both spatial position and the degree of membrane order for all detected molecules, which were aggregated to produce a marked point pattern.

4.3.2 Detecting domains in simulated marked point patterns and determining statistical significance

To validate the algorithmic performance of P-Check, 1000 distributions were simulated from which each mark was assigned from one of two possible Normal distributions (see **Materials and Methods**). The overlap was recorded for each pair of distributions (see **Appendix** for formula). A Gaussian decision boundary was calculated from the distribution of all marks and the point pattern was binarised by thresholding, subject to this decision boundary. For each simulated data set, we recorded the overlap, the domain radius, the value of P_0, and whether or not the ROI passed P-Check. Results suggest that the value of P_0 tends to decrease with higher overlap (**Figure 4.4a-c**). Linear regression analysis suggests a linear relationship with intercept at 0.878 and coefficient -0.108 (**Figure 4.4d**). Simultaneously, the probability of passing P-Check becomes lower, and results suggest a linear trendline with intercept 0.999 and coefficient -0.136 (**Figure 4.4e**). The value of P_0 was found to increase approximately with domain radius (**Figure 4.5a-c**), suggesting a linear trendline with intercept 0.768 and coefficient 2.272×10^{-5} (**Figure 4.5d**). Analogously, the probability of passing

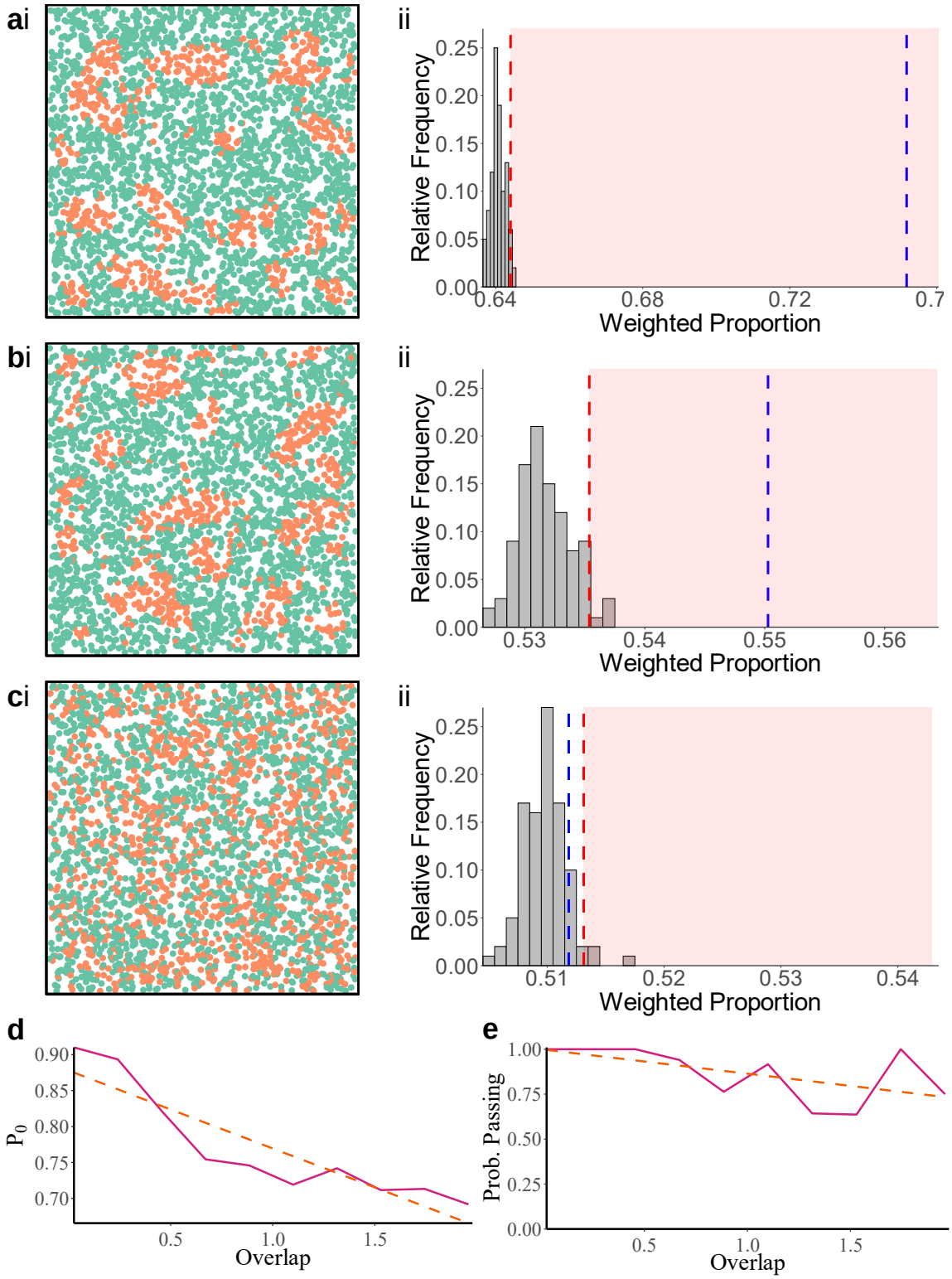


Figure 4.4: The results of P-Check on simulated MPP data sets with increasing overlap between the background and domain GP distributions. Overlaps for **a-c** are 0.5, 1.0 and 2.0, respectively. In each simulation, the discretised marked point pattern

(i) and histogram of weighted proportions derived from P-Check (ii) is shown. All ROI sizes: 1μm by 1μm. Binsize for all histograms: 0.001. The box marked in red signifies the top 5% of weighted proportions and represents the region at which the value of P_0 is statistically significant at the 5% level. As overlap increases, the value of P_0 (blue dashed line) shifts closer to the centre of the histogram. For the highest overlap, there is no statistically significant evidence of partitioning at the 5% level. **d** Mean value of P_0 returned from simulated data sets versus overlap (magenta). Orange dashed line represents line of best fit recovered from linear regression, with intercept at 0.878 and coefficient -0.108. **e** Probability of ROI passing P-Check versus overlap (magenta). Orange dashed line represents line of best fit with intercept 0.999 and coefficient -0.136.

P-Check increased linearly with intercept 0.933 and coefficient 3.803×10^{-5} (**Figure 4.5e**). In general, statistical significance is more likely to be detected over point clouds with larger domains and lower overlap between the distribution of marks in each phase. Notably, for all domains below 200nm, there was at least a 93% chance that they would be detected by P-Check. Furthermore, for all overlap values between 0 and 2, there was at least a 64% chance that domains would be detected. For a simulated data set comprising 2000 points, P-Check took no longer than 1 minute to run on a single processor.

4.3.3 Demonstration of P-Check on artificial membrane data

Synthetic membranes were generated using established GUV preparation methods[323] for liquid-ordered and liquid-disordered phases using DPPC with 30 mol% cholesterol and pure DOPC, respectively. GUVs were osmotically popped to give a planar membrane patch[323]. Di-4-ANEPPDHQ was added to membrane patches at a concentration of 20nM. At each insertion, a blink is detected by the camera in both red and green channels. Molecules are localised to determine x,y-coordinates, and GP

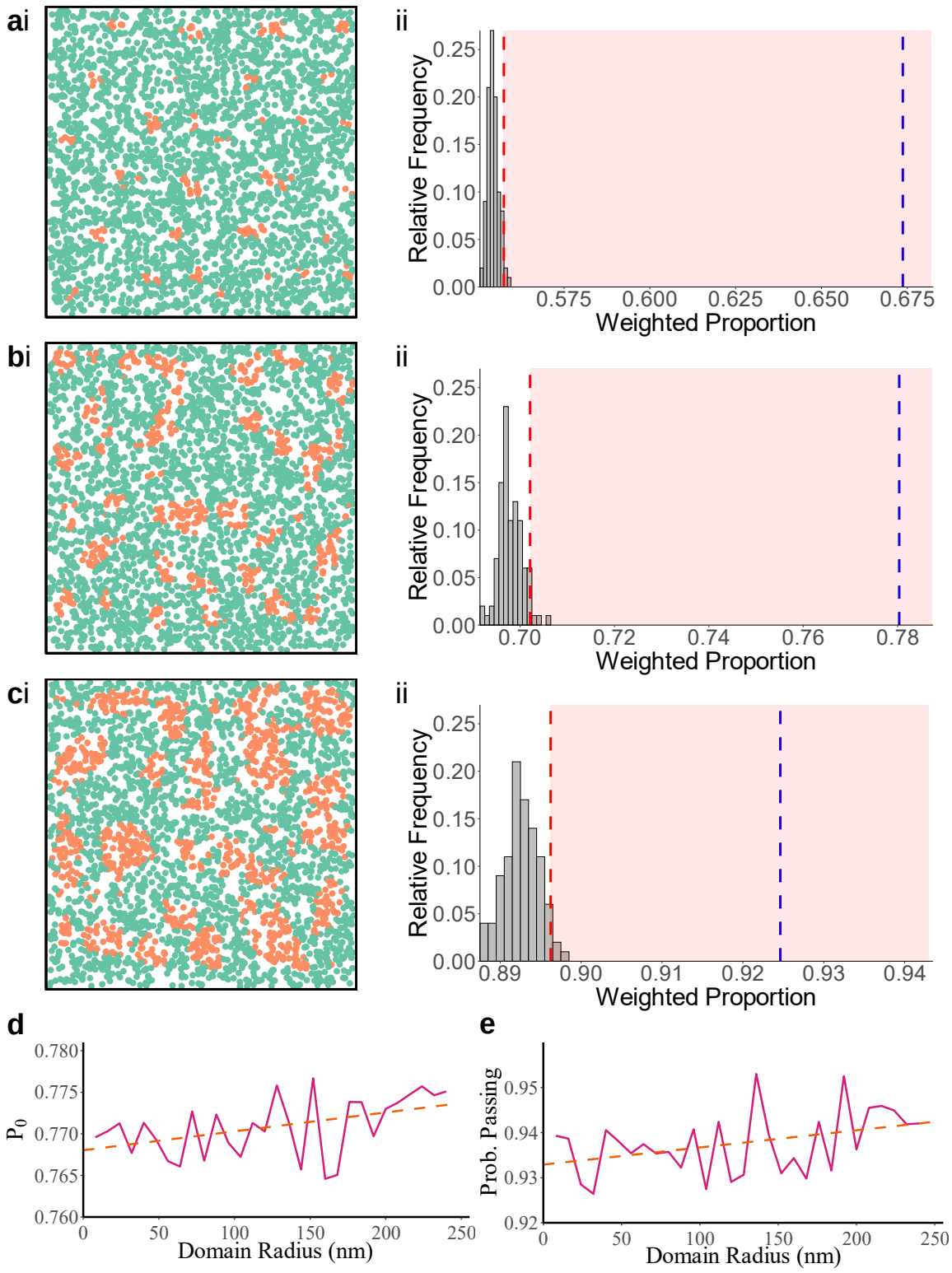


Figure 4.5: The results of P-Check on simulated MPP data sets with increasing domain radii. Domain radii for **a-c** are 40nm, 60nm and 80nm, respectively. In each simulation, the discretised marked point pattern (i) and histogram of weighted proportions derived from P-Check (ii) is shown. All ROI sizes: 1µm by 1µm. Binsize for

all histograms: 0.001. The box marked in red signifies the top 5% of weighted proportions and represents the region at which the value of P_0 is statistically significant at the 5% level. **d** Mean value of P_0 returned from simulated data sets versus domain radius (magenta). Orange dashed line represents line of best fit recovered from linear regression, with intercept 0.768 and coefficient 2.272×10^{-5}. **e** Probability of ROI passing P-Check versus domain radius (magenta). Orange dashed line represents linear line of best fit with intercept 0.933 and coefficient 3.803×10^{-5}.

values are calculated for each localisation from the derived photon count in the two channels. GP values determined from both DPPC and DOPC membrane patches form two distinct Gaussian distributions (**Figure 4.6a-b**). The point of intersection of these Gaussians was taken as the threshold for discretisation, and all marked point patterns were binarised. Notably, the overlap of the distributions corresponding to these phases was 1.455. Spatially-averaged 2D histograms (included in the PLASMA software package) from DOPC and DPPC point patterns are given in **Figure 4.6c and 4.6e**, respectively. The corresponding distributions of P-Check proportional values, P_k, are given in **Figure 4.6d** and **4.6f** and do not suggest statistically significant heterogeneity. This is in line with expectation, since monophase model bilayers are homogeneous in lipid composition, with little variation in hydration across the membrane patches. Therefore, in model bilayers where large differences in GP are not expected, P-Check does not detect domains.

4.3.4 Detecting domains in marked point patterns from RAMA27 data

Having demonstrated P-Check on synthetic membranes, the next step was to probe nanoscale membrane order of live cells. Ratiometric PAINT data was acquired analogously on RAMA27 cells in the presence of an 80nM concentration of di-4-

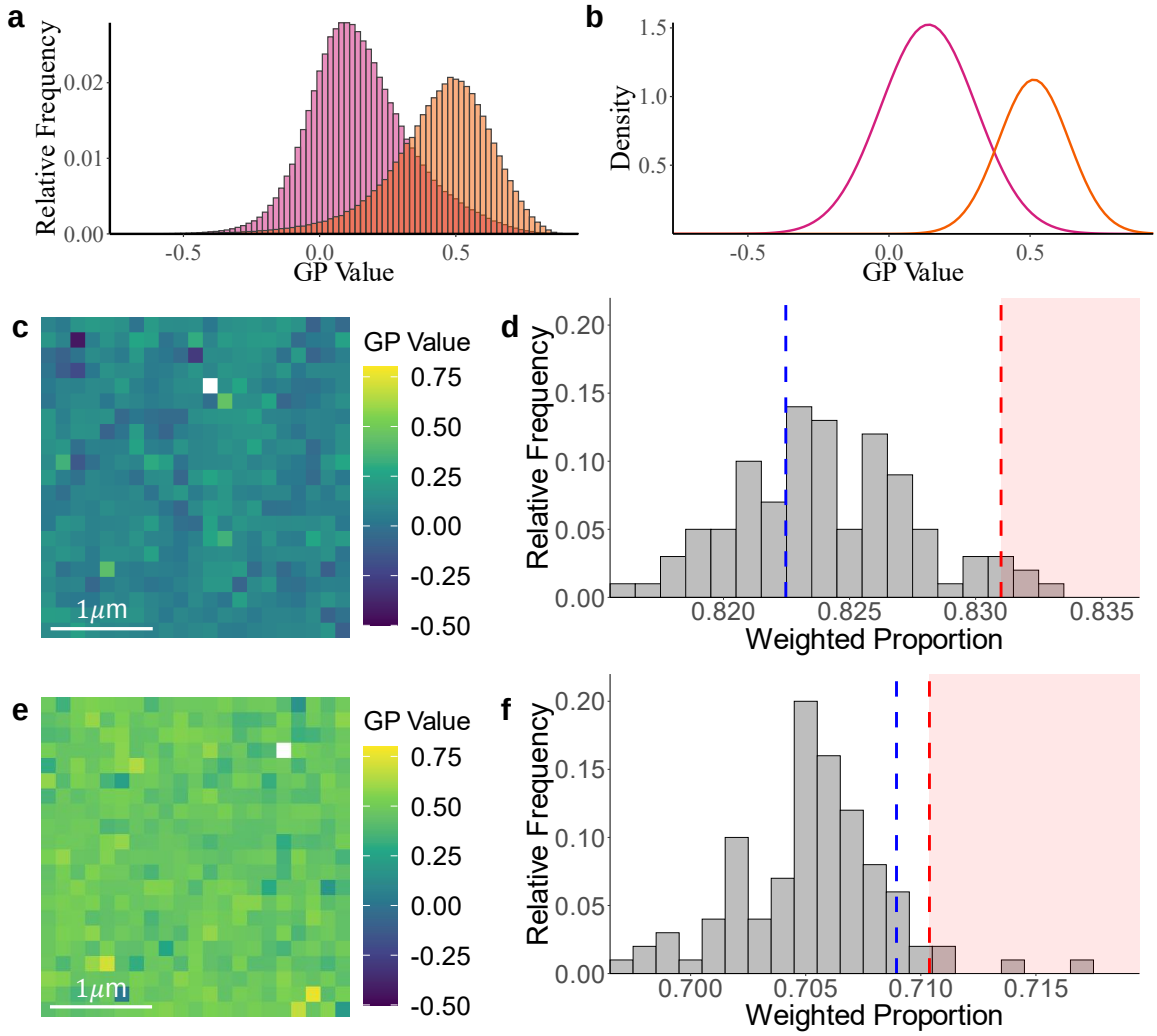


Figure 4.6: Applications to synthetic membranes. **a** Distribution of GP values from both the DOPC data (magenta) and DPPC (orange) data. Binsize: 0.02. **b** Gaussian fit for distributions of GP values from DOPC (magenta) and DPPC (orange) data. **c-d** The membrane order map (GP histogram) of an example DOPC data set and the corresponding P-Check results. **e-f** The GP histogram of an example DPPC data set and the corresponding P-Check results. Analysis does not find any statistically significant evidence of separation in either case. All ROI sizes: 3μm by 3μm. For **b**, **d**, the box marked in red signifies the top 5% of weighted proportions and represents the region at which the value of P_0 (blue dashed line) is statistically significant at the 5% level. Binsize for all P-Check histograms: 0.001.

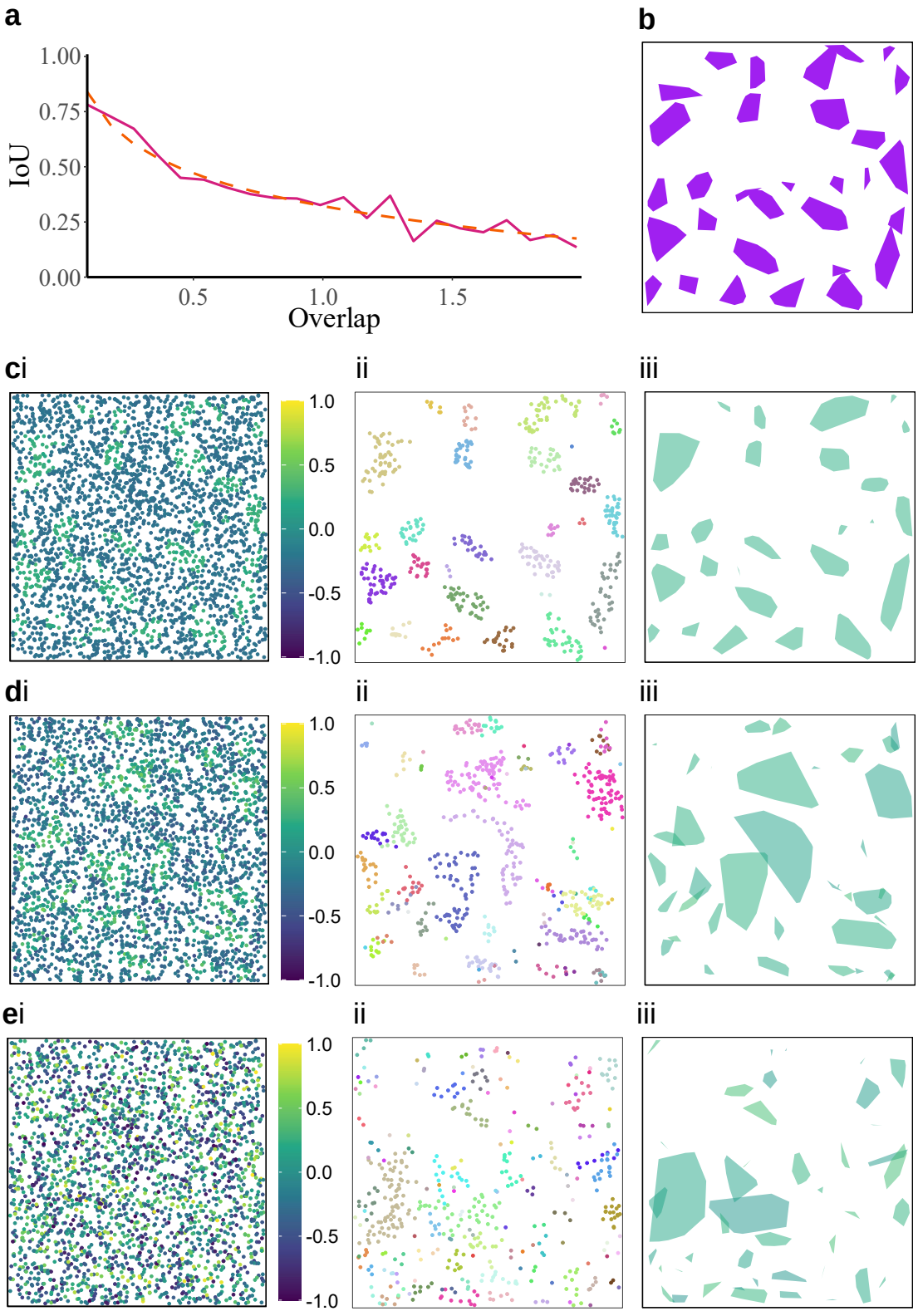


Figure 4.7: The results of JOSEPH on simulated MPP data sets with increasing overlap between the background and domain GP distributions. **a** Mean IoU returned from simulated data sets versus overlap (magenta). Orange dashed line represents

logarithmic line of best fit, with intercept 0.322 and coefficient -0.215. **b** Ground truth hulls over which exemplar MPPs are simulated. Example MPPs are given in panels **c-e**, with overlaps 0.5, 1.0, and 2.0, respectively. Each row depicts (i) the MPP simulated over the ground truth hulls in **b**, (ii) the clusters identified by JOSEPH (with low-GP background of GP < 0.0 filtered out) and (iii) the convex hulls of the clusters, with each hull's colour corresponding to its average GP. All ROI sizes: 1μm by 1μm. Each simulation is comprised of 3000 points with an average domain radius of ~60nm. Mean background GP = -0.25, mean domain GP = 0.25, with standard deviation of both = 0.03125, giving an overlap value of 0.5. Colour bars represent simulated GP value.

ANEPPDHQ in DMEM. In all other aspects, the microscope setup and analysis pipeline were the same as for synthetic membranes. P-Check identified non-random heterogeneity in 194 of 316 ROIs from RAMA27 cell data. This suggests that statistically significant non-random distributions of GP values are present across most ROIs, and that there is heterogeneity in lipid order across RAMA27 cell plasma membranes.

4.3.5 Quantitative mapping of domains in simulated marked point patterns

JOSEPH's performance was validated on simulated data sets with varied mark overlap values (**Figure 4.7**), domain radii (**Figure 4.8**) and point densities (**Figure 4.9**) as ground truth (see **Materials and Methods** for quantification). For each simulation, the domains identified by JOSEPH were recorded and the IoU score was calculated between results and the ground truth[332]. The IoU quantifies the degree of overlap between domains and the ground truth, while penalising overfitting (see **Appendix** for formula)[196]. Here, a complete lack of overlap yields a value of 0, while perfect overlap gives a value of 1. From this, we identify three variables which impact JOSEPH's performance. Firstly, the IoU score decreases as mark overlap increases, and results

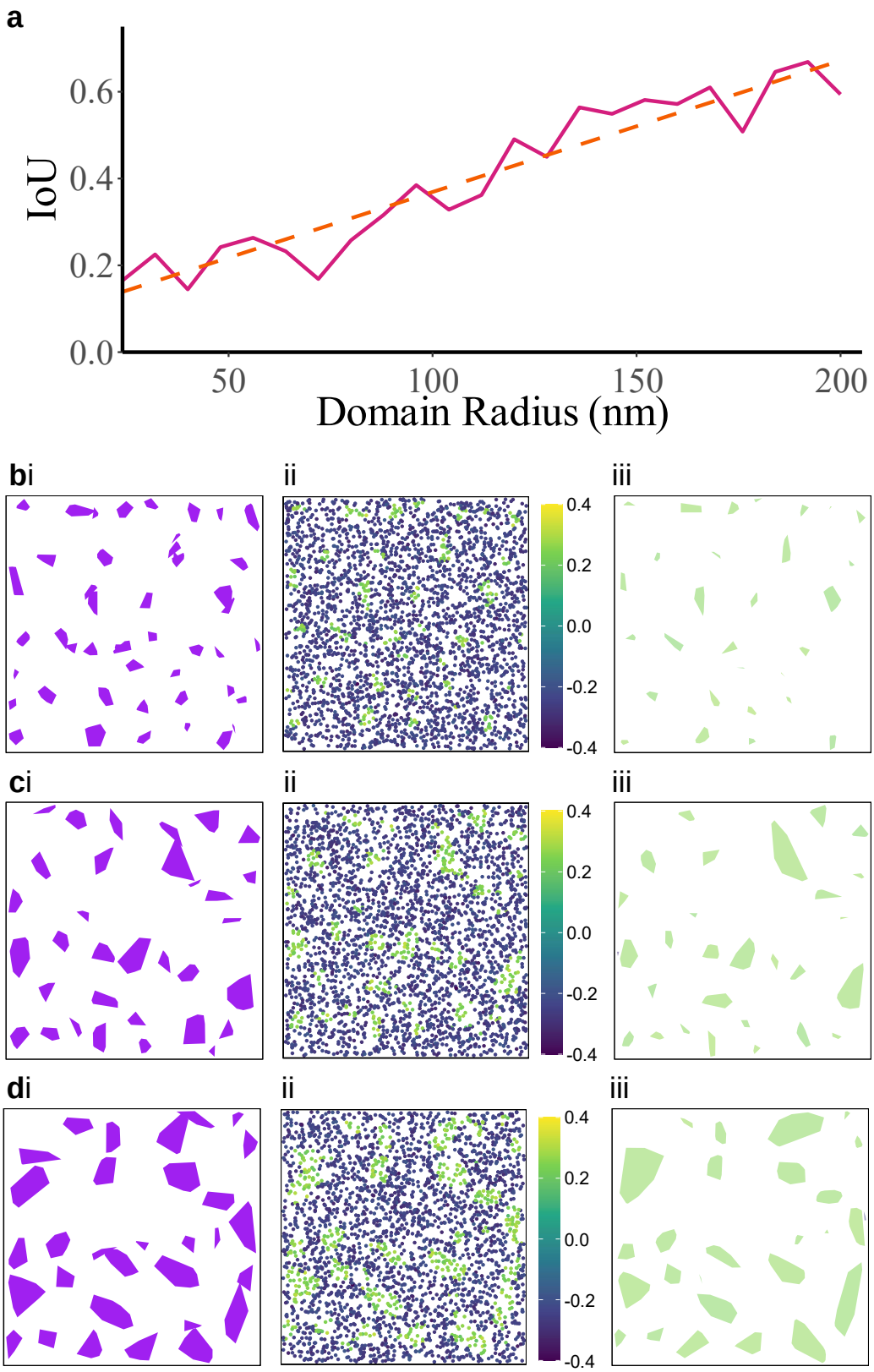

a
IoU
0.0
0.2
0.4
0.6
50
100
150
200
Domain Radius (nm)
bi
ii
iii
ci
ii
iii
di
ii
iii
0.4
0.2
0.0
-0.2
-0.4

Figure 4.8: The results of JOSEPH on simulated MPP data sets of varied domain size. **a** Mean IoU returned from simulated data sets versus domain radius (magenta). Orange dashed line represents line of best fit recovered from linear regression, with intercept 0.066 and coefficient 0.003. Average cluster sizes for **b-d** are 40nm, 50nm, and 60nm, respectively. Each row depicts (i) the convex hulls of the ground truth domains, (ii) the simulated MPP, and (iv) the convex hulls of the clusters, with each hull's colour corresponding to its average GP (low-GP background hulls of GP < 0.0 are filtered out). All ROI sizes: 1μm by 1μm. Each simulation is comprised of 3000 points. Mean background GP = -0.25, mean domain GP = 0.25, with standard deviation of both = 0.03125, giving an overlap value of 0.5. Colour bars represent simulated GP value.

suggest a logarithmic trendline with intercept 0.322 and coefficient -0.215 (**Figure 4.7a**). This suggests that the quality of the partition becomes increasingly poor at higher overlap values. Exemplar simulated data sets of varied overlap are visualised in **Figure 4.7b-e**, alongside the outputs of JOSEPH. Secondly, the IoU value tends to increase with greater domain radius, and regression analysis suggests a linear relationship with intercept at 0.066 and coefficient 0.003 (**Figure 4.8a**). Exemplar simulated data sets with varied domain sizes are visualised in **Figure 4.8b-d**, alongside the outputs of JOSEPH. This relationship may arise as a result of lower point counts within smaller domains, which is inherent of CSR distributions. To quantify the impact of point density, we recorded the IoU versus proportion of points assigned to domains, p_d, on non-CSR data sets only. As the value of p_d increases, so too does the IoU (**Figure 4.9a**), suggesting a logarithmic trendline with intercept 0.688 and coefficient 0.232. These results suggest that domains are less likely to be detected if the density of points within them is low. Exemplar simulated data sets with varied values of p_d are visualised in **Figure 4.9b-e**, alongside the outputs of JOSEPH. In summary, the quality of the partition returned from JOSEPH (as quantified by the IoU

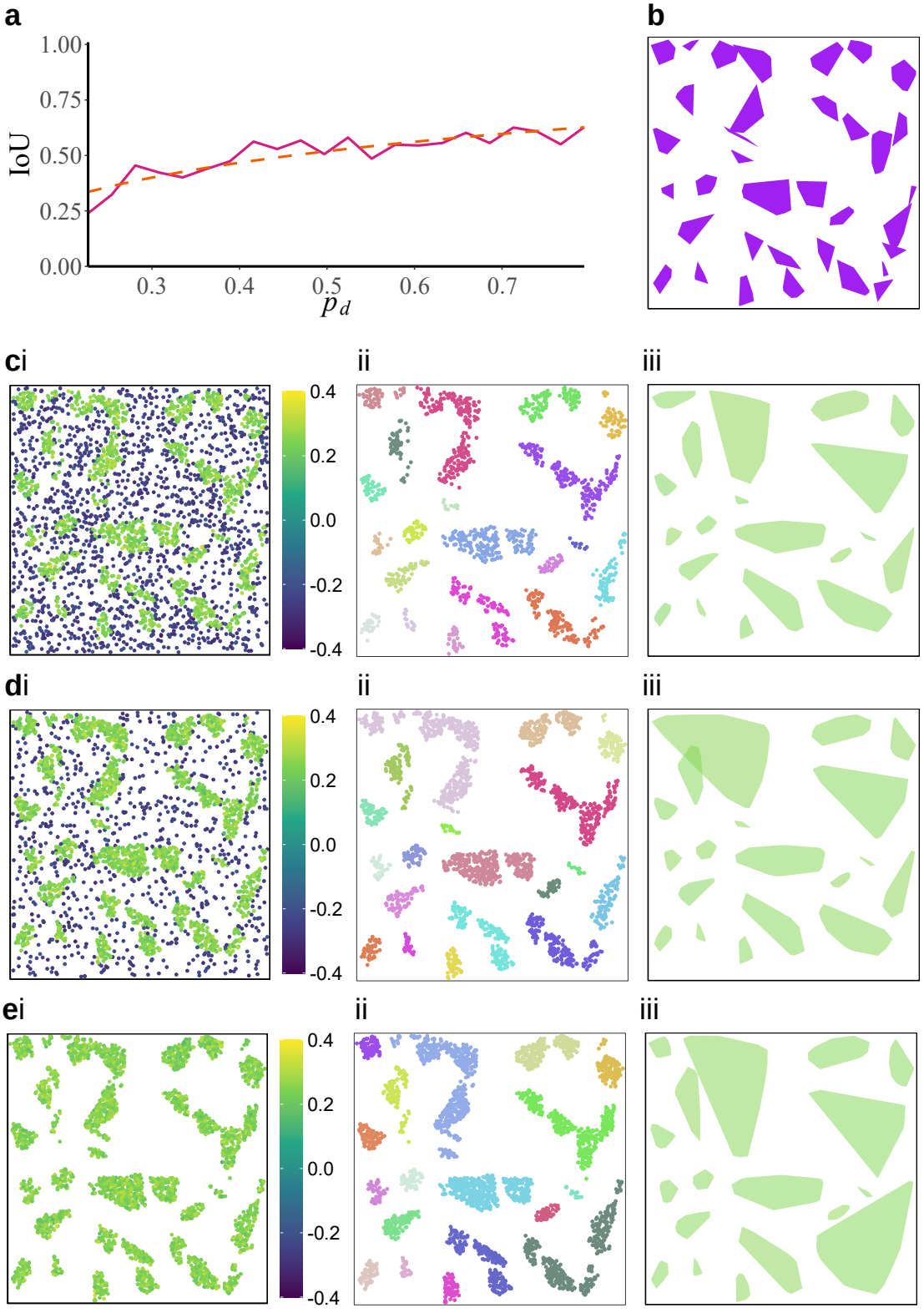


Figure 4.9: The results of JOSEPH on simulated MPP data sets, varying the proportion of points assigned to domains, p_d. **a** Mean IoU returned from simulated data sets versus p_d (magenta). Orange dashed line represents logarithmic line of best

fit, with intercept 0.688 and coefficient 0.232. **b** Ground truth hulls over which exemplar MPPs are simulated. Example MPPs are given in panels **c-e**, with $p_d = 0.5$, 0.75, and 1.0, respectively. Each row depicts (i) the MPP simulated over the ground truth hulls in **b**, (ii) the clusters identified by JOSEPH (with low-GP background of GP < 0.0 filtered out) and (iii) the convex hulls of the clusters, with each hull's colour corresponding to its average GP. All ROI sizes: 1μm by 1μm. Each simulation is comprised of 3000 points with an average domain radius of ~60nm. Mean background GP = -0.25, mean domain GP = 0.25, with standard deviation of both = 0.03125, giving an overlap value of 0.5. Colour bars represent simulated GP value.

score) improves with lower overlap and greater point density within domains. For a simulated data set comprising 2000 points, JOSEPH took no longer than 10 second to run on a single processor.

4.3.6 Measuring nanoscale membrane order in live cell membranes with PLASMA

ROIs from live RAMA27 cell data which exhibited statistically significant evidence of mark heterogeneity at the 5% level (determined via P-Check) were carried forward and analysed with JOSEPH (**Figure 4.10a-c**). For each cluster found, the average GP of all cluster points was compared to the global average GP within the ROI to determine the difference in GP values across domains (**Figure 4.10d**). Results show two peaks comprising clusters with GP values above and below the global average, respectively. Furthermore, the area of each convex hull was recorded across all analysed ROIs (**Figure 4.10e**). On average, domains occupied an area of approximately 0.15μm^2 with 90% of all domains falling between 0.034μm^2 and 0.293μm^2. Assuming domain circularity, this corresponds to radial values of around 100-300nm. This may suggest that both lipid-ordered and lipid-disordered nanodomains are present within RAMA27 cells.

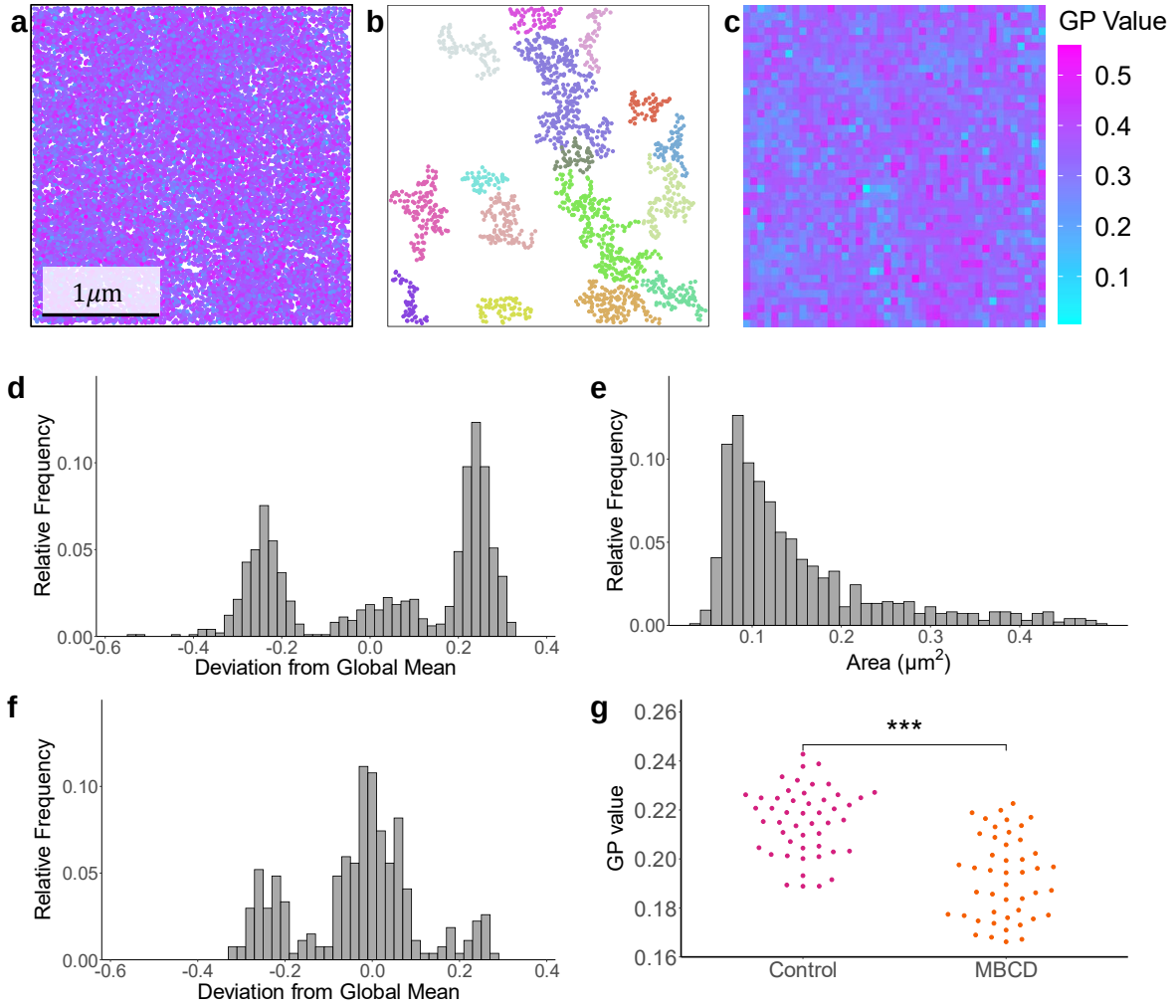


Figure 4.10: Live cell membrane results. **a** A marked point pattern acquired from 2-channel PAINT SMLM on RAMA27 cell lines. **b** Clusters identified by JOSEPH with average GP values which were significantly different to the global mean. **c** The membrane order map of the ROI. **d** Histogram of the relative frequency which domains identified by JOSEPH fell into intervals above (ordered) or below (disordered) the global average GP value (~0.22) for untreated cells. **e** The areas (in μm^2) of all domains identified by JOSEPH in untreated cells. **f** Histogram of the relative frequency of domains identified by JOSEPH which fell into intervals above (ordered) or below (disordered) the global average GP value (~0.17) for cells treated with MβCD. **g** Beeswarm plot of mean GP value for all domains within selected ROIs for untreated (Control, magenta) and methyl-β-cyclodextrin (MβCD, orange). Statistical significance was determined via two sample t-test over 57 control ROIs and 52 MβCD ROIs.

Significance ranking: n.s - not significant, * - $p < 0.05$, ** - $p < 0.01$, *** - $p < 0.001$. All binsizes: 0.02.

Further, to demonstrate the method's sensitivity to changes in lipid order, we used a well-established cholesterol binding molecule, methyl-β-cyclodextrin (MβCD)[333], to remove cholesterol from the plasma membrane. Cells were treated with MβCD and compared to untreated (control) cells using P-Check and JOSEPH (**Figure 4.10f-g**). P-Check identified non-random heterogeneity in 49 of 57 ROIs from untreated cell data and 46 of 52 ROIs from MβCD-treated cell data. We recorded the difference between global mean GP and the mean GP of domains identified by JOSEPH for MβCD-treated cells (**Figure 4.10f**). In this case, we observed that domains now elicited a central peak and two smaller lobes relating to the lower and higher order domains identified. Notably, the relative frequency of domains above the global mean GP has decreased when compared with **Figure 4.10d**. Further, we observed a decrease in the average GP value per ROI for MβCD treated cells when compared with the untreated cells (**Figure 4.10g**). A relative shift towards low-order domains is observed upon treatment with MβCD, which suggests greater lipid disorder within both the domains and the whole membrane.

4.4 Discussion

In this work, we have introduced a series of topological data analysis techniques, under the PLASMA umbrella package, for detecting heterogeneity in marked point patterns. By applying these methods to simulated data with known ground truths, we have demonstrated their capacity to determine the existence and location of underlying domains. In particular, we find that P-Check is more likely to detect statistically significant heterogeneity over point patterns with larger domains and lower

overlap between the distribution of marks in each phase. Among these simulations, for all domains below 200nm, there was at least a 93% chance of detection, and for all overlap values between 0 and 2, there was at least a 64% chance of detection. For JOSEPH, we find that the quality of the partition returned (as quantified by the IoU score) improves with lower overlap and greater point density within domains.

4.4.1 Results from ratiometric PAINT data of GUVs and RAMA27 cells

Using a ratiometric PAINT approach with the solvatochromic probe di-4-ANEPPDHQ, we produced marked point patterns of the distribution of membrane order, represented by GP values, in GUVs and RAMA27 cells. Notably, the overlap of the distributions corresponding to ordered and disordered phases in GUVs was 1.455, which (according to simulations) would suggest a ~65% of detecting domains. We applied P-Check to live-cell RAMA27 data and identified non-random heterogeneity in 194 of 316 ROIs (~62%) at the 5% significance level. Then, after passing these ROIs through JOSEPH, we find significant deviation between the average GP values of identified domains and the global average GP. Furthermore, we calculate the area of all convex hulls and find that 90% of all domains are characterised by a radius between 100-300nm. However, since these quantities were derived from convex hulls, rather than concave, they may be overestimates of the true size. Ultimately, this could suggest a degree of regulated lipid heterogeneity across the plasma membrane. Furthermore, we are able to quantify changes in membrane order upon manipulation of the plasma membrane with methyl-β-cyclodextrin, where a shift towards the disordered domain population was observed.

4.4.2 Applicability of PLASMA and advances in bioinformatics

As discussed, methods for partitioning marked point patterns have yet to keep pace with more generalised spatial cluster analysis algorithms. The PLASMA package could bridge this gap by providing a segmentation pipeline for marked pattern data. Only two input parameters are required for PLASMA analysis: a search radius and a mark deviance threshold, for which we have supplied additional methods for parameter estimation, built into the software package by default. As a TDA tool, JOSEPH is applicable to MPPs which are spatially anisotropic. Notably, JOSEPH relies on the local mean of each point's neighbourhood as a measure of similarity, although this may not provide an accurate summary statistic for data sets of particularly low density. However, in this specific application to SMLM, data sets are often sufficiently dense, so it is easy to overcome this limitation[129,130]. JOSEPH is an unsupervised machine learning approach, which may be prone to hallucinate clusters. Since JOSEPH does not filter the connected components returned from persistent homology, every point is assigned to a cluster by default. Therefore, it is recommended to first conduct P-Check in order to determine if domains may actually be present within the data. For this reason, it is important to consider both algorithms in the PLASMA package.

4.4.3 Concluding remarks

In conclusion, we have developed a software package for detecting heterogeneity in marked point patterns and partitioning marked data sets in the style of traditional cluster analysis. We find these methods are applicable to data extracted from 2-channel ratiometric PAINT SMLM acquisition using the solvatochromic probe di-4-ANEPPDHQ. In conjunction with cluster and colocalisation analysis, PLASMA may serve as a basis for probing, mapping and quantifying the nanoscale spatial

organisation of lipid phase domains in the cell plasma membrane. The data presented suggest that large scale segregation of domains may not be ubiquitous, but that both liquid-ordered and liquid-disordered lipid domains are present at the nanoscale (~100-300nm) in RAMA27 cells. Notably, the ordered domains identified here may be characteristic of lipid rafts. As imaging modalities become increasingly optimised, ratiometric SMLM methods produce greater precision in localisation and GP determination. Therefore, analysis packages such as PLASMA could play a central role in determining the existence and distribution of nanoscale lipid phase domains and may offer an avenue for visualising and quantifying lipid raft properties. By employing PLASMA over SMLM data acquired with different environmentally-sensitive probes, we may be able to map other properties of interest (e.g., membrane charge or tension[334]), and give insights into how these properties interplay with the lipidome. This work summarises a novel approach for MPP analysis and highlights the potential for further membrane mapping techniques, specifically towards SMLM with environmentally-sensitive dyes.

Chapter 5: Discussion and conclusion

The development of analytic techniques for quantifying biophysical membrane properties remains at the forefront of active research for biologists, immunologists, data scientists and microscopists. Protein aggregation on the plasma membrane underpins essential cellular processes, such as signalling and communication, and it is now known that lipid packing correlates with transmembrane protein motility[64,148]. Central to this notion is the spatial oligomerisation or clustering of transmembrane proteins, which is hypothesised to convert analogue extracellular signals into digital intracellular counterparts[78,289]. While the existence, prevalence and function of lipid rafts remain controversial topics, it is now generally accepted that oligomerisation associates with membrane lipid order and that protein clusters are tuned to optimise the transmission of signalling information[335]. To further research into the morphological properties and physiological relevance of ordered lipid domains, it is essential to devise scalable spatially-descriptive statistics to interpret the complex data of SMLM and its evolving imaging modalities. This thesis presents a series of topology-based analytic tools for image segmentation, modelling protein aggregation, and point pattern partitioning, designed for use with conventional and super-resolved FM under environmentally-sensitive probes. TOBLERONE, a topological image analysis tool, is developed for classical cell segmentation in confocal microscopy (**Chapter 2**) and identifies regions of varied lipid packing across members of the fungal species *C. gattii*. Moving towards the nanoscale, ASMODEUS is introduced as a PAD simulator with applications to generalising transmembrane protein motility and generating realistic SMLM data (**Chapter 3**). These simulations are then implemented in **Chapter 4** to validate the analytic methods comprising the PLASMA package, which are used to identify heterogeneity in MPP data. The techniques presented here may serve as

platforms for studying properties of the membrane nano-environment in the evolving landscape of SMLM and big data quantification.

5.1 Summary of results

The central contributions of this thesis come in the form of novel topological data analysis techniques for probing biophysical data, derived from different forms of microscopy, with specific applications to the cell plasma membrane. In **Chapter 2**, we show that TOBLERONE segmentation offers a trade-off between sensitivity and specificity which may not be achievable with pre-existing classical methods. We quantify algorithmic performance under simulated image artefacts, Gaussian noise and blur, and determine that sensitivity experiences a greater decrement with noise, while specificity is more sensitive to blur. Generally, we find that most pixels belonging to an object are correctly identified and most pixels belonging to the background are correctly ignored by TOBLERONE. In addition, we demonstrate TOBLERONE on confocal FM data. We quantify membrane order in *C. gattii* with di-4-ANEPPDHQ and find statistically significant evidence at the 1% level that treatment with 2OHOA or 7-ketosterol can lower the average GP value across the *C. gattii* cell plasma membrane. For 3D cell segmentation, we find that Jurkat T cells span a diameter of approximately 15.5μm on average, in accordance with existing literature[282-284]. Further, we track *S. pombe* dynamics using tempTOBLERONE, and determine that progression from telophase to interphase in the nuclear division of *S. pombe* can occur in under a minute, in line with existing reports[285].

In **Chapter 3**, we shift focus to nanoscale protein aggregation and develop a simulator for protein map data. We find that an ABM approach, using the Ripley's K function as a metric of cluster affinity, can induce point pattern aggregation with pre-defined

cluster properties. For single populations of molecular species, all 9600 simulations reached convergence within the model runtime of 2.5 minutes, and experienced a 6-fold reduction in error. In addition, all associated cluster properties fell to within 10% of their target values at convergence. Furthermore, we find that nucleation can increase the likelihood of cluster formation and that the probability of convergence increases with greater protein recruitment, even at a relatively low number of nucleation sites. This same result holds true for simulations of CD3ζ distributions in Jurkat T cells, with induced nucleation increasing the likelihood of favouring a distribution associated with activated conditions as opposed to non-activated. For multiple population models, we find that an instantaneous switch in clustering regime can cause simulated activators and agents to co-cluster or disperse. This elicits a sudden shift in agent activation, and induces a response analogous to digitisation. ASMODEUS therefore offers a dual purpose in generating static point clouds for validating analytic techniques and dynamic simulations for comparing molecular interactions across hypothetical distributions.

In **Chapter 4**, we implemented 2-channel PAINT SMLM with solvachromatic probes to explore the distribution of membrane order, represented by point patterns marked with GP values, at resolutions below the diffraction limit of light. We implemented a software package, PLASMA, for identifying heterogeneity and partitioning these data sets. Results on simulated data, derived from ASMODEUS, suggest that P-Check has a 93% chance of detecting domains below 200nm in size, and at least a 64% chance of detecting domains with overlap values between 0 and 2. Further, we find that the quality of the partition returned by JOSEPH (quantified by the IoU score) improves with lower overlap and greater point density within domains. In experimental data, results suggest that lipid domains exist in RAMA27 membranes and are more

abundant than in synthetic membrane data. The domains identified in live cells exhibited mean GP values that deviated significantly from the global average of their host ROI. This could suggest a degree of regulated heterogeneity in lipid packing across the plasma membrane. Domains ranged in sizes between 0.034μm^2 and 0.293μm^2 in area, suggesting radii of around 100-300nm, assuming circularity. As these length scales are typically observed close to the diffraction limit of light, this form of analysis may offer a unique and detailed insight into membrane heterogeneity and the lipidome.

5.2 Suitability of methods

5.2.1 TOBLERONE

State-of-the-art methods in cell and organelle segmentation rely on machine and deep learning frameworks. Such frameworks are typically built upon convolutional neural networks (CellSeg[250], CellPose[251,273] and 3DCellSeg[217]) and may impose explicit cell geometries (StarDist[249] and 3DStarDist[336]). With the advent of weakly-supervised training regimes[337] and efficient image processing tools, such as SegmentAnything[252], it is likely that artificial intelligence (AI) will see increased integration into desktop analysis pipelines in future. However, AI-based models warrant greater processing power and annotated training data, which may not always be available. Classical methods may be used in lieu of machine learning (ML), and may involve either simple binarisation methods (Otsu, Huang or Ray thresholding[338,339]), binary map post-processing (seed-point extraction[340], Watershed algorithm[278,341], gradient-based detection[338]), or some combination of both (Spot Segmentation and 3D Simple Segmentation[264]). That said, classical methods can present analogous drawbacks, in terms of increased parameterisation and geometric specificity.

As discussed in **Chapter 2**, TOBLERONE draws topological image processing to the attention of biologists and microscopists, as an alternative to existing classical segmentation methods. This allows for complete, geometry-free cell segmentation on the images, z-stacks and t-stacks acquired from conventional fluorescence microscopy. We note that the algorithmic performance of TOBLERONE is dependent on image quality, meaning that imaging artefacts may skew results. However, this is a shared property among all image analysis tools and can be improved with pre-processing techniques, such as denoising, image smoothing and image sharpening. All implementations of TOBLERONE require a single parameter, and detailed instructions for parameter estimation are listed in the methods (see **Chapter 2**). As a method of TIA, TOBLERONE is invariant of geometric object properties and applicable to images with a high degree of between-cell variation[342,343].

TOBLERONE's capacity to detect objects across a range of fluorescence intensities is particularly applicable to segmenting FM under polarity-sensitive dyes. Such images present a unique data type in that fluorescent emissions are taken from two channels and are heterogeneous across both. As seen in **Chapter 2**, the algorithm has already been used to compare differences in membrane order across *C. gattii* under different growth media. Further, since development, TOBLERONE has been used in probing membrane order among the fission yeast, *Schizosaccharomyces japonicus* and its sister species *Schizosaccharomyces pombe*[37]. The production of unsaturated acyl tails relies on eukaryotic desaturases, which require molecular oxygen, and therefore only occurs in aerobic environments[37]. TOBLERONE has shown that *S. japonicus* can regulate its own membrane fluidity by exploiting phospholipid acyl tail length asymmetry, in the absence of unsaturated fatty acids in the membrane lipidome[37]. This

evolutionary principle allows *S. japonicus* to grow in both aerobic and anaerobic conditions, whereas *S. pombe* cannot[37].

To summarise, we have here shown that topological image analysis techniques can be used for segmentation of FM data. As a subset of TDA, the algorithms introduced can probe for arbitrarily-shaped biological structures of any underlying topology. This permits semi-automatic, reproducible segmentation of stainable cellular components even in data sets with varying pixel intensities. While TOBLERONE and its subsidiaries require greater parameter tuning and are less capable of distinguishing adjacent cells than ML methods, the algorithms provide a suitable alternative for researchers who lack computational resources and want to avoid the black box approach presented by ML.

5.2.2 ASMODEUS

In **Chapter 3**, we introduce a novel method for simulating transmembrane PAD which generates protein maps that match spatially-descriptive statistics derived from experimental data. Existing SMLM simulation models (such as SuReSim[299], FluoSim[300], SMeagol[301], TestSTORM[302], ThunderSTORM[179], LocMoFit[167] and SMIS[303]) generate localisations built on underlying geometries and ground truth structures specified by user input. These methods simulate artefacts based on the photophysical properties of fluorescent molecules to derive realistic SMLM data, however, they require manual user input regarding point pattern distribution. Virtual-SMLM[344] offers an interactive simulator for SMLM acquisition, rather than simply creating data, but is also limited to known ground-truth geometries. Simulators of this form are specific to certain SMLM imaging modalities and require understanding of

photophysical properties, such as photo-activation, bleaching and switching rates. These are often taken to be constant, which is not necessarily a valid assumption[93].

ABMs model PAD by imposing biophysical behaviours on the trajectories of individual agents, such as fixed-rate or variable diffusivity[300,301,303]. Other agent-based modelling approaches have directly simulated lipid rafts as regions of the membrane which inhibit protein diffusivity across multiple molecular species[345]. These simulations found success in simulating dimerisation of the platelet- and megakaryocyte- specific receptor for collagen glycoprotein VI, however, they are limited by the binary diffusivity coefficient used in the model[345]. Additionally, they assume that lipid domains are fixed in space and time, which contrasts with the literature[346-348]. These models also require, at minimum, knowledge of biophysical parameters such as molecule size[345]. Very few of these simulators take account of molecular interactions, but even then, they are still limited by imposed geometry[300,345].

With the advent of ASMODEUS, we can simulate localisation data with generalisable geometries. We show that the simulator is capable of recapitulating the cluster properties of input target distributions from simulated data. Further, the model can track multiple populations of interacting molecular species and generate perturbations which serve as proxies for natural and induced biophysical disruptions. Using ASMODEUS as a spatiotemporal PAD simulator, we can optimise nucleation parameters required to elicit clustering or a phase shift of distribution in simulated data. This simulation technique can be used to generate SMLM data that represents both non-activated and activated distributions, whilst also modelling a shift from one distribution to another during simulation. Research into therapeutic mediators which induce or disrupt TCR aggregation is ongoing[13,296,297], and *in silico* modelling platforms such as ASMODEUS could be used to inform mediator structure.

In order to improve generalisability, we would need to take account of the explicit biophysical processes, occurring on the plasma membrane, which force molecules to adopt configurations defined by Ripley's K function and other spatially-descriptive statistics. As the temporal resolution of SMLM is increasingly improved, we afford greater availability of data to validate new agent-based models and predict PAD. With these developments, we may be able to determine important factors leading to lipid raft formation and redistribution, while also predicting protein dynamics based on phase affinity. With further finetuning, *in silico* modelling could be used for simulating the effect of nanotherapeutics, which may lead to novel clinical applications.

5.2.3 PLASMA

Current methods of marked point pattern analysis rely on mark discretisation and measures of colocalisation. This may be achieved through nearest-neighbour distance analysis, the coordinate-based colocalisation value[239], or scalable metrics such as SODA[234], Clus-DoC[349] and LAMA[241]. The principle drawback of this approach is that the quantitative information associated with each GP value, and therefore the relative differences between adjacent molecules, is lost upon discretisation. Some second-order characteristics can be used to determine correlations between the spatial positions and marks of points in MPPs, such as Isham's mark correlation function, Stoyan's k_{mm} function (normed), the mark variogram and the mark covariance function (unnormed)[233]. These determine relative association between marked points as a function of both distance and difference in marks. In general, the problem of non-spatial clustering remains an open question in point pattern analysis. As such, in **Chapter 4**, we introduce the PLASMA package, which offers a method of MPP partitioning and systems for parameter estimation. This allows for parameter-free

cluster analysis of marked point patterns, characterised by data with three or more dimensions, of which not all are necessarily spatial.

We have endeavoured to overcome any drawbacks associated with the algorithms introduced through PLASMA. Only two input parameters are required across the analysis pipeline, which are automatically estimated from the input data by Gaussian mixture modelling. P-Check is used as a measure of co-clustering, which functions on discrete point patterns, and can determine whether underlying domains are present within a data set. Furthermore, we develop a framework for applying TDA, and specifically persistent homology, to marked point patterns. This is summarised in the second algorithm, JOSEPH, which interprets the local mean mark value as a measure of similarity between points. Notably, this summary statistic loses precision in sparser data sets – however, SMLM often provides rich data sets[129,130]. Although the workflow presented here was initially developed for SMLM data, all algorithms in PLASMA are applicable to generic MPP data, such as those acquired in seismology, agriculture and spatial omics[350-352].

With advancements in super-resolution imaging, SMLM methods are producing increasingly information-rich data sets. Owing to the development of new environmentally-sensitive probes and imaging methods, localisation data can now be encoded with biophysical membrane properties, such as pH, membrane tension, and quantification of lipid packing. As new polarity-sensitive fluorescent probes are continuously developed and improved, there is greater potential for generating marked point patterns which exceed the standard structural information typically obtained via SMLM and allow for real-time visualisation of lipid-ordered nanodomains. This archetype of big data necessitates specialised analytic techniques for quantification of

marked point patterns, and so the methodologies introduced in PLASMA may be of increasing interest in the foreseeable future.

In order to fully understand the nature of lipid rafts, SMLM modalities must be finetuned to achieve greater precision and accuracy, in terms of both spatial localisation and in photon detection across multiple channels. Further, by developing polarity-sensitive probes with improved quantum efficiency, we will be better able to estimate lipid packing across the plasma membrane, which will ultimately permit more accurate data partitioning and nanodomain identification. As membrane probes with greater balance between binding and dissociation rates are developed, and methods in SMLM become increasingly computationally-efficient, image acquisition times are reduced. This will allow for faster imaging of dense marked point patterns, which provides a clearer representation of atypical lipid packing in live cells, without the bias of experimental artefacts such as drift. In turn, this allows for more precise detection and localisation of lipid nanodomains, and for tracking their dynamics. Through this, we may be able to determine the role and behaviour of lipid rafts in biophysical processes, such as signal transduction, mounting of the immune response and cancer progression, which could have notable clinical impact.

5.3 Long-term outlook

Despite advances in experimental and analytic methods, several fundamental questions regarding the extent and impact of membrane order remain unanswered. The complex interactions between lipid packing, membrane order and membrane curvature, among other biophysical and continuum properties, remain largely unmapped. To further advancement in theoretical and computational membrane modelling, a precise functional relationship must be derived. This is complicated by

the impact of substance and signal transport, across and along the membrane, on membrane lipid order. Cholesterol has recently been shown to have a differential effect on phospholipid and protein classes, with some molecules more sensitive than others[89]. A detailed analysis of the molecular mechanisms underlying the effect of cholesterol-induced membrane re-ordering, inter-leaflet translocation, and membrane asymmetry stabilisation awaits future investigations. To provide in-depth studies of these principles, further development of microscopic, computational, and experimental systems is required. There is no universal solution for experimental design, and each target requires careful probe selection and workflow optimisation – from a biochemical perspective, systemic microscopy would benefit from small, efficient, and quantitative labels, and this will require major collaborative efforts[353]. The increase in spatial resolution afforded by super-resolution microscopy brings trade-offs in other image quality metrics that are equally important for dissecting bioprocesses, such as temporal resolution, sensitivity, imaging speed and phototoxicity[354]. For SMLM, further research is required to increase speed and live cell compatibility, validate deep learning approaches and develop unsupervised machine learning for generalised image reconstruction[355].

The implementation of intelligent hardware systems, such as adaptive optics, may compensate for aberration and prevent excessive imaging artefacts by reducing light doses through selective illumination of the relevant parts of the sample. Self-driving microscopes, which make use of deep reinforcement algorithms and probabilistic optimisation methods, show promise for long-term assay planning and optimal decision making. This will enable rapid, real-time modulation of acquisitions and ensure that sufficient data are captured for all cell states, regardless of underlying heterogeneity[354]. Synergising hardware with deep-learning models will enable

superior pattern recognition, image reconstruction, denoising and information inference. However, more rigorous validation methodologies will be required to establish biological authenticity of computationally-augmented images and the integrity of recovered signal intensities from techniques that enhance data synthetically. Furthermore, strategies to effectively combine diverse datasets, while retaining the specificity of individual experimental conditions, must be developed. This will require major concerted efforts across interdisciplinary research domains. At the very least, these studies await the advancement of both microscope hardware designs and computational software systems, which can leverage deep learning while accurately interpreting model outputs[185]. As modern bio-image data becomes increasingly complex, quantitative analyses and machine learning models are quickly becoming the primary viable avenue for objective study.

The advancement of experimental and computational methods is necessary for further derivation of clinically-relevant results. For instance, correlative analysis may enable research into disorders such as multiple sclerosis, which are characterised by defects in T cell signalling, as well as lipid, cholesterol, and glycosphingolipid metabolism[121]. The precise mechanism by which lipid packing disrupts and engages antigen sensing and receptor signalling remains unclear. Future studies are required to determine the relative contribution of associated cholesterol and lipid species in orchestrating membrane order. Dynamic marked point pattern analysis may enable tracking of lipid rafts, shedding light on the principles of domain migration in the immunological synapse and how this influences the immune response. Further, accurate mapping of receptor nanodomains which interact with cytoskeletal structures, related to the synaptic scaffold, may elucidate understanding of signalling cascades[353]. How membrane remodelling impacts the frequency of interactions between

immunostimulatory receptors, such as Lck, and negative regulators, such as Lck-inhibiting protein Csk, remains to be determined[70]. These experimental challenges will only be overcome with advances to imaging, spectroscopy and controlled cell-activating systems. The aforementioned results would help establish the degree to which lipid-mediated receptor organisation can influence signalling outcome in T cells, and perhaps even its relation to malfunctions such as immune disorders. The impact of receptor nanoscale organisation is now well-established, but immunomodulatory nanomedicine analogues lag behind. Dynamic mathematical modelling of T cell receptors, parameterised by real experimental data, could present clinical applications, such as computationally-optimised immunotherapies.

5.4 Concluding remarks

In this work, we have aimed to develop TDA techniques for microscopy data from micro- to nano-scales and to highlight the benefits of TDA in bioinformatics and microscopy. Analytic method have been introduced for segmenting, modelling and quantifying FM data across the range of spatial resolutions afforded by state-of-the-art imaging modalities. We find that topological image analysis is a viable avenue for segmentation and analysis of data derived from diffraction-limited fluorescence microscopy. With ASMODEUS, we use agent-based modelling to produce advanced PAD simulations which match distributions to target statistics derived from real data. This generates new spatial point patterns which recapitulate the cluster properties of realistic SMLM data and can be used as a framework for predicting hypothetical interactions between molecular species. We have demonstrated that 2-channel ratiometric PAINT SMLM acquisition combined with statistical cluster and colocalisation analysis can serve as a basis for probing and mapping the nanoscale

spatial organisation of ordered lipid domains in mammalian cell membranes. Results suggest that ordered lipid domains segregate from lipid-disordered regions at the nanoscale (~100-300nm). This highlights the potential for further membrane mapping approaches and lays the groundwork for analysis of generalised marked point patterns acquired by SMLM. In summary, this work highlights the importance of topological quantification in non-spatial microscopy data, introduces novel analytic techniques for identifying variations in biophysical membrane properties, and offers further insight into both protein aggregation dynamics and the impact of lipid packing on the plasma membrane.

Bibliography

1 Grecco, Hernán E., Schmick, M. & Bastiaens, Philippe I. H. Signaling from the Living Plasma Membrane. *Cell* **144**, 897-909, doi:https://doi.org/10.1016/j.cell.2011.01.029 (2011).

2 Lin, W.-C. *et al.* H-Ras forms dimers on membrane surfaces via a protein–protein interface. *Proceedings of the National Academy of Sciences* **111**, 2996-3001, doi:10.1073/pnas.1321155111 (2014).

3 Lorent, J. H. *et al.* Plasma membranes are asymmetric in lipid unsaturation, packing and protein shape. *Nat. Chem. Biol.* **16**, 644-652, doi:10.1038/s41589-020-0529-6 (2020).

4 Levental, K. R. *et al.* Lipidomic and biophysical homeostasis of mammalian membranes counteracts dietary lipid perturbations to maintain cellular fitness. *Nature Communications* **11**, 1339, doi:10.1038/s41467-020-15203-1 (2020).

5 Singer, S. J. & Nicolson, G. L. The Fluid Mosaic Model of the Structure of Cell Membranes. *Science* **175**, 720-731, doi:10.1126/science.175.4023.720 (1972).

6 Kazantzis, M. & Stahl, A. Fatty acid transport proteins, implications in physiology and disease. *Biochim. Biophys. Acta* **1821**, 852-857, doi:https://doi.org/10.1016/j.bbalip.2011.09.010 (2012).

7 Breidigan, J., Krzyzanowski, N., Liu, Y., Porcar, L. & Perez-Salas, U. Influence of the membrane environment on cholesterol transfer. *J. Lipid Res.* **58**, 2255–2263, doi:10.1194/jlr.M077909 (2017).

8 Brown, A. J. & Sharpe, L. J. in *Biochemistry of Lipids, Lipoproteins and Membranes (Sixth Edition)* (eds Neale D. Ridgway & Roger S. McLeod) 327-358 (Elsevier, 2016).

9 Li, Y., Park, M. J., Ye, S.-K., Kim, C.-W. & Kim, Y.-N. Elevated Levels of Cholesterol-Rich Lipid Rafts in Cancer Cells Are Correlated with Apoptosis Sensitivity Induced by Cholesterol-Depleting Agents. *The American journal of pathology* **168**, 1107-1118;, doi:10.2353/ajpath.2006.050959 (2006).

10 Cooper, G. M. *The Cell: A Molecular Approach. 2nd edition*. (Sinauer Associates 2000, 2000).

11 Krahn, M. P. Phospholipids of the Plasma Membrane – Regulators or Consequence of Cell Polarity? *Frontiers in Cell and Developmental Biology* **8** (2020).

12 Li, W., Li, F., Zhang, X., Lin, H.-K. & Xu, C. Insights into the post-translational modification and its emerging role in shaping the tumor microenvironment. *Signal Transduction and Targeted Therapy* **6**, 422, doi:10.1038/s41392-021-00825-8 (2021).

13 Goyette, J., Nieves, D. J., Ma, Y. & Gaus, K. How does T cell receptor clustering impact on signal transduction? *J. Cell Sci.* **132**, jcs226423, doi:10.1242/jcs.226423 (2019).

14 Sengupta, P. *et al.* Probing protein heterogeneity in the plasma membrane using PALM and pair correlation analysis. *Nat. Methods* **8**, 969-975, doi:10.1038/nmeth.1704 (2011).

15 Sezgin, E. *et al.* Elucidating membrane structure and protein behavior using giant plasma membrane vesicles. *Nat. Protoc.* **7**, 1042-1051, doi:10.1038/nprot.2012.059 (2012).

16 Kholodenko, B. N., Hancock, J. F. & Kolch, W. Signalling ballet in space and time. *Nature Reviews Molecular Cell Biology* **11**, 414-426, doi:10.1038/nrm2901 (2010).
17 Lillemeier, B. & Klammt, C. How membrane structures control T cell signaling. *Front. Immunol.* **3** (2012).
18 Jaffrès, P.-A. *et al.* Alkyl ether lipids, ion channels and lipid raft reorganization in cancer therapy. *Pharmacol. Ther.* **165**, 114-131, doi:10.1016/j.pharmthera.2016.06.003 (2016).
19 Rossy, J., Ma, Y. & Gaus, K. The organisation of the cell membrane: do proteins rule lipids? *Curr. Opin. Chem. Biol.* **20**, 54-59, doi:https://doi.org/10.1016/j.cbpa.2014.04.009 (2014).
20 Lagache, T., Lang, G., Sauvonnet, N. & Olivo-Marin, J.-C. Analysis of the Spatial Organization of Molecules with Robust Statistics. *PLoS One* **8**, doi:10.1371/journal.pone.0080914 (2013).
21 Nicolson, G. L. & Ferreira de Mattos, G. Fifty Years of the Fluid-Mosaic Model of Biomembrane Structure and Organization and Its Importance in Biomedicine with Particular Emphasis on Membrane Lipid Replacement. *Biomedicines* **10**, 1711 (2022).
22 Leung, S. S. W., Brewer, J., Bagatolli, L. A. & Thewalt, J. L. Measuring molecular order for lipid membrane phase studies: Linear relationship between Laurdan generalized polarization and deuterium NMR order parameter. *Biochim. Biophys. Acta* **1861**, doi:https://doi.org/10.1016/j.bbamem.2019.183053 (2019).
23 Steinkühler, J., Sezgin, E., Urbančič, I., Eggeling, C. & Dimova, R. Mechanical properties of plasma membrane vesicles correlate with lipid order, viscosity and cell density. *Communications Biology* **2**, 337, doi:10.1038/s42003-019-0583-3 (2019).
24 Kusumi, A., Suzuki, K. G. N., Kasai, R. S., Ritchie, K. & Fujiwara, T. K. Hierarchical mesoscale domain organization of the plasma membrane. *Trends in Biochemical Sciences* **36**, 604-615, doi:https://doi.org/10.1016/j.tibs.2011.08.001 (2011).
25 Kusumi, A. & Sako, Y. Cell surface organization by the membrane skeleton. *Curr. Opin. Cell Biol.* **8**, 566-574, doi:https://doi.org/10.1016/S0955-0674(96)80036-6 (1996).
26 Levental, I., Levental, K. R. & Heberle, F. A. Lipid Rafts: Controversies Resolved, Mysteries Remain. *Trends Cell Biol.* **30**, 341-353, doi:https://doi.org/10.1016/j.tcb.2020.01.009 (2020).
27 Levental, I. Lipid rafts come of age. *Nature Reviews Molecular Cell Biology* **21**, 420-420, doi:10.1038/s41580-020-0252-x (2020).
28 Schmid, E. *et al.* Size-dependent protein segregation at membrane interfaces. *Nature Physics* **12**, 704–711, doi:10.1038/nphys3678 (2016).
29 Lillemeier, B. F., Pfeiffer, J. R., Surviladze, Z., Wilson, B. S. & Davis, M. M. Plasma membrane-associated proteins are clustered into islands attached to the cytoskeleton. *Proceedings of the National Academy of Sciences* **103**, 18992-18997, doi:10.1073/pnas.0609009103 (2006).
30 Trimble, W. & Grinstein, S. Barriers to the free diffusion of proteins and lipids in the plasma membrane. *The Journal of Cell Biology* **208**, 259-271, doi:10.1083/jcb.201410071 (2015).

31 Raghupathy, R. *et al.* Transbilayer Lipid Interactions Mediate Nanoclustering of Lipid-Anchored Proteins. *Cell* **161**, 581-594, doi:https://doi.org/10.1016/j.cell.2015.03.048 (2015).
32 Lingwood, D. & Simons, K. Lipid Rafts As a Membrane-Organizing Principle. *Science* **327**, 46-50, doi:10.1126/science.1174621 (2010).
33 Ardito, F., Giuliani, M., Perrone, D., Troiano, G. & Lo Muzio, L. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *Int. J. Mol. Med.* **40**, 271-280, doi:10.3892/ijmm.2017.3036 (2017).
34 Bond, C., Santiago-Ruiz, A. N., Tang, Q. & Lakadamyali, M. Technological advances in super-resolution microscopy to study cellular processes. *Mol. Cell* **82**, 315-332, doi:https://doi.org/10.1016/j.molcel.2021.12.022 (2022).
35 Preta, G. New Insights Into Targeting Membrane Lipids for Cancer Therapy. *Frontiers in Cell and Developmental Biology* **8**, doi:10.3389/fcell.2020.571237 (2020).
36 Kaiser, H.-J. *et al.* Order of lipid phases in model and plasma membranes. *Proceedings of the National Academy of Sciences* **106**, 16645-16650, doi:10.1073/pnas.0908987106 (2009).
37 Panconi, L., Lorenz, C. D., May, R. C., Owen, D. M. & Makarova, M. Phospholipid tail asymmetry allows cellular adaptation to anoxic environments. *J. Biol. Chem.* **299**, doi:10.1016/j.jbc.2023.105134 (2023).
38 Quinn, P. J. & Wolf, C. The liquid-ordered phase in membranes. *Biochim. Biophys. Acta* **1788**, 33-46, doi:https://doi.org/10.1016/j.bbamem.2008.08.005 (2009).
39 Marinescu, D. C. & Marinescu, G. M. in *Classical and Quantum Information* (eds Dan C. Marinescu & Gabriela M. Marinescu) 221-344 (Academic Press, 2012).
40 Chan, L., Morris, G. M. & Hutchison, G. R. Understanding Conformational Entropy in Small Molecules. *J. Chem. Theory Comput.* **17**, 2099-2106, doi:10.1021/acs.jctc.0c01213 (2021).
41 Zhang, X., Barraza, K. M. & Beauchamp, J. L. Cholesterol provides nonsacrificial protection of membrane lipids from chemical damage at air–water interface. *Proceedings of the National Academy of Sciences* **115**, 3255-3260, doi:10.1073/pnas.1722323115 (2018).
42 Yu, W., So, P. T., French, T. & Gratton, E. Fluorescence generalized polarization of cell membranes: a two-photon scanning microscopy approach. *Biophys. J.* **70**, 626-636, doi:https://doi.org/10.1016/S0006-3495(96)79646-7 (1996).
43 Niemelä, P. S. *et al.* Membrane Proteins Diffuse as Dynamic Complexes with Lipids. *Journal of the American Chemical Society* **132**, 7574-7575, doi:10.1021/ja101481b (2010).
44 Lorent, J. H. *et al.* Structural determinants and functional consequences of protein affinity for membrane rafts. *Nature Communications* **8**, 1219, doi:10.1038/s41467-017-01328-3 (2017).
45 Gurdap, C. O., Wedemann, L., Sych, T. & Sezgin, E. Influence of the extracellular domain size on the dynamic behavior of membrane proteins. *Biophys. J.* **121**, 3826-3836, doi:https://doi.org/10.1016/j.bpj.2022.09.010 (2022).
46 Simons, K. & Ikonen, E. Functional rafts in cell membranes. *Nature* **387**, 569-572, doi:10.1038/42408 (1997).

47 Baumgart, T. *et al.* Large-scale fluid/fluid phase separation of proteins and lipids in giant plasma membrane vesicles. *Proceedings of the National Academy of Sciences* **104**, 3165-3170, doi:10.1073/pnas.0611357104 (2007).
48 Sezgin, E., Levental, I., Mayor, S. & Eggeling, C. The mystery of membrane organization: composition, regulation and roles of lipid rafts. *Nature Reviews Molecular Cell Biology* **18**, 361-374, doi:10.1038/nrm.2017.16 (2017).
49 Levental, K. R. & Levental, I. in *Curr. Top. Membr.* Vol. 75 (ed Anne K. Kenworthy) 25-57 (Academic Press, 2015).
50 Saka, S. K. *et al.* Multi-protein assemblies underlie the mesoscale organization of the plasma membrane. *Nature Communications* **5**, 4509, doi:10.1038/ncomms5509 (2014).
51 Sezgin, E. *et al.* Adaptive Lipid Packing and Bioactivity in Membrane Domains. *PLoS One* **10**, e0123930, doi:10.1371/journal.pone.0123930 (2015).
52 Tulodziecka, K. *et al.* Remodeling of the postsynaptic plasma membrane during neural development. *Mol. Biol. Cell* **27**, 3480-3489, doi:10.1091/mbc.E16-06-0420 (2016).
53 Baumgart, T., Hess, S. T. & Webb, W. W. Imaging coexisting fluid domains in biomembrane models coupling curvature and line tension. *Nature* **425**, 821-824, doi:10.1038/nature02013 (2003).
54 Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nature Reviews Molecular Cell Biology* **18**, 285-298, doi:10.1038/nrm.2017.7 (2017).
55 Snead, W. T. *et al.* Membrane surfaces regulate assembly of ribonucleoprotein condensates. *Nat. Cell Biol.* **24**, 461-470, doi:10.1038/s41556-022-00882-3 (2022).
56 Mangiarotti, A. *et al.* Biomolecular condensates modulate membrane lipid packing and hydration. *Nature Communications* **14**, 6081, doi:10.1038/s41467-023-41709-5 (2023).
57 Mangiarotti, A., Chen, N., Zhao, Z., Lipowsky, R. & Dimova, R. Wetting and complex remodeling of membranes by biomolecular condensates. *Nature Communications* **14**, 2809, doi:10.1038/s41467-023-37955-2 (2023).
58 Parton, R. G. & Hancock, J. F. Lipid rafts and plasma membrane microorganization: insights from Ras. *Trends Cell Biol.* **14 3**, 141-147 (2004).
59 Szlasa, W., Zendran, I., Zalesińska, A., Tarek, M. & Kulbacka, J. Lipid composition of the cancer cell membrane. *Journal of Bioenergetics and Biomembranes* **52**, 321–342, doi:10.1007/s10863-020-09846-4 (2020).
60 Beloribi-Djefaflia, S., Vasseur, S. & Guillaumond, F. Lipid metabolic reprogramming in cancer cells. *Oncogenesis* **5**, 189, doi:10.1038/oncsis.2015.49 (2016).
61 Ros-Baró, A. *et al.* Lipid rafts are required for GLUT4 internalization in adipose cells. *Proceedings of the National Academy of Sciences* **98**, 12050-12055, doi:10.1073/pnas.211341698 (2001).
62 Chamberlain, L. H., Burgoyne, R. D. & Gould, G. W. SNARE proteins are highly enriched in lipid rafts in PC12 cells: Implications for the spatial control of exocytosis. *Proceedings of the National Academy of Sciences* **98**, 5619-5624, doi:10.1073/pnas.091502398 (2001).
63 Lajoie, P. & Nabi, I. R. Regulation of raft-dependent endocytosis. *Journal of Cellular and Molecular Medicine* **11**, 644 - 653 (2007).
64 Lamerton, R. E., Lightfoot, A., Nieves, D. J. & Owen, D. M. The Role of Protein and Lipid Clustering in Lymphocyte Activation. *Front. Immunol.* **12** (2021).

65 Purtic, B., Pitcher, L. A., van Oers, N. S. C. & Wülfing, C. T cell receptor (TCR) clustering in the immunological synapse integrates TCR and costimulatory signaling in selected T cells. *Proceedings of the National Academy of Sciences* **102**, 2904-2909, doi:10.1073/pnas.0406867102 (2005).
66 Sherman, E., Barr, V. & Samelson, L. E. Super-resolution characterization of TCR-dependent signaling clusters. *Immunol. Rev.* **251**, 21-35, doi:https://doi.org/10.1111/imr.12010 (2013).
67 Yuan, Y. *et al.* Single-Molecule Super-Resolution Imaging of T-Cell Plasma Membrane CD4 Redistribution upon HIV-1 Binding. *Viruses* **13**, 142 (2021).
68 Gaus, K., Chklovskaia, E., Fazekas de St. Groth, B., Jessup, W. & Harder, T. Condensation of the plasma membrane at the site of T lymphocyte activation. *J. Cell Biol.* **171**, 121-131, doi:10.1083/jcb.200505047 (2005).
69 Courtney, A. H. *et al.* CD45 functions as a signaling gatekeeper in T cells. *Science Signaling* **12**, 8151, doi:10.1126/scisignal.aaw8151 (2019).
70 Urbančič, I. *et al.* Aggregation and mobility of membrane proteins interplay with local lipid order in the plasma membrane of T cells. *FEBS Lett.* **595**, 2127-2146, doi:https://doi.org/10.1002/1873-3468.14153 (2021).
71 Janes, P. W., Ley, S. C. & Magee, A. I. Aggregation of Lipid Rafts Accompanies Signaling via the T Cell Antigen Receptor. *J. Cell Biol.* **147**, 447-461, doi:10.1083/jcb.147.2.447 (1999).
72 Wang, H.-Y. *et al.* Coupling of protein condensates to ordered lipid domains determines functional membrane organization. *Science Advances* **9**, 6205, doi:10.1126/sciadv.adf6205 (2023).
73 Schieffer, D., Naware, S., Bakun, W. & Bamezai, A. K. Lipid raft-based membrane order is important for antigen-specific clonal expansion of CD4+ T lymphocytes. *BMC Immunol.* **15**, 58, doi:10.1186/s12865-014-0058-8 (2014).
74 Fang, T. *et al.* Spatial Regulation of T-Cell Signaling by Programmed Death-Ligand 1 on Wireframe DNA Origami Flat Sheets. *ACS Nano* **15**, 3441-3452, doi:10.1021/acsnano.0c10632 (2021).
75 Fernandes, R. A. *et al.* A cell topography-based mechanism for ligand discrimination by the T cell receptor. *Proceedings of the National Academy of Sciences* **116**, 14002-14010, doi:10.1073/pnas.1817255116 (2019).
76 Simons, K. & Toomre, D. Lipid rafts and signal transduction. *Nature Reviews Molecular Cell Biology* **1**, 31-39, doi:10.1038/35036052 (2000).
77 Shivanandan, A., Unnikrishnan, J. & Radenovic, A. Accounting for Limited Detection Efficiency and Localization Precision in Cluster Analysis in Single Molecule Localization Microscopy. *PLoS One* **10**, e0118767, doi:10.1371/journal.pone.0118767 (2015).
78 Mugler, A., Tostevin, F. & ten Wolde, P. R. Spatial partitioning improves the reliability of biochemical signaling. *Proceedings of the National Academy of Sciences* **110**, 5927-5932, doi:10.1073/pnas.1218301110 (2013).
79 Ramadurai, S. *et al.* Lateral Diffusion of Membrane Proteins. *Journal of the American Chemical Society* **131**, 12650-12656, doi:10.1021/ja902853g (2009).
80 Filippov, A., Orädd, G. & Lindblom, G. Lipid Lateral Diffusion in Ordered and Disordered Phases in Raft Mixtures. *Biophys. J.* **86**, 891-896, doi:10.1016/S0006-3495(04)74164-8 (2004).
81 Dietrich, C. *et al.* Lipid Rafts Reconstituted in Model Membranes. *Biophys. J.* **80**, 1417-1428, doi:10.1016/S0006-3495(01)76114-0 (2001).
82 Beckers, D., Urbancic, D. & Sezgin, E. Impact of Nanoscale Hindrances on the Relationship between Lipid Packing and Diffusion in Model Membranes. *The*

Journal of Physical Chemistry B **124**, 1487-1494, doi:10.1021/acs.jpcb.0c00445 (2020).
83 Buwaneka, P., Ralko, A., Liu, S.-L. & Cho, W. Evaluation of the available cholesterol concentration in the inner leaflet of the plasma membrane of mammalian cells. *J. Lipid Res.* **62**, 100084, doi:10.1016/j.jlr.2021.100084 (2021).
84 Ogasawara, F. *et al.* Changes in the asymmetric distribution of cholesterol in the plasma membrane influence streptolysin O pore formation. *Sci. Rep.* **9**, 4548, doi:10.1038/s41598-019-39973-x (2019).
85 Pinkwart, K. *et al.* Nanoscale dynamics of cholesterol in the cell membrane. *J. Biol. Chem.* **294**, 12599-12609, doi:https://doi.org/10.1074/jbc.RA119.009683 (2019).
86 Buyan, A., Allender, D. W., Corry, B. & Schick, M. Lipid redistribution in the highly curved footprint of Piezo1. *Biophys. J.* **122**, 1900-1913, doi:10.1016/j.bpj.2022.07.022 (2023).
87 Girard, M. & Bereau, T. Induced asymmetries in membranes. *Biophys. J.* **122**, 2092-2098, doi:10.1016/j.bpj.2022.12.004 (2023).
88 Risselada, H. Cholesterol: The Plasma Membrane's Constituent that Chooses Sides. *Biophys. J.* **116**, 2235–2236, doi:10.1016/j.bpj.2019.05.003 (2019).
89 Menon, I. *et al.* A cholesterol switch controls phospholipid scrambling by G-protein-coupled receptors. *J. Biol. Chem.* **300**, 105649, doi:10.1016/j.jbc.2024.105649 (2024).
90 Castillo, S. R., Nguyen, M. H. L., DiPasquale, M., Kelley, E. G. & Marquardt, D. Mitocans induce lipid flip-flop and permeabilize the membrane to signal apoptosis. *Biophys. J.* **122**, 2353-2366, doi:https://doi.org/10.1016/j.bpj.2023.03.039 (2023).
91 Shibata, N. *et al.* Atomic resolution electron microscopy in a magnetic field free environment. *Nature Communications* **10**, 2308, doi:10.1038/s41467-019-10281-2 (2019).
92 Oreopoulos, J. & Yip, C. M. Probing Membrane Order and Topography in Supported Lipid Bilayers by Combined Polarized Total Internal Reflection Fluorescence-Atomic Force Microscopy. *Biophys. J.* **96**, 1970-1984, doi:10.1016/j.bpj.2008.11.041 (2009).
93 Lakowicz, J. *Principles of Fluorescence Spectroscopy*. Vol. 1 (2006).
94 Endesfelder, U., van de Linde, S., Wolter, S., Sauer, M. & Heilemann, M. Subdiffraction-Resolution Fluorescence Microscopy of Myosin–Actin Motility. *Chemphyschem* **11**, 836-840, doi:https://doi.org/10.1002/cphc.200900944 (2010).
95 Lekha, P. *et al.* A hidden Markov model approach to characterizing the photo-switching behavior of fluorophores. *The Annals of Applied Statistics* **13**, 1397-1429, doi:10.1214/19-AOAS1240 (2019).
96 Lelek, M. *et al.* Single-molecule localization microscopy. *Nature Reviews Methods Primers* **1**, 39, doi:10.1038/s43586-021-00038-x (2021).
97 Platzer, R. *et al.* Unscrambling fluorophore blinking for comprehensive cluster detection via photoactivated localization microscopy. *Nature Communications* **11**, 4993, doi:10.1038/s41467-020-18726-9 (2020).
98 Waters, J. C. Accuracy and precision in quantitative fluorescence microscopy. *J. Cell Biol.* **185**, 1135-1148, doi:10.1083/jcb.200903097 (2009).

99 Juhasz, J., Davis, James H. & Sharom, Frances J. Fluorescent probe partitioning in giant unilamellar vesicles of 'lipid raft' mixtures. *Biochemical Journal* **430**, 415-423, doi:10.1042/BJ20100516 (2010).
100 Jin, L., Millard, A., Wuskell, J., Clark, H. & Loew, L. Cholesterol-Enriched Lipid Domains Can Be Visualized by di-4-ANEPPDHQ with Linear and Nonlinear Optics. *Biophys. J.* **89**, 4-6, doi:10.1529/biophysj.105.064816 (2005).
101 Sezgin, E. *et al.* Partitioning, Diffusion, and Ligand Binding of Raft Lipid Analogs in Model and Cellular Plasma Membranes. *Biochim. Biophys. Acta* **1818**, 1777-1784, doi:10.1016/j.bbamem.2012.03.007 (2012).
102 Loudet, A. & Burgess, K. BODIPY Dyes and Their Derivatives: Syntheses and Spectroscopic Properties. *Chem. Rev.* **107**, 4891-4932, doi:10.1021/cr078381n (2007).
103 Lindner, B., Burkard, T. & Schuler, M. Phagocytosis assays with different pH-sensitive fluorescent particles and various readouts. *Biotechniques* **68**, 245-250, doi:10.2144/btn-2020-0003 (2020).
104 Deo, C. Hybrid Fluorescent Probes for Imaging Membrane Tension Inside Living Cells. *ACS Central Science* **6**, 1285-1287, doi:10.1021/acscentsci.0c00977 (2020).
105 Owen, D. & Gaus, K. Optimized time-gated generalized polarization imaging of Laurdan and di-4-ANEPPDHQ for membrane order image contrast enhancement. *Microscopy research and technique* **73**, 618-622, doi:10.1002/jemt.20801 (2009).
106 Ashdown, G. & Owen, D. Imaging Membrane Order Using Environmentally Sensitive Fluorophores. *Methods in molecular biology (Clifton, N.J.)* **1232**, 115-122, doi:10.1007/978-1-4939-1752-5_10 (2015).
107 Danylchuk, D. I., Jouard, P.-H. & Klymchenko, A. S. Targeted Solvatochromic Fluorescent Probes for Imaging Lipid Order in Organelles under Oxidative and Mechanical Stress. *Journal of the American Chemical Society* **143**, 912-924, doi:10.1021/jacs.0c10972 (2021).
108 Owen, D. M., Rentero, C., Magenau, A., Abu-Siniyeh, A. & Gaus, K. Quantitative imaging of membrane lipid order in cells and organisms. *Nat. Protoc.* **7**, 24-35 (2011).
109 Sackett, D. L. & Wolff, J. Nile red as a polarity-sensitive fluorescent probe of hydrophobic protein surfaces. *Anal. Biochem.* **167**, 228-234, doi:https://doi.org/10.1016/0003-2697(87)90157-6 (1987).
110 Zhanghao, K. *et al.* High-dimensional super-resolution imaging reveals heterogeneity and dynamics of subcellular lipid membranes. *Nature Communications* **11**, 5890, doi:10.1038/s41467-020-19747-0 (2020).
111 Kucherak, O. A. *et al.* Switchable Nile Red-Based Probe for Cholesterol and Lipid Order at the Outer Leaflet of Biomembranes. *Journal of the American Chemical Society* **132**, 4907-4916, doi:10.1021/ja100351w (2010).
112 Moon, S. *et al.* Spectrally Resolved, Functional Super-Resolution Microscopy Reveals Nanoscale Compositional Heterogeneity in Live-Cell Membranes. *Journal of the American Chemical Society* **139**, 10944-10947, doi:10.1021/jacs.7b03846 (2017).
113 Iaea, D. B. & Maxfield, F. R. Membrane order in the plasma membrane and endocytic recycling compartment. *PLoS One* **12**, doi:10.1371/journal.pone.0188041 (2017).

114 Mazeres, S., Joly, E., López Albores, A. & Tardin, C. Characterization of M-laurdan, a versatile probe to explore order in lipid membranes. *F1000Research* **3**, 172, doi:10.12688/f1000research.4805.2 (2014).
115 Viard, M. *et al.* Laurdan solvatochromism: solvent dielectric relaxation and intramolecular excited-state reaction. *Biophys. J.* **73**, 2221-2234, doi:10.1016/s0006-3495(97)78253-5 (1997).
116 Dodes Traian, M. M., Flecha, F. L. G. & Levi, V. Imaging lipid lateral organization in membranes with C-laurdan in a confocal microscope. *J. Lipid Res.* **53**, 609 - 616 (2012).
117 Sezgin, E., Sadowski, T. & Simons, K. Measuring Lipid Packing of Model and Cellular Membranes with Environment Sensitive Probes. *Langmuir* **30**, 8160-8166, doi:10.1021/la501226v (2014).
118 Färber, N. & Westerhausen, C. Broad lipid phase transitions in mammalian cell membranes measured by Laurdan fluorescence spectroscopy. *Biochim. Biophys. Acta* **1864**, 183794, doi:https://doi.org/10.1016/j.bbamem.2021.183794 (2022).
119 Jin, L. *et al.* Characterization and application of a new optical probe for membrane lipid domains. *Biophys J* **90**, 2563-2575, doi:10.1529/biophysj.105.072884 (2006).
120 Sengupta, S., Karsalia, R., Morrissey, A. & Bamezai, A. K. Cholesterol-dependent plasma membrane order (Lo) is critical for antigen-specific clonal expansion of CD4+ T cells. *Sci. Rep.* **11**, 13970, doi:10.1038/s41598-021-93403-5 (2021).
121 Waddington, K. E. *et al.* LXR directly regulates glycosphingolipid synthesis and affects human CD4+ T cell function. *Proceedings of the National Academy of Sciences* **118**, doi:10.1073/pnas.2017394118 (2021).
122 Panconi, L., Makarova, M., Lambert, E. R., May, R. C. & Owen, D. M. Topology-based fluorescence image analysis for automated cell identification and segmentation. *Journal of Biophotonics* **16**, doi:https://doi.org/10.1002/jbio.202200199 (2022).
123 Amaro, M., Reina, F., Hof, M., Eggeling, C. & Sezgin, E. Laurdan and Di-4-ANEPPDHQ probe different properties of the membrane. *J. Phys. D: Appl. Phys.* **50**, 134004, doi:10.1088/1361-6463/aa5dbc (2017).
124 Feigenson, G. W. & Buboltz, J. T. Ternary Phase Diagram of Dipalmitoyl-PC/Dilauroyl-PC/Cholesterol: Nanoscopic Domain Formation Driven by Cholesterol. *Biophys. J.* **80**, 2775-2788, doi:https://doi.org/10.1016/S0006-3495(01)76245-5 (2001).
125 Ragaller, F. *et al.* Dissecting the mechanisms of environment sensitivity of smart probes for quantitative assessment of membrane properties. *Open Biology* **12**, 220175, doi:10.1098/rsob.220175 (2022).
126 Dinic, J., Biverståhl, H., Mäler, L. & Parmryd, I. Laurdan and di-4-ANEPPDHQ do not respond to membrane-inserted peptides and are good probes for lipid packing. *Biochim. Biophys. Acta* **1808**, 298-306, doi:https://doi.org/10.1016/j.bbamem.2010.10.002 (2011).
127 Renz, M. Fluorescence microscopy—A historical and technical perspective. *Cytometry Part A* **83**, 767-779, doi:https://doi.org/10.1002/cyto.a.22295 (2013).
128 Khater, I. M., Nabi, I. R. & Hamarneh, G. A Review of Super-Resolution Single-Molecule Localization Microscopy Cluster Analysis and Quantification Methods. *Patterns* **1**, 100038, doi:https://doi.org/10.1016/j.patter.2020.100038 (2020).

129 Nieves, D. J., Gaus, K. & Baker, M. A. B. DNA-Based Super-Resolution Microscopy: DNA-PAINT. *Genes* **9**, 621 (2018).
130 Dai, M., Jungmann, R. & Yin, P. Optical imaging of individual biomolecules in densely packed clusters. *Nature Nanotechnology* **11**, 798-807, doi:10.1038/nnano.2016.95 (2016).
131 Ries, J. SMAP: a modular super-resolution microscopy analysis platform for SMLM data. *Nat Methods* **17**, 870-872, doi:10.1038/s41592-020-0938-1 (2020).
132 Gwosch, K. C. *et al.* MINFLUX nanoscopy delivers 3D multicolor nanometer resolution in cells. *Nat. Methods* **17**, 217-224, doi:10.1038/s41592-019-0688-0 (2020).
133 Young, M. R. Principles and Technique of Fluorescence Microscopy. *J. Cell Sci.* **102**, 419-449, doi:10.1242/jcs.s3-102.60.419 (1961).
134 Fero, M. & Pogliano, K. Automated Quantitative Live Cell Fluorescence Microscopy. *Cold Spring Harb. Perspect. Biol.* **2**, doi:10.1101/cshperspect.a000455 (2010).
135 Sanderson, M., Smith, I., Parker, I. & Bootman, M. Fluorescence Microscopy. *Cold Spring Harbor Protocols* **10**, doi:10.1101/pdb.top071795 (2014).
136 Zhao, X., Li, R., Lu, C., Baluška, F. & Wan, Y. Di-4-ANEPPDHQ, a fluorescent probe for the visualisation of membrane microdomains in living Arabidopsis thaliana cells. *Plant Physiology and Biochemistry* **87**, 53-60, doi:https://doi.org/10.1016/j.plaphy.2014.12.015 (2015).
137 Blázquez-Castro, A. & Stockert, J. *Fluorescence Microscopy in Life Sciences*. (2017).
138 Peters, R. *Quantification of actin nanoarchitecture at the T cell immunological synapse using single molecule localisation microscopy*, King's College London, (2018).
139 Huang, B., Bates, M. & Zhuang, X. Super-Resolution Fluorescence Microscopy. *Annual Review of Biochemistry* **78**, 993-1016, doi:https://doi.org/10.1146/annurev.biochem.77.061906.092014 (2009).
140 Bianco, B. & Diaspro, A. Analysis of three-dimensional cell imaging obtained with optical microscopy techniques based on defocusing. *Cell Biophys.* **15**, 189-199, doi:10.1007/BF02989683 (1989).
141 Heilemann, M. *et al.* Subdiffraction-Resolution Fluorescence Imaging with Conventional Fluorescent Probes. *Angew. Chem. Int. Ed.* **47**, 6172-6176, doi:https://doi.org/10.1002/anie.200802376 (2008).
142 Gray, N. Knowing the limit. *Nat. Cell Biol.* **11**, doi:10.1038/ncb1940 (2009).
143 Stemmer, A., Beck, M. & Fiolka, R. Widefield fluorescence microscopy with extended resolution. *Histochem. Cell Biol.* **130**, 807-817, doi:10.1007/s00418-008-0506-8 (2008).
144 Neil, M. A. A., Juškaitis, R. & Wilson, T. Real time 3D fluorescence microscopy by two beam interference illumination. *Opt. Commun.* **153**, 1-4, doi:https://doi.org/10.1016/S0030-4018(98)00210-7 (1998).
145 Mondal, P. P. Temporal resolution in fluorescence imaging. *Frontiers in Molecular Biosciences* **1**, 11 (2014).
146 Fouquet, C. *et al.* Improving Axial Resolution in Confocal Microscopy with New High Refractive Index Mounting Media. *PLoS One* **10**, doi:10.1371/journal.pone.0121096 (2015).

147 Wu, Y.-L., Tschanz, A., Krupnik, L. & Ries, J. Quantitative Data Analysis in Single-Molecule Localization Microscopy. *Trends Cell Biol.* **30**, 837-851, doi:10.1016/j.tcb.2020.07.005 (2020).
148 Goyette, J., Nieves, D., Ma, Y. & Gaus, K. How does T cell receptor clustering impact on signal transduction? *J. Cell Sci.* **132**, doi:10.1242/jcs.226423 (2019).
149 Lidke, D. S. *et al.* Quantum dot ligands provide new insights into erbB/HER receptor–mediated signal transduction. *Nat. Biotechnol.* **22**, 198-203, doi:10.1038/nbt929 (2004).
150 Thompson, R. E., Larson, D. R. & Webb, W. W. Precise Nanometer Localization Analysis for Individual Fluorescent Probes. *Biophys. J.* **82**, 2775-2783, doi:https://doi.org/10.1016/S0006-3495(02)75618-X (2002).
151 Ober, R., Ram, S. & Ward, E. Localization Accuracy in Single-Molecule Microscopy. *Biophys. J.* **86**, 1185-1200, doi:10.1016/S0006-3495(04)74193-4 (2004).
152 Klymchenko, A. S. Solvatochromic and Fluorogenic Dyes as Environment-Sensitive Probes: Design and Biological Applications. *Acc. Chem. Res.* **50**, 366-375, doi:10.1021/acs.accounts.6b00517 (2017).
153 Owen, D. M., Williamson, D., Magenau, A., Rossy, J. & Gaus, K. in *Methods Enzymol.* Vol. 504 (ed P. Michael conn) 221-235 (Academic Press, 2012).
154 Betzig, E. *et al.* Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science* **313**, 1642-1645, doi:10.1126/science.1127344 (2006).
155 van de Linde, S. *et al.* Direct stochastic optical reconstruction microscopy with standard fluorescent probes. *Nat. Protoc.* **6**, 991-1009, doi:10.1038/nprot.2011.336 (2011).
156 Colombo, J. *et al.* A functional family of fluorescent nucleotide analogues to investigate actin dynamics and energetics. *Nature Communications* **12**, 548, doi:10.1038/s41467-020-20827-4 (2021).
157 Albertazzi, L. & Heilemann, M. When Weak Is Strong: A Plea for Low-Affinity Binders for Optical Microscopy. *Angew. Chem. Int. Ed.* **62**, doi:https://doi.org/10.1002/anie.202303390 (2023).
158 Jungmann, R. *et al.* Quantitative super-resolution imaging with qPAINT. *Nat. Methods* **13**, 439-442, doi:10.1038/nmeth.3804 (2016).
159 Jungmann, R. *et al.* Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. *Nat. Methods* **11**, 313-318, doi:10.1038/nmeth.2835 (2014).
160 Schueder, F. *et al.* Super-Resolution Spatial Proximity Detection with Proximity-PAINT. *Angew. Chem. Int. Ed.* **60**, 716-720, doi:https://doi.org/10.1002/anie.202009031 (2021).
161 Ostersehlt, L. M. *et al.* DNA-PAINT MINFLUX nanoscopy. *Nat. Methods* **19**, 1072-1075, doi:10.1038/s41592-022-01577-1 (2022).
162 Narayanasamy, K. K., Rahm, J. V., Tourani, S. & Heilemann, M. Fast DNA-PAINT imaging using a deep neural network. *Nature Communications* **13**, 5047, doi:10.1038/s41467-022-32626-0 (2022).
163 Kessler, L. F. *et al.* Self-quenched Fluorophore Dimers for DNA-PAINT and STED Microscopy. *Angew. Chem. Int. Ed.* **62**, doi:https://doi.org/10.1002/anie.202307538 (2023).
164 Schueder, F. *et al.* An order of magnitude faster DNA-PAINT imaging by optimized sequence design and buffer conditions. *Nat. Methods* **16**, 1101-1104, doi:10.1038/s41592-019-0584-7 (2019).

165 Jang, S. *et al.* Neural network-assisted single-molecule localization microscopy with a weak-affinity protein tag. *Biophysical Reports* **3**, 100123, doi:https://doi.org/10.1016/j.bpr.2023.100123 (2023).
166 Manders, E. M. M., Verbeek, F. J. & Aten, J. A. Measurement of co-localization of objects in dual-colour confocal images. *J. Microsc.* **169**, 375-382, doi:https://doi.org/10.1111/j.1365-2818.1993.tb03313.x (1993).
167 Wu, Y.-L. *et al.* Maximum-likelihood model fitting for quantitative analysis of SMLM data. *Nat. Methods* **20**, 139-148, doi:10.1038/s41592-022-01676-z (2023).
168 Bates, M. *et al.* Optimal precision and accuracy in 4Pi-STORM using dynamic spline PSF models. *Nat. Methods* **19**, 603-612, doi:10.1038/s41592-022-01465-8 (2022).
169 Endesfelder, U., Malkusch, S., Fricke, F. & Heilemann, M. A simple method to estimate the average localization precision of a single-molecule localization microscopy experiment. *Histochem. Cell Biol.* **141**, 629-638, doi:10.1007/s00418-014-1192-3 (2014).
170 Cohen, E. A. K., Abraham, A. V., Ramakrishnan, S. & Ober, R. J. Resolution limit of image analysis algorithms. *Nature Communications* **10**, 793, doi:10.1038/s41467-019-08689-x (2019).
171 Bongiovanni, M. N. *et al.* Multi-dimensional super-resolution imaging enables surface hydrophobicity mapping. *Nature Communications* **7**, 13544, doi:10.1038/ncomms13544 (2016).
172 Small, A. Theoretical limits on errors and acquisition rates in localizing switchable fluorophores. *Biophys. J.* **96**, 16-18 (2008).
173 Demmerle, J., Wegel, E., Schermelleh, L. & Dobbie, I. M. Assessing resolution in super-resolution imaging. *Methods* **88**, 3-10, doi:https://doi.org/10.1016/j.ymeth.2015.07.001 (2015).
174 Wester, M. J. *et al.* Robust, fiducial-free drift correction for super-resolution imaging. *Sci. Rep.* **11**, 23672, doi:10.1038/s41598-021-02850-7 (2021).
175 Leung, B. O. & Chou, K. C. Review of Super-Resolution Fluorescence Microscopy for Biology. *Appl. Spectrosc.* **65**, 967-980, doi:10.1366/11-06398 (2011).
176 Jensen, L. G. *et al.* Correction of multiple-blinking artifacts in photoactivated localization microscopy. *Nat. Methods* **19**, 594-602, doi:10.1038/s41592-022-01463-w (2022).
177 Shivanandan, A., Deschout, H., Scarselli, M. & Radenovic, A. Challenges in quantitative single molecule localization microscopy. *FEBS Lett.* **588**, 3595-3602, doi:https://doi.org/10.1016/j.febslet.2014.06.014 (2014).
178 Fazel, M. *et al.* Bayesian Multiple Emitter Fitting using Reversible Jump Markov Chain Monte Carlo. *Sci. Rep.* **9**, 13791, doi:10.1038/s41598-019-50232-x (2019).
179 Ovesný, M., Křížek, P., Borkovec, J., Švindrych, Z. & Hagen, G. M. ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics* **30**, 2389-2390, doi:10.1093/bioinformatics/btu202 (2014).
180 Nyquist, H. Certain Topics in Telegraph Transmission Theory. *Transactions of the American Institute of Electrical Engineers* **47**, 617-644, doi:10.1109/T-AIEE.1928.5055024 (1928).
181 Shannon, C. E. Communication in the Presence of Noise. *Proceedings of the IRE* **37**, 10-21, doi:10.1109/JRPROC.1949.232969 (1949).

182 Speiser, A. *et al.* Deep learning enables fast and dense single-molecule localization with high accuracy. *Nat. Methods* **18**, 1082-1090, doi:10.1038/s41592-021-01236-x (2021).
183 Saguy, A. *et al.* DBlink: dynamic localization microscopy in super spatiotemporal resolution via deep learning. *Nat. Methods* **20**, 1939-1948, doi:10.1038/s41592-023-01966-0 (2023).
184 Möckl, L., Roy, A. R. & Moerner, W. E. Deep learning in single-molecule microscopy: fundamentals, caveats, and recent developments [Invited]. *Biomed. Opt. Express* **11**, 1633-1661, doi:10.1364/BOE.386361 (2020).
185 G'omez-de-Mariscal, E., Del Rosario, M., Pylvänäinen, J. W., Jacquemet, G. & Henriques, R. Harnessing artificial intelligence to reduce phototoxicity in live imaging. *J. Cell Sci.* **137** (2023).
186 von Chamier, L. *et al.* Democratising deep learning for microscopy with ZeroCostDL4Mic. *Nature Communications* **12**, 2276, doi:10.1038/s41467-021-22518-0 (2021).
187 Eggeling, C. *et al.* Direct observation of the nanoscale dynamics of membrane lipids in a living cell. *Nature* **457**, 1159-1162, doi:10.1038/nature07596 (2009).
188 Sevcsik, E. *et al.* GPI-anchored proteins do not reside in ordered domains in the live cell plasma membrane. *Nature Communications* **6**, 6969, doi:10.1038/ncomms7969 (2015).
189 Liu, Y. *et al.* The involvement of lipid rafts in epidermal growth factor-induced chemotaxis of breast cancer cells. *Mol. Membr. Biol.* **24**, 91-101, doi:10.1080/10929080600990500 (2007).
190 Sako, Y., Minoghchi, S. & Yanagida, T. Single-molecule imaging of EGFR signalling on the surface of living cells. *Nat. Cell Biol.* **2**, 168-172, doi:10.1038/35004044 (2000).
191 Owen, D. *et al.* PALM imaging and cluster analysis of protein heterogeneity at the cell surface. *Journal of Biophotonics* **3**, 446-454, doi:10.1002/jbio.200900089 (2010).
192 Kiskowski, M., Hancock, J. & Kenworthy, A. On the Use of Ripley's K-Function and Its Derivatives to Analyze Domain Size. *Biophys. J.* **97**, 1095-1103, doi:10.1016/j.bpj.2009.05.039 (2009).
193 Sengupta, P., Jovanovic-Talisman, T. & Lippincott-Schwartz, J. Quantifying spatial organization in point-localization superresolution images using pair correlation analysis. *Nat. Protoc.* **8**, 345-354, doi:10.1038/nprot.2013.005 (2013).
194 Xie, J., Girshick, R. & Farhadi, A. Unsupervised deep embedding for clustering analysis. *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, 478–487 (2016).
195 Shivanandan, A., Unnikrishnan, J. & Radenovic, A. On characterizing protein spatial clusters with correlation approaches. *Sci. Rep.* **6**, 31164, doi:10.1038/srep31164 (2016).
196 Nieves, D. J. *et al.* A framework for evaluating the performance of SMLM cluster analysis algorithms. *Nat. Methods* **20**, 259-267, doi:10.1038/s41592-022-01750-6 (2023).
197 Goyette, J. & Gaus, K. Mechanisms of protein nanoscale clustering. *Curr. Opin. Cell Biol.* **44**, 86-92, doi:https://doi.org/10.1016/j.ceb.2016.09.004 (2017).
198 Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. in *Knowledge Discovery and Data Mining.* 226–231.

199 Pike, J. A. *et al.* Topological data analysis quantifies biological nano-structure from single molecule localization microscopy. *Bioinformatics* **36**, 1614-1621, doi:10.1093/bioinformatics/btz788 (2020).
200 Mondal, P. Probabilistic Optically-Selective Single-molecule Imaging Based Localization Encoded (POSSIBLE) microscopy for ultra-superresolution imaging. *PLoS One* **15**, doi:10.1371/journal.pone.0242452 (2020).
201 Fujita, A., Takahashi, D. Y. & Patriota, A. G. A non-parametric method to estimate the number of clusters. *Comput. Stat. Data Anal.* **73**, 27-39, doi:https://doi.org/10.1016/j.csda.2013.11.012 (2014).
202 Dudoit, S. & Fridlyand, J. A Prediction-Based Resampling Method for Estimating the Number of Clusters in a Data Set. *Genome Biol.* **3**, doi:10.1186/gb-2002-3-7-research0036 (2002).
203 Veatch, S. L. *et al.* Correlation Functions Quantify Super-Resolution Images and Estimate Apparent Clustering Due to Over-Counting. *PLoS One* **7**, e31457, doi:10.1371/journal.pone.0031457 (2012).
204 Rand, W. M. Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association* **66**, 846-850, doi:10.2307/2284239 (1971).
205 Hubert, L. & Arabie, P. Comparing partitions. *Journal of Classification* **2**, 193-218, doi:10.1007/BF01908075 (1985).
206 Margalit, A. & Knott, G. An algorithm for computing the union, intersection or difference of two polygons. *Computers & Graphics* **13**, 167-183, doi:10.1016/0097-8493(89)90059-9 (1989).
207 Wakefield, D. L. *et al.* Using quantitative single molecule localization microscopy to optimize multivalent HER2-targeting ligands. *Frontiers in Medicine* **10** (2023).
208 Griffié, J. *et al.* A Bayesian cluster analysis method for single-molecule localization microscopy data. *Nat. Protoc.* **11**, 2499-2514, doi:10.1038/nprot.2016.149 (2016).
209 Rubin-Delanchy, P. *et al.* Bayesian cluster identification in single-molecule localization microscopy data. *Nat. Methods* **12**, 1072-1076, doi:10.1038/nmeth.3612 (2015).
210 Griffié, J. *et al.* 3D Bayesian cluster analysis of super-resolution data reveals LAT recruitment to the T cell synapse. *Sci. Rep.* **7**, 4077, doi:10.1038/s41598-017-04450-w (2017).
211 Matioli, L. C., Santos, S. R., Kleina, M. & Leite, E. A. A new algorithm for clustering based on kernel density estimation. *Journal of Applied Statistics* **45**, 347-366, doi:10.1080/02664763.2016.1277191 (2018).
212 Mazouchi, A. & Milstein, J. N. Fast Optimized Cluster Algorithm for Localizations (FOCAL): a spatial cluster analysis for super-resolved microscopy. *Bioinformatics* **32**, 747-754, doi:10.1093/bioinformatics/btv630 (2016).
213 Joseph, M., Bort, E., Grose, R., McCormick, P. & Simoncelli, S. Quantitative Super-Resolution Imaging for the Analysis of GPCR Oligomerization. *Biomolecules* **11**, 1503, doi:10.3390/biom11101503 (2021).
214 Griffié, J., Boelen, L., Burn, G., Cope, A. P. & Owen, D. M. Topographic prominence as a method for cluster identification in single-molecule localisation data. *Journal of Biophotonics* **8**, 925-934, doi:https://doi.org/10.1002/jbio.201400127 (2015).

215 Okabe, A., Boots, B., Sugihara, K. & Chiu, S. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Vol. 43 (2000).
216 Levet, F. *et al.* SR-Tesseler: a method to segment and quantify localization-based super-resolution microscopy data. *Nat. Methods* **12**, 1065-1071, doi:10.1038/nmeth.3579 (2015).
217 Wang, A. *et al.* A novel deep learning-based 3D cell segmentation framework for future image-based disease detection. *Sci. Rep.* **12**, 342, doi:10.1038/s41598-021-04048-3 (2022).
218 Williamson, D. J. *et al.* Machine learning for cluster analysis of localization microscopy data. *Nature Communications* **11**, 1493, doi:10.1038/s41467-020-15293-x (2020).
219 Chazal, F. & Michel, B. An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists. *Frontiers in Artificial Intelligence* **4**, doi:10.3389/frai.2021.667963 (2021).
220 Zomorodian, A. Fast construction of the Vietoris-Rips complex. *Computers & Graphics* **34**, 263-271, doi:https://doi.org/10.1016/j.cag.2010.03.007 (2010).
221 Chazal, F., Guibas, L., Oudot, S. & Skraba, P. Persistence-Based Clustering in Riemannian Manifolds. *Journal of the ACM* **60**, doi:10.1145/1998196.1998212 (2011).
222 Meng, Z., Anand, D. V., Lu, Y., Wu, J. & Xia, K. Weighted persistent homology for biomolecular data analysis. *Sci. Rep.* **10**, 2079, doi:10.1038/s41598-019-55660-3 (2020).
223 Vipond, O. *et al.* Multiparameter persistent homology landscapes identify immune cell spatial patterns in tumors. *Proceedings of the National Academy of Sciences* **118**, doi:10.1073/pnas.2102166118 (2021).
224 Aktas, M. E., Akbas, E. & Fatmaoui, A. E. Persistence homology of networks: methods and applications. *Applied Network Science* **4**, 61, doi:10.1007/s41109-019-0179-3 (2019).
225 Byrne, N., Clough, J. R., Valverde, I., Montana, G. & King, A. P. A persistent homology-based topological loss for CNN-based multi-class segmentation of CMR. *ArXiv* **abs/2107.12689** (2021).
226 Adams, H. *et al.* Persistence images: A stable vector representation of persistent homology. **18**, 1-35 (2017).
227 Edelsbrunner, H. & Morozov, D. Persistent homology: theory and practice. *European Congress of Mathematics*, 31-50 (2014).
228 Edwards, P. *et al.* TDAExplore: quantitative image analysis through topology-based machine learning. *bioRxiv*, 2021.2006.2013.448249, doi:10.1101/2021.06.13.448249 (2021).
229 Botsch, M., Pajarola, R., Singh, G. K. C., Mémoli, F. & Carlsson, G. E. Eurographics Symposium on Point-based Graphics (2007) Topological Methods for the Analysis of High Dimensional Data Sets and 3d Object Recognition
230 Baker, A. P. *et al.* Fast transient networks in spontaneous human brain activity. *eLife* **3**, e01867, doi:10.7554/eLife.01867 (2014).
231 Nicolau, M., Levine, A. J. & Carlsson, G. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Sciences* **108**, 7265-7270, doi:10.1073/pnas.1102826108 (2011).

232 Yao, Y. *et al.* Topological Methods for Exploring Low-density States in Biomolecular Folding Pathways. *The Journal of chemical physics* **130**, 144115, doi:10.1063/1.3103496 (2009).
233 Schlather, M. On the Second-Order Characteristics of Marked Point Processes. *Bernoulli* **7**, 99-117, doi:10.2307/3318604 (2001).
234 Lagache, T. *et al.* Mapping molecular assemblies with fluorescence microscopy and object-based spatial statistics. *Nature Communications* **9**, 698, doi:10.1038/s41467-018-03053-x (2018).
235 Willems, J. & MacGillavry, H. D. A coordinate-based co-localization index to quantify and visualize spatial associations in single-molecule localization microscopy. *Sci. Rep.* **12**, 4676, doi:10.1038/s41598-022-08746-4 (2022).
236 Wang, S. *et al.* Spatially Adaptive Colocalization Analysis in Dual-Color Fluorescence Microscopy. *IEEE Transactions on Image Processing* **28**, 4471-4485, doi:10.1109/TIP.2019.2909194 (2019).
237 Dunn, K., Kamocka, M. & McDonald, J. A practical guide to evaluating colocalization in biological microscopy. *American journal of physiology. Cell physiology* **300**, 723-742, doi:10.1152/ajpcell.00462.2010 (2011).
238 Ronneberger, O. *et al.* Spatial quantitative analysis of fluorescently labeled nuclear structures: Problems, methods, pitfalls. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **16**, 523-562, doi:10.1007/s10577-008-1236-4 (2008).
239 Malkusch, S. *et al.* Coordinate-based colocalization analysis of single-molecule localization microscopy data. *Histochem. Cell Biol.* **137**, 1-10, doi:10.1007/s00418-011-0880-5 (2012).
240 Rossy, J., Cohen, E., Gaus, K. & Owen, D. Method for co-cluster analysis in multichannel single-molecule localisation data. *Histochem. Cell Biol.* **141**, 605-612, doi:10.1007/s00418-014-1208-z (2014).
241 Malkusch, S. & Heilemann, M. Extracting quantitative information from single-molecule super-resolution imaging data with LAMA – LocAlization Microscopy Analyzer. *Sci. Rep.* **6**, 34486, doi:10.1038/srep34486 (2016).
242 Owen, D. M., Williamson, D., Rentero, C. & Gaus, K. Quantitative microscopy: protein dynamics and membrane organisation. *Traffic* **10**, 962-972 (2009).
243 Peters, R. *et al.* Quantification of fibrous spatial point patterns from single-molecule localization microscopy (SMLM) data. *Bioinformatics* **33**, 1703-1711, doi:10.1093/bioinformatics/btx026 (2017).
244 Adams, J. K. *et al.* Single-frame 3D fluorescence microscopy with ultraminiature lensless FlatScope. *Science Advances* **3**, doi:10.1126/sciadv.1701548 (2017).
245 Lorbeer, R.-A. *et al.* Highly efficient 3D fluorescence microscopy with a scanning laser optical tomograph. *Opt. Express* **19**, 5419-5430, doi:10.1364/OE.19.005419 (2011).
246 Vu, Q. D. *et al.* Methods for Segmentation and Classification of Digital Microscopy Tissue Images. *Frontiers in Bioengineering and Biotechnology* **7** (2019).
247 Long, F. Microscopy cell nuclei segmentation with enhanced U-Net. *BMC Bioinformatics* **21**, 8, doi:10.1186/s12859-019-3332-1 (2020).
248 Masubuchi, S. *et al.* Deep-learning-based image segmentation integrated with optical microscopy for automatically searching for two-dimensional materials. *npj 2D Materials and Applications* **4**, 3, doi:10.1038/s41699-020-0137-z (2020).
249 Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. *Cell Detection with Star-Convex Polygons*. Vol. 11071 (Springer International Publishing, 2018).

250 Lee, M. Y. *et al.* CellSeg: a robust, pre-trained nucleus segmentation and pixel quantification software for highly multiplexed fluorescence images. *BMC Bioinformatics* **23**, 46, doi:10.1186/s12859-022-04570-9 (2022).
251 Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100-106, doi:10.1038/s41592-020-01018-x (2021).
252 Kirillov, A. *et al.* Segment Anything. *ArXiv* **abs/2304.02643** (2023).
253 Belthangady, C. & Royer, L. A. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nat. Methods* **16**, 1215-1225, doi:10.1038/s41592-019-0458-z (2019).
254 Biderman, S. & Scheirer, W. J. in *Proceedings on "I Can't Believe It's Not Better!" at NeurIPS Workshops* Vol. 137 (eds Forde Jessica Zosa, Ruiz Francisco, F. Pradier Melanie, & Schein Aaron) 106--117 (PMLR, Proceedings of Machine Learning Research, 2020).
255 Chen, J., Yang, L., Zhang, Y., Alber, M. & Chen, D. Combining Fully Convolutional and Recurrent Neural Networks for 3D Biomedical Image Segmentation. *ArXiv* **abs/1609.01006** (2016).
256 Jiang, J., Kao, P. Y., Belteton, S. A., Szymanski, D. B. & Manjunath, B. S. in *2019 IEEE International Conference on Image Processing (ICIP).* 1555-1559.
257 Wirjadi, O. Survey of 3D image segmentation methods. *ITWM Rep.* **123** (2007).
258 Taha, A. A. & Hanbury, A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med. Imaging* **15**, 29, doi:10.1186/s12880-015-0068-x (2015).
259 Kornilov, A. S. & Safonov, I. V. An Overview of Watershed Algorithm Implementations in Open Source Libraries. *Journal of Imaging* **4**, 123 (2018).
260 Sun, H., Yang, J. & Ren, M. A fast watershed algorithm based on chain code and its application in image segmentation. *Pattern Recog. Lett.* **26**, 1266-1274, doi:https://doi.org/10.1016/j.patrec.2004.11.007 (2005).
261 Raman, S., Maxwell Ca Fau - Barcellos-Hoff, M. H., Barcellos-Hoff Mh Fau - Parvin, B. & Parvin, B. Geometric approach to segmentation and protein localization in cell culture assays. *J. Microsc.* **225**, 22-30 (2007).
262 Zinchuk, V. & Grossenbacher, O. Machine Learning for Analysis of Microscopy Images: A Practical Guide. *Curr. Protoc. Cell Biol.* **86**, doi:10.1002/cpcb.101 (2020).
263 Chen, P. C. & Pavlidis, T. Image segmentation as an estimation problem. *Computer Graphics and Image Processing* **12**, 153-172, doi:https://doi.org/10.1016/0146-664X(80)90009-X (1980).
264 Ollion, J., Cochennec, J., Loll, F., Escudé, C. & Boudier, T. TANGO: a generic tool for high-throughput 3D image analysis for studying nuclear organization. *Bioinformatics* **29**, 1840-1841, doi:10.1093/bioinformatics/btt276 (2013).
265 Lin, Z., Jin, J. & Talbot, H. Unseeded region growing for 3D image segmentation. *ACM International Conference Proceeding Series* (2002).
266 Ershov, D. *et al.* TrackMate 7: integrating state-of-the-art segmentation algorithms into tracking pipelines. *Nat. Methods* **19**, 829-832, doi:10.1038/s41592-022-01507-1 (2022).
267 Otter, N., Porter, M. A., Tillmann, U., Grindrod, P. & Harrington, H. A. A roadmap for the computation of persistent homology. *EPJ Data Science* **6**, 17, doi:10.1140/epjds/s13688-017-0109-5 (2017).
268 Kong, J. *et al.* in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI).* 1212-1215.

269 Li, G. *et al.* 3D cell nuclei segmentation based on gradient flow tracking. *BMC Cell Biol.* **8**, 40, doi:10.1186/1471-2121-8-40 (2007).
270 Maška, M. *et al.* The Cell Tracking Challenge: 10 years of objective benchmarking. *Nat. Methods* **20**, 1010-1020, doi:10.1038/s41592-023-01879-y (2023).
271 Moreno, S., Klar, A. & Nurse, P. in *Methods Enzymol.* Vol. 194 795-823 (Academic Press, 1991).
272 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676-682, doi:10.1038/nmeth.2019 (2012).
273 Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nat. Methods* **19**, 1634-1641, doi:10.1038/s41592-022-01663-4 (2022).
274 Pau, G., Fuchs, F., Sklyar, O., Boutros, M. & Huber, W. EBImage—an R package for image processing with applications to cellular phenotypes. *Bioinformatics* **26**, 979-981, doi:10.1093/bioinformatics/btq046 (2010).
275 Laine, R. F., Jacquemet, G. & Krull, A. Imaging in focus: An introduction to denoising bioimages in the era of deep learning. *The International Journal of Biochemistry & Cell Biology* **140**, 106077, doi:https://doi.org/10.1016/j.biocel.2021.106077 (2021).
276 Tinevez, J.-Y. *et al.* TrackMate: An open and extensible platform for single-particle tracking. *Methods* **115**, 80-90, doi:https://doi.org/10.1016/j.ymeth.2016.09.016 (2017).
277 Panconi, L., Tansell, A., Collins, A. J., Makarova, M. & Owen, D. M. Three-dimensional topology-based analysis segments volumetric and spatiotemporal fluorescence microscopy. *Biological Imaging* **4**, e1, doi:10.1017/S2633903X23000260 (2024).
278 K V, L., R, A., Michahial, S. & M, D. Implementation of Watershed Segmentation. *IJARCCE* **5**, 196-199, doi:10.17148/IJARCCE.2016.51243 (2016).
279 Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics* **9**, 62-66, doi:10.1109/TSMC.1979.4310076 (1979).
280 Ahirwar, D. A. Study of Techniques used for Medical Image Segmentation and Computation of Statistical Test for Region Classification of Brain MRI. *I.J. Information Technology and Computer Science, 2013, 05, 44-53* **5**, doi:10.5815/ijitcs.2013.05.06 (2013).
281 Panconi, L., Makarova, M., Lambert, E. R., May, R. C. & Owen, D. M. Topology-based fluorescence image analysis for automated cell identification and segmentation. *bioRxiv*, 2022.2006.2022.497179, doi:10.1101/2022.06.22.497179 (2022).
282 Rosenbluth, M., Lam, W. & Fletcher, D. Force Microscopy of Nonadherent Cells: A Comparison of Leukemia Cell Deformability. *Biophys. J.* **90**, 2994-3003, doi:10.1529/biophysj.105.067496 (2006).
283 Yang *et al.* Determination of the Membrane Transport Properties of Jurkat Cells with a Microfluidic Device. *Micromachines* **10**, 832, doi:10.3390/mi10120832 (2019).
284 Litvinova, L. *et al.* Content of Nucleic Acids in the Adhesive T-Lymphoblast Jurkat Line and Their Mobility In Vitro. *Cell and Tissue Biol.* **15**, 445-454, doi:10.1134/S1990519X21050072 (2021).

285 Zhu, Z. *et al.* Time-lapse electrical impedance spectroscopy for monitoring the cell cycle of single immobilized S. pombe cells. *Sci. Rep.* **5**, 17180, doi:10.1038/srep17180 (2015).
286 Bártová, E. *et al.* Recruitment of Oct4 Protein to UV-Damaged Chromatin in Embryonic Stem Cells. *PLoS One* **6**, doi:10.1371/journal.pone.0027281 (2011).
287 Wickham, H. *Chapman & Hall/CRC the R series (CRC Press)* (CRC Press, Boca Raton, FL, 2015).
288 Awan, M. & Shin, J. Weakly supervised multi-class semantic video segmentation for road scenes. *Comput. Vision Image Understanding* **230**, 103664, doi:https://doi.org/10.1016/j.cviu.2023.103664 (2023).
289 Roob, E., Trendel, N., Rein ten Wolde, P. & Mugler, A. Cooperative Clustering Digitizes Biochemical Signaling and Enhances its Fidelity. *Biophys. J.* **110**, 1661-1669, doi:https://doi.org/10.1016/j.bpj.2016.02.031 (2016).
290 Sieber, J. J. *et al.* Anatomy and Dynamics of a Supramolecular Membrane Protein Cluster. *Science* **317**, 1072-1076, doi:10.1126/science.1141727 (2007).
291 Lang, T. & Rizzoli, S. O. Membrane protein clusters at nanoscale resolution: more than pretty pictures. *Physiology* **25**, 116-124 (2010).
292 Hartman, N. C. & Groves, J. T. Signaling clusters in the cell membrane. *Curr. Opin. Cell Biol.* **23**, 370-376, doi:https://doi.org/10.1016/j.ceb.2011.05.003 (2011).
293 Simoncelli, S. *et al.* Multi-color Molecular Visualization of Signaling Proteins Reveals How C-Terminal Src Kinase Nanoclusters Regulate T Cell Receptor Activation. *Cell Rep.* **33**, 108523, doi:https://doi.org/10.1016/j.celrep.2020.108523 (2020).
294 Davis, S. J. & van der Merwe, P. A. The kinetic-segregation model: TCR triggering and beyond. *Nat. Immunol.* **7**, 803-809, doi:10.1038/ni1369 (2006).
295 Dong, D. *et al.* Structural basis of assembly of the human T cell receptor–CD3 complex. *Nature* **573**, 546-552, doi:10.1038/s41586-019-1537-0 (2019).
296 Chen, K. Y. *et al.* Trapping or slowing the diffusion of T cell receptors at close contacts initiates T cell signaling. *Proceedings of the National Academy of Sciences* **118**, e2024250118, doi:10.1073/pnas.2024250118 (2021).
297 Nieves, D. J. *et al.* The T cell receptor displays lateral signal propagation involving non-engaged receptors. *Nanoscale* **14**, 3513-3526, doi:10.1039/D1NR05855J (2022).
298 An, G., Mi, Q., Dutta-Moscato, J. & Vodovotz, Y. Agent-based models in translational systems biology. *WIREs Systems Biology and Medicine* **1**, 159-171, doi:https://doi.org/10.1002/wsbm.45 (2009).
299 Venkataramani, V., Herrmannsdörfer, F., Heilemann, M. & Kuner, T. SuReSim: simulating localization microscopy experiments from ground truth models. *Nat. Methods* **13**, 319-321, doi:10.1038/nmeth.3775 (2016).
300 Lagardère, M., Chamma, I., Bouilhol, E., Nikolski, M. & Thoumine, O. FluoSim: simulator of single molecule dynamics for fluorescence live-cell and super-resolution imaging of membrane proteins. *Sci. Rep.* **10**, 19954, doi:10.1038/s41598-020-75814-y (2020).
301 Lindén, M., Ćurić, V., Boucharin, A., Fange, D. & Elf, J. Simulated single molecule microscopy with SMeagol. *Bioinformatics* **32**, 2394-2395, doi:10.1093/bioinformatics/btw109 (2016).

302 Novák, T., Gajdos, T., Sinkó, J., Szabó, G. & Erdélyi, M. TestSTORM: Versatile simulator software for multimodal super-resolution localization fluorescence microscopy. *Sci. Rep.* **7**, 951, doi:10.1038/s41598-017-01122-7 (2017).
303 Bourgeois, D. Single molecule imaging simulations with advanced fluorophore photophysics. *Communications Biology* **6**, 53, doi:10.1038/s42003-023-04432-x (2023).
304 Pertsinidis, A. *et al.* Ultrahigh-resolution imaging reveals formation of neuronal SNARE/Munc18 complexes in situ. *Proceedings of the National Academy of Sciences* **110**, 2812-2820, doi:10.1073/pnas.1310654110 (2013).
305 Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-Liquid Phase Separation in Biology. *Annual Review of Cell and Developmental Biology* **30**, 39-58, doi:10.1146/annurev-cellbio-100913-013325 (2014).
306 Owen, D. M., Sauer, M. & Gaus, K. Fluorescence localization microscopy. *Commun. Integr. Biol.* **5**, 345-349, doi:10.4161/cib.20348 (2012).
307 Griffié, J., Peters, R. & Owen, D. M. An agent-based model of molecular aggregation at the cell membrane. *PLoS One* **15**, doi:10.1371/journal.pone.0226825 (2020).
308 Schnitzbauer, J., Strauss, M. T., Schlichthaerle, T., Schueder, F. & Jungmann, R. Super-resolution microscopy with DNA-PAINT. *Nat. Protoc.* **12**, 1198-1228, doi:10.1038/nprot.2017.024 (2017).
309 Young, M. E., Carroad, P. A. & Bell, R. L. Estimation of diffusion coefficients of proteins. *Biotechnology and Bioengineering* **22**, 947-955, doi:https://doi.org/10.1002/bit.260220504 (1980).
310 Ernst, D. & Köhler, J. Measuring a diffusion coefficient by single-particle tracking: statistical analysis of experimental mean squared displacement curves. *Physical Chemistry Chemical Physics* **15**, 845-849, doi:10.1039/C2CP43433D (2013).
311 Das, A. A., Ajayakumar Darsana, T. & Jacob, E. Agent-based re-engineering of ErbB signaling: a modeling pipeline for integrative systems biology. *Bioinformatics* **33**, 726-732, doi:10.1093/bioinformatics/btw709 (2017).
312 An, G. Introduction of an agent-based multi-scale modular architecture for dynamic knowledge representation of acute inflammation. *Theoretical Biology and Medical Modelling* **5**, 11, doi:10.1186/1742-4682-5-11 (2008).
313 Bankes, S. C. Agent-based modeling: A revolution? *Proceedings of the National Academy of Sciences* **99**, 7199-7200, doi:10.1073/pnas.072081299 (2002).
314 Rysman, E. *et al.* De novo Lipogenesis Protects Cancer Cells from Free Radicals and Chemotherapeutics by Promoting Membrane Lipid Saturation. *Cancer Res.* **70**, 8117-8126, doi:10.1158/0008-5472.CAN-09-3871 (2010).
315 Adriá-Cebrián, J. *et al.* MCF-7 Drug Resistant Cell Lines Switch Their Lipid Metabolism to Triple Negative Breast Cancer Signature. *Cancers (Basel)* **13**, 5871 (2021).
316 Chen, X. *et al.* Plasma lipidomics profiling identified lipid biomarkers in distinguishing early-stage breast cancer from benign lesions. *Oncotarget* **7**, 36622-36631, doi:10.18632/oncotarget.9124 (2016).
317 Sezgin, E., Levental, I., Mayor, S. & Eggeling, C. The mystery of membrane organization: composition, regulation and roles of lipid rafts. *Nat Rev Mol Cell Biol* **18**, 361-374, doi:10.1038/nrm.2017.16 (2017).

318 Levental, I., Levental, K. R. & Heberle, F. A. Lipid Rafts: Controversies Resolved, Mysteries Remain. *Trends Cell Biol* **30**, 341-353, doi:10.1016/j.tcb.2020.01.009 (2020).
319 Sharonov, A. & Hochstrasser, R. M. Wide-field subdiffraction imaging by accumulated binding of diffusing probes. *Proc Natl Acad Sci U S A* **103**, 18911-18916, doi:10.1073/pnas.0609643104 (2006).
320 Danylchuk, D. I., Moon, S., Xu, K. & Klymchenko, A. S. Switchable Solvatochromic Probes for Live-Cell Super-resolution Imaging of Plasma Membrane Organization. *Angew Chem Int Ed Engl* **58**, 14920-14924, doi:10.1002/anie.201907690 (2019).
321 Barbotin, A. *et al.* z-STED Imaging and Spectroscopy to Investigate Nanoscale Membrane Structure and Dynamics. *Biophys J* **118**, 2448-2457, doi:10.1016/j.bpj.2020.04.006 (2020).
322 Rudland, P. S., Twiston Davies, A. C. & Tsao, S. W. Rat mammary preadipocytes in culture produce a trophic agent for mammary epithelia-prostaglandin E2. *J Cell Physiol* **120**, 364-376, doi:10.1002/jcp.1041200315 (1984).
323 Morales-Penningston, N. F. *et al.* GUV preparation and imaging: Minimizing artifacts. *Biochim. Biophys. Acta* **1798**, 1324-1332, doi:https://doi.org/10.1016/j.bbamem.2010.03.011 (2010).
324 Griffie, J., Peters, R. & Owen, D. M. An agent-based model of molecular aggregation at the cell membrane. *PLoS One* **15**, e0226825, doi:10.1371/journal.pone.0226825 (2020).
325 Stein, M. Large Sample Properties of Simulations Using Latin Hypercube Sampling. *Technometrics* **29**, 143-151, doi:Doi 10.2307/1269769 (1987).
326 Pesarin, F. & Salmaso, L. The permutation testing approach: a review. *STATISTICA* **70**, 481-509, doi:10.6092/issn.1973-2201/3599 (1993).
327 Pebesma, E. & Bivand, R. *Spatial Data Science: With Applications in R*. (2023).
328 Ripley, B. D. The Second-Order Analysis of Stationary Point Processes. *J. Appl. Probab.* **13**, 255-266, doi:10.2307/3212829 (1976).
329 Murtagh, F. & Farid, M. Pattern Classification, by Richard O. Duda, Peter E. Hart, and David G. Stork. *J. Classification* **18**, 273-275 (2001).
330 Scrucca, L., Fop, M., Murphy, T. & Raftery, A. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *The R Journal* **8**, 205-233, doi:10.32614/RJ-2016-021 (2016).
331 Massey, F. J. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association* **46**, 68-78, doi:10.2307/2280095 (1951).
332 Margalit, A. & Knott, G. D. An algorithm for computing the union, intersection or difference of two polygons. *Comput. graph* **13**, 167-183 (1989).
333 Jeon, J. H. *et al.* Lipid raft modulation inhibits NSCLC cell migration through delocalization of the focal adhesion complex. *Lung Cancer* **69**, 165-171, doi:https://doi.org/10.1016/j.lungcan.2009.10.014 (2010).
334 Goujon, A. *et al.* Mechanosensitive Fluorescent Probes to Image Membrane Tension in Mitochondria, Endoplasmic Reticulum, and Lysosomes. *J Am Chem Soc* **141**, 3380-3384, doi:10.1021/jacs.8b13189 (2019).
335 Mollinedo, F. & Gajate, C. Lipid rafts as major platforms for signaling regulation in cancer. *Advances in biological regulation* **57**, 130-146, doi:10.1016/j.jbior.2014.10.003 (2014).
336 Weigert, M., Schmidt, U., Haase, R., Sugawara, K. & Myers, G. in *2020 IEEE Winter Conference on Applications of Computer Vision (WACV).* 3655-3662.

337 Shrestha, P., Kuang, N. & Yu, J. Efficient end-to-end learning for cell segmentation with machine generated weak annotations. *Communications Biology* **6**, 232, doi:10.1038/s42003-023-04608-5 (2023).
338 Stegmaier, J. *et al.* Fast Segmentation of Stained Nuclei in Terabyte-Scale, Time Resolved 3D Microscopy Image Stacks. *PLoS One* **9**, doi:10.1371/journal.pone.0090036 (2014).
339 Giedt, R. J., Koch, P. D. & Weissleder, R. Single Cell Analysis of Drug Distribution by Intravital Imaging. *PLoS One* **8**, doi:10.1371/journal.pone.0060988 (2013).
340 Vicar, T. *et al.* Cell segmentation methods for label-free contrast microscopy: review and comprehensive comparison. *BMC Bioinformatics* **20**, 360, doi:10.1186/s12859-019-2880-8 (2019).
341 Chalfoun, J. *et al.* FogBank: a single cell segmentation across multiple cell lines and image modalities. *BMC Bioinformatics* **15**, 431, doi:10.1186/s12859-014-0431-x (2014).
342 Rojas-Moraleda, R. *et al.* Robust detection and segmentation of cell nuclei in biomedical images based on a computational topology framework. *Med. Image Anal.* **38**, 90-103, doi:https://doi.org/10.1016/j.media.2017.02.009 (2017).
343 Assaf, R., Goupil, A., Kacim, M. & Vrabie, V. *Topological persistence based on pixels for object segmentation in biomedical images*. (IEEE, 2017).
344 Griffié, J. *et al.* Virtual-SMLM, a virtual environment for real-time interactive SMLM acquisition. *bioRxiv*, 2020.2003.2005.967893, doi:10.1101/2020.03.05.967893 (2020).
345 Tantiwong, C. *et al.* An agent-based approach for modelling and simulation of glycoprotein VI receptor diffusion, localisation and dimerisation in platelet lipid rafts. *Sci. Rep.* **13**, 3906, doi:10.1038/s41598-023-30884-6 (2023).
346 Bolmatov, D. *et al.* Molecular Picture of the Transient Nature of Lipid Rafts. *Langmuir* **36**, 4887-4896, doi:10.1021/acs.langmuir.0c00125 (2020).
347 Owen, D. & Gaus, K. Imaging lipid domains in cell membranes: the advent of super-resolution fluorescence microscopy. *Frontiers in Plant Science* **4**, 503 (2013).
348 Schütte, O. M. *et al.* Size and mobility of lipid domains tuned by geometrical constraints. *Proceedings of the National Academy of Sciences* **114**, 6064-6071, doi:10.1073/pnas.1704199114 (2017).
349 Pageon, S. V., Nicovich, P. R., Mollazade, M., Tabarin, T. & Gaus, K. Clus-DoC: a combined cluster detection and colocalization analysis for single-molecule localization microscopy data. *Mol. Biol. Cell* **27**, 3627-3636, doi:10.1091/mbc.E16-07-0478 (2016).
350 Vasudevan, K., Eckel, S., Fleischer, F., Schmidt, V. & Cook, F. A. Statistical analysis of spatial point patterns on deep seismic reflection data: a preliminary test. *Geophysical Journal International* **171**, 823-840, doi:10.1111/j.1365-246X.2007.03572.x (2007).
351 Xin, H. *et al.* Spatial pattern analysis of forest trees based on the vectorial mark. *Journal of Forestry Research* **33**, 1301-1315, doi:10.1007/s11676-021-01417-6 (2022).
352 Martinelli, A. L. & Rapsomaniki, M. A. ATHENA: analysis of tumor heterogeneity from spatial omics measurements. *Bioinformatics* **38**, 3151-3153, doi:10.1093/bioinformatics/btac303 (2022).
353 Unterauer, E. M. & Jungmann, R. Quantitative Imaging With DNA-PAINT for Applications in Synaptic Neuroscience. *Front. Synaptic Neurosci.* **13** (2022).

354 Andrews, B. *et al.* Imaging cell biology. *Nat. Cell Biol.* **24**, 1180-1185, doi:10.1038/s41556-022-00960-6 (2022).
355 Alvelid, J. & Testa, I. Fluorescence microscopy at the molecular scale. *Current Opinion in Biomedical Engineering* **12**, 34-42, doi:https://doi.org/10.1016/j.cobme.2019.09.009 (2019).
356 Owen, D. M., Rentero, C., Magenau, A., Abu-Siniyeh, A. & Gaus, K. Quantitative imaging of membrane lipid order in cells and organisms. *Nat. Protoc.* **7**, 24-35, doi:10.1038/nprot.2011.419 (2012).
357 Baddeley, A. & Turner, R. spatstat: An R Package for Analyzing Spatial Point Patterns. *Journal of Statistical Software* **12**, 1 - 42, doi:10.18637/jss.v012.i06 (2005).
358 Xu, S. L., Cai, Y. W. & Cheng, G. D. Volume preserving nonlinear density filter based on heaviside functions. *Struct Multidiscip O* **41**, 495-505, doi:10.1007/s00158-009-0452-7 (2010).

Appendix

7.1 Formulas

Calculation of Generalised Polarisation in Fluorescence Microscopy

The GP image is calculated under the following equation:

$$GP = \frac{I_1 - G \times I_2}{I_1 + G \times I_2}.$$

Here I_j is the intensity profile in the image acquired in spectral channel j, with the first channel corresponding to the ordered phase, and G is the calibration factor or G factor. This is equivalent to,

$$GP = \frac{I_{400-460} - G \times I_{470-530}}{I_{400-460} + G \times I_{470-530}} \text{ and } GP = \frac{I_{500-580} - G \times I_{620-750}}{I_{500-580} + G \times I_{620-750}},$$

for Laurdan and di-4-ANEPPDHQ respectively[356]. The G factor is used in the GP value calculation to compensate for differences in the efficiency of photon collection in the two channels. This calibration is performed by imaging dyes in a standard solution under standard conditions such that the ratio of fluorescence in each channel is always constant. The G factor is then calculated as,

$$G = \frac{GP_{ref} + GP_{ref}GP_{mes} - GP_{mes} - 1}{GP_{mes} + GP_{ref}GP_{mes} - GP_{ref} - 1},$$

where GP_{mes} is the GP value of the respective dye in pure dimethyl sulfoxide (DMSO – a chemical which dissolves polar and non-polar compounds), measured with the same microscope setup and settings as those used for the real sample, and GP_{ref} is a reference value for the dye in DMSO[356]. The GP_{ref} may have a conventional value already assigned. For example, under Laurdan, GP_{ref} = 0.207[356]. Generally, this value

must be chosen so that the GP values for model membranes with liquid-ordered and -disordered phases are separated at around GP = 0.

Calculation of sensitivity and specificity

Consider the matrix representation of a binary image, I. For each pixel (i,j), we denote the pixel intensity in the matrix representation as I_{ij}. We denote pixel (i,j) as active if $I_{ij} = 1$ and inactive if $I_{ij} = 0$. Consider a ground-truth image matrix G and a binarised segmentation of G, denoted S. Then we define pixel (i,j) as a true positive if $G_{ij} = 1$ and $S_{ij} = 1$, true negative if $G_{ij} = 0$ and $S_{ij} = 0$, false positive if $G_{ij} = 0$ and $S_{ij} = 1$, and false negative if $G_{ij} = 1$ and $S_{ij} = 0$. Let TP, TN, FP and FN denote the number of true positives, true negatives, false positives and false negatives, respectively. Then we define sensitivity as,

$$Sens = \frac{TP}{TP + FN},$$

or the fraction of true positive pixels over all active pixels, and define specificity as,

$$Spec = \frac{TN}{TN + FP},$$

or the fraction of true negative pixels over all inactive pixels[281].

Derivation of error and step size for ASMODEUS

The average ratiometric error of point i is given by,

$$E_i = \frac{1}{|R|}\sum_{r \in R} e_i(r),$$

where R is the set of radial values to iterate over (derived during calculation of the Ripley's K function[357]) and $e_i(r)$ is the scalar error at radius r, given by,

$$e_i(r) = \begin{cases} \dfrac{|K_t(r) - K_i(r)|}{K_t(r) + K_i(r)}, & K_i(r) > K_t(r) \\ 1 - \dfrac{K_i(r) - \pi r^2}{K_t(r) - \pi r^2} & K_t(r) \geq K_i(r) > \pi r^2 \\ 1, & K_i(r) \leq \pi r^2 \end{cases}.$$

Here, $K_t(r)$ is the global target K function. The lower bound is set at πr^2 as this is the expected distribution of the K function for a CSR point pattern[328]. The step size of point i is then calculated as,

$$D_i = (D_{max} - D_{min})E_i^2 + D_{min}.$$

A quadratically decreasing step size was chosen based on preliminary results[307].

Weighted proportion for P-Check

Let $N \in \mathbb{N}$ be a finite number (here, the total number of random trials used in the permutation test of P-Check). Then, for each $k = 0,1, \ldots, N$, let $M_k \subset \mathbb{R}^3$ be a marked point pattern such that $|M_k| = n$ (M_k contains n points) and let $X_k = \{(x, y): (x, y, m) \in M_k\}$ be the spatial point pattern over which M_k is defined. Then we define the weighted proportion of M_k as,

$$P_k = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\mathcal{H}(|\boldsymbol{x}_i - \boldsymbol{x}_j| - r)(1 - \mathcal{H}(|m_i - m_j|))}{\sum_{i=1}^{n}\sum_{j=1}^{n}\mathcal{H}\left(|\boldsymbol{x}_i - \boldsymbol{x}_j| - r\right)},$$

where $\boldsymbol{x}_i \in X_k$ is the spatial coordinate of point i, m_i is the associated mark of point i, r is the search radius, and $\mathcal{H}$ is the Heaviside function, which equates to 1 when its input is positive and 0 otherwise[358].

Intersection over union

Let $M \subset \mathbb{R}^3$ be a marked point pattern and let $X = \{(x, y): (x, y, m) \in M\}$ be the spatial point pattern over which M is defined. Suppose $g_1, g_2, \ldots, g_k \subseteq X$ and

$c_1, c_2, \dots, c_n \subseteq X$ (for finite k, n), with $g_i \cap g_j = \emptyset \; \forall \; i \neq j$ and $c_i \cap c_j = \emptyset \; \forall \; i \neq j$. Here, we may interpret $g_1, g_2, \dots, g_k$ to be the ground truth clusters in X and $c_1, c_2, \dots, c_n$ to be the clusters recovered from M (e.g. by JOSEPH). For any subset $y = \{p_1, p_2, \dots, p_N\} \subseteq X$, we define the convex hull[336] of y as the set,

$$Conv(y) = \left\{ \sum_{j=1}^{N} \lambda_j p_j : \lambda_j \geq 0 \; \forall \; j \text{ and } \sum_{j=1}^{N} \lambda_j = 1 \right\}.$$

Let G and C be the union of all convex hulls of $g_1, g_2, \dots, g_k$ and $c_1, c_2, \dots, c_n$, respectively. That is,

$$G = \bigcup_{i=1}^{k} Conv(g_i),$$

$$C = \bigcup_{i=1}^{n} Conv(c_i).$$

Then we define the intersection over union, I, of subsets $g_1, g_2, \dots, g_k$ and $c_1, c_2, \dots, c_n$ as,

$$I = \frac{G \cap C}{G \cup C}.$$

Formula for overlap of two Normal distributions

For two Normal distributions defined by $f_1 = N(\mu_1, \sigma_1)$ and $f_2 = N(\mu_2, \sigma_2)$ the overlap between f_1 and f_2 is quantified as,

$$o = \frac{\sqrt{\sigma_1 + \sigma_2}}{|\mu_1 - \mu_2|}.$$

7.2 Schematics

Figure 7.1: Conceptual diagram of 2D and 3D TOBLERONE algorithm.

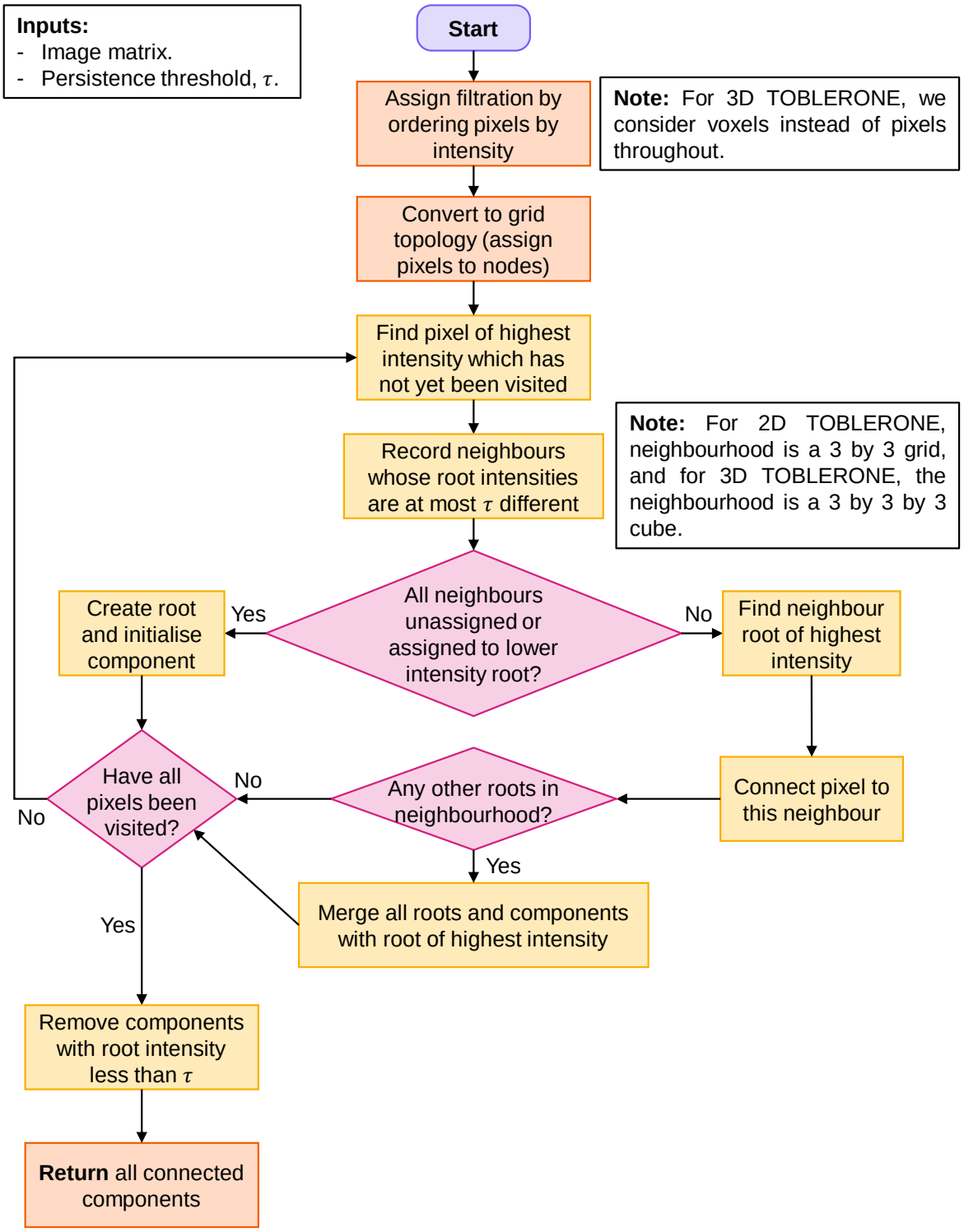


Figure 7.2: Conceptual diagram of tempTOBLERONE algorithm.

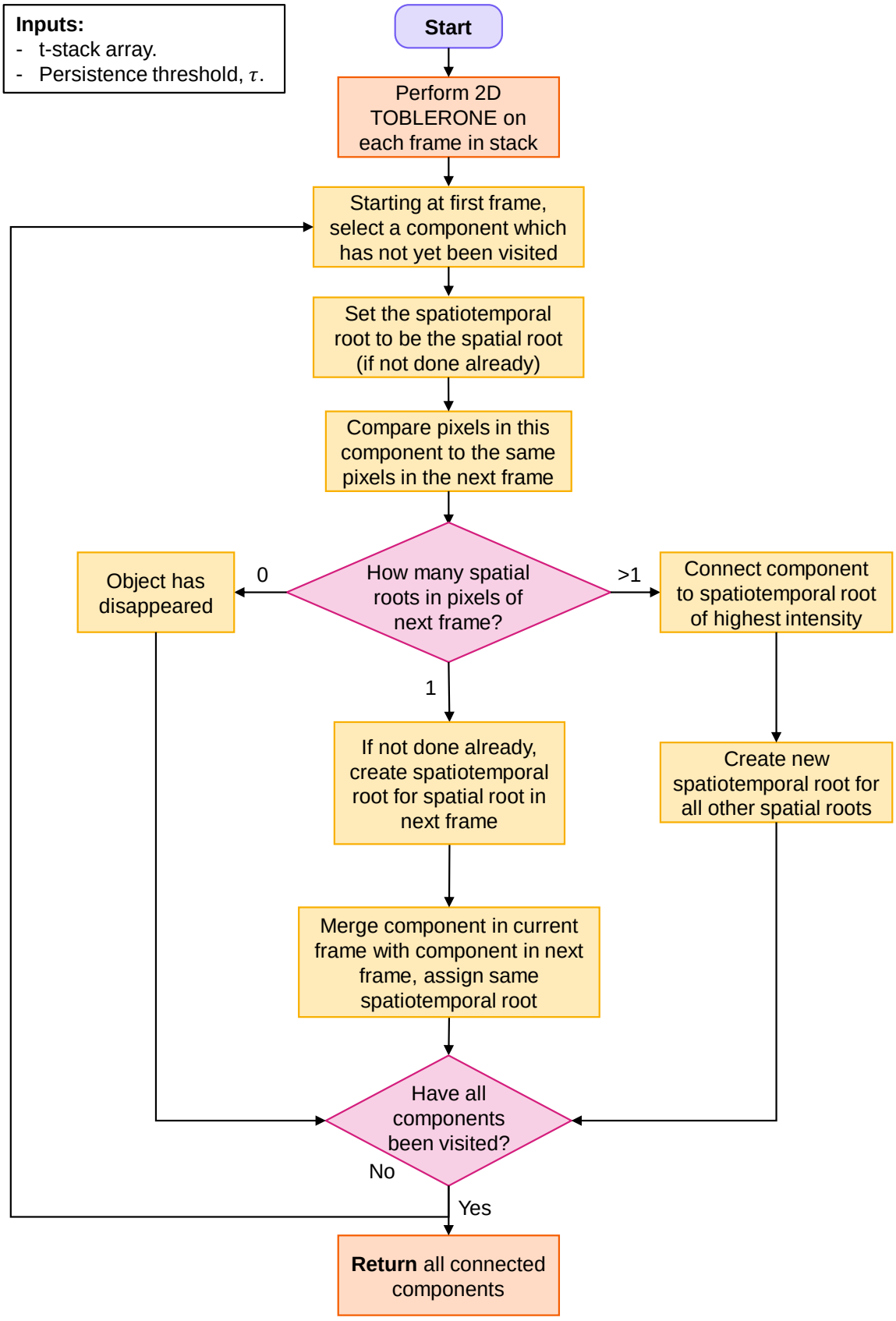


Figure 7.3: Conceptual diagram of ASMODEUS algorithm.

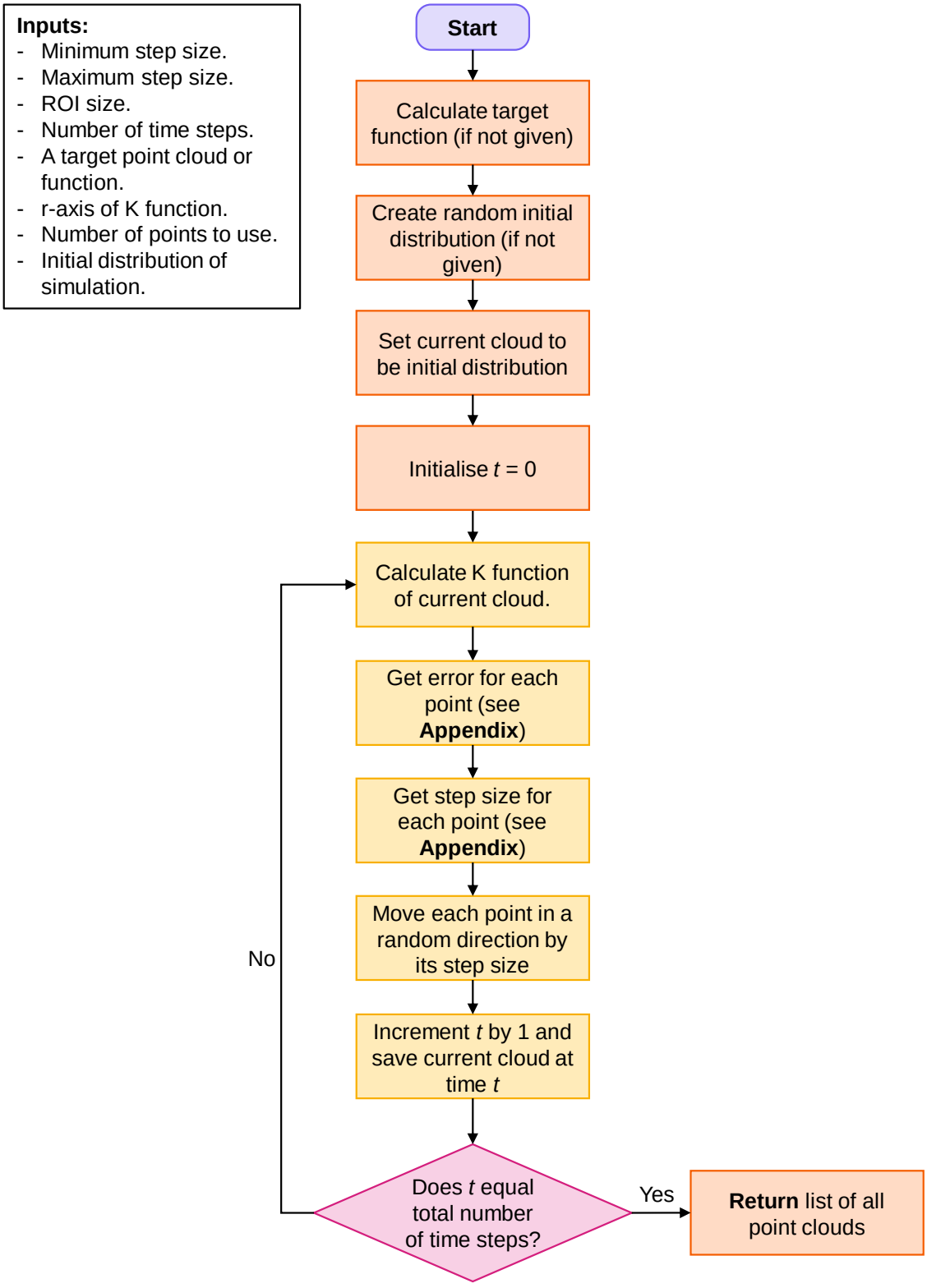


Figure 7.4: Conceptual diagram of P-Check algorithm.

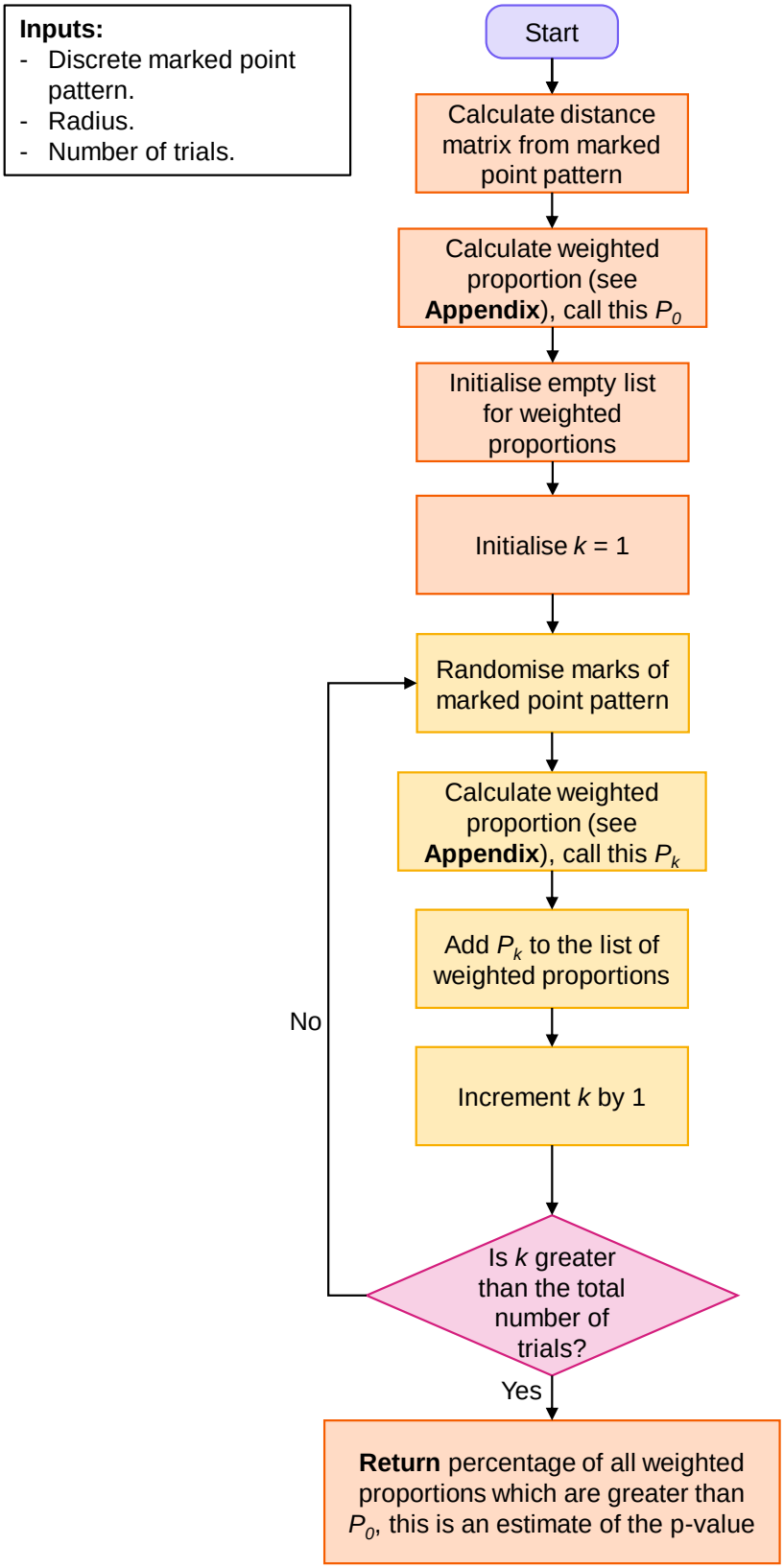
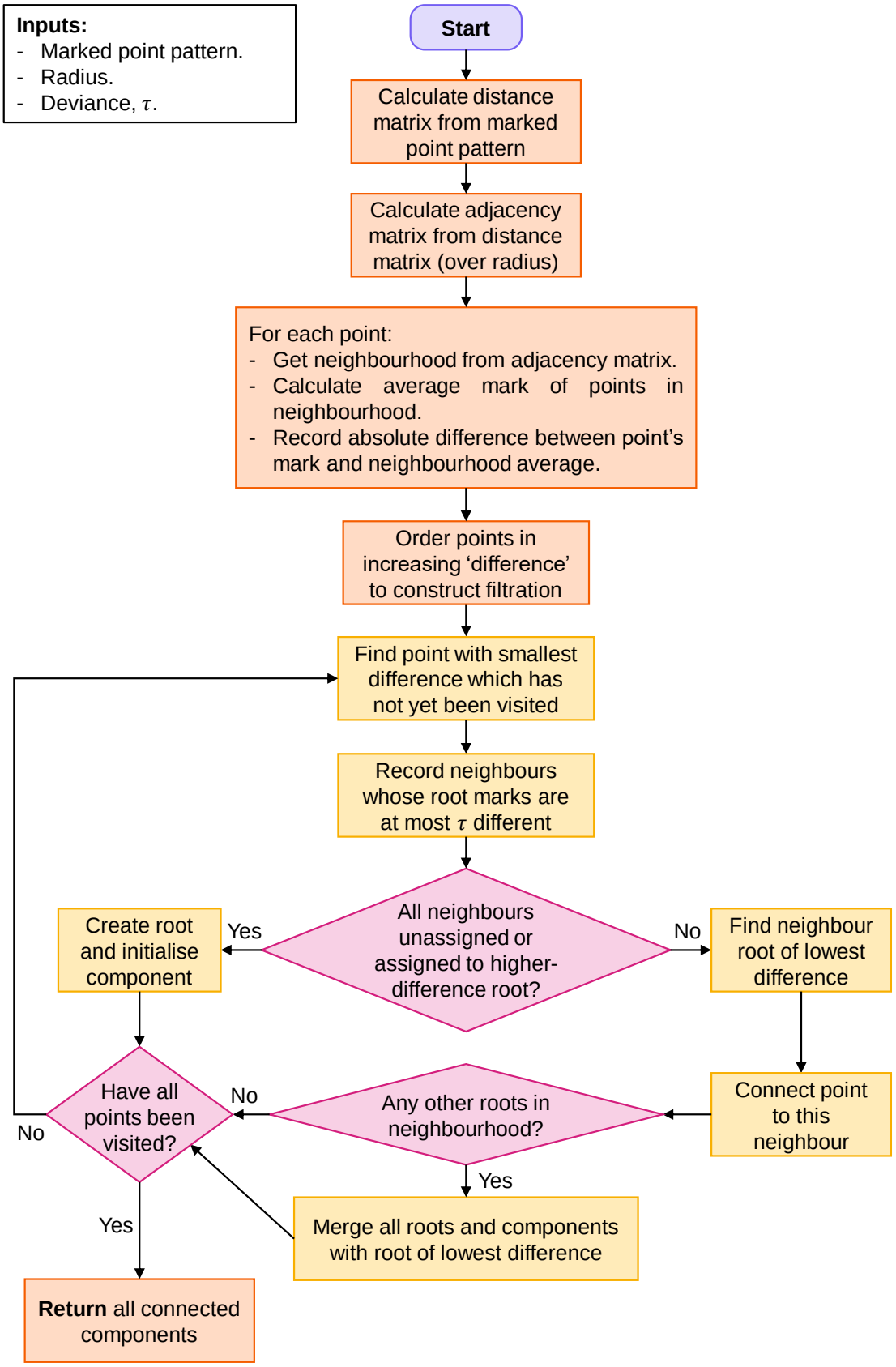


Figure 7.5: Conceptual diagram of JOSEPH algorithm.



7.3 Pseudocode

7.2.1 2DTOBLERONE

```
#Initialise:
#image is a matrix, for which each entry contains the intensity
of a pixel.
#p is the persistence threshold.
INPUT image, p
CALL rescale on image, setting minimum intensity to 0 and maximum
intensity to 1
#Calculate filtration - order pixels by intensity.
CALL as.vector on image to get imageVector
CALL order on imageVector to get filtration
CALL as.matrix on filtration to get filtrationMatrix
#Initialise union-find data structure:
#Segmentation stores several lists of pixels, with each list
corresponding to one connected component.
#r stores the root pixel of all pixels in a connected component.
#e stores the index in Segmentation of all pixels in a connected
component.
#m stores the total number of connected components.
INITIALISE empty list Segmentation
INITIALISE zero vectors r and e with length equal to number of
pixels in image
INITIALISE m to 0
#Perform persistent homology.
FOR each pixel index i in filtration
      GET pixel coordinates of i from image
      GET coordinates of all pixels in 3 by 3 neighbourhood around
      pixel i
      SET filtrationValues as values of all neighbouring pixels
      from filtrationMatrix
      SET N to be the set of neighbours of i with entries in
      filtrationValues lower than i
      IF N is empty
            #This pixel is a root.
            INCREMENT m by 1
            INITIALISE new list at index m in Segmentation
            containing pixel i
            SET value of e at position i to be m
            SET value of r at position i to be i
      ELSE
      #This pixel may be a root.
      SET R to be the set of roots for all pixels in N (i.e. R =
      r[N])
      SET potentials to be the set of all roots in R whose intensity
      in image is at most p different from the intensity of i in
      image
            IF potentials is empty
```

```
                #This pixel cannot connect to any existing
                component in Segmentation.
                IF intensity of pixel i in image is greater than
                p
                      #This pixel is a root.
                      INCREMENT m by 1
                      INITIALISE new list at index m in
                      Segmentation containing pixel i
                      SET value of e at position i to be m
                      SET value of r at position i to be i
          ELSE
                #This pixel connects to the root of highest
                intensity.
                SET j to be the argmax of the intensities of all
                roots in R (i.e. the root with highest intensity
                in image)
                SET n to be the entry of j in e (i.e. n = e[j])
                APPEND i to the list in Segmentation at index n
                SET value of e at position i to be n
                SET value of r at position i to be j
                #Merge all other components represented by roots
                in R.
                FOR each other root k in R
                      SET q to be the entry of k in e (i.e. q = e[k])
                      SET P to be the list in Segmentation at index
                      q
                      APPEND P to the list in Segmentation at index
                      n
                      SET the list in Segmentation at index q to be
                      empty
                      FOR each pixel z in P
                            SET value of e at position z to be n
                            SET value of r at position z to be j
#Delete empty lists.
FOR each entry k in Segmentation
     IF k is empty
          DELETE k from Segmentation
#Each entry in Segmentation now contains a list of pixels,
corresponding to each connected component.
#The vector r is returned for tempTOBLERONE, and may otherwise
be ignored.
RETURN Segmentation, r
```

7.2.2 3DTOBLERONE

```
#Initialise:
#stack is a three-dimensional array (x, y, z), for which each entry contains the intensity of a voxel.
#p is the persistence threshold.
INPUT stack, p
CALL rescale on stack, setting minimum intensity to 0 and maximum intensity to 1
#Calculate filtration - order voxels by intensity.
CALL as.vector on stack to get stackVector
CALL order on stackVector to get filtration
CALL as.array on filtration to get filtrationArray
#Initialise union-find data structure:
#Segmentation stores several lists of voxels, with each list corresponding to one connected component.
#r stores the root voxel of all voxels in a connected component.
#e stores the index in Segmentation of all voxels in a connected component.
#m stores the total number of connected components.
INITIALISE empty list Segmentation
INITIALISE zero vectors r and e with length equal to number of voxels in stack
INITIALISE m to 0
#Perform persistent homology.
FOR each voxel index i in filtration
    GET voxel coordinates of i from stack
    GET coordinates of all voxels in 3 by 3 by 3 neighbourhood around voxel i
    SET filtrationValues as values of all neighbouring voxels from filtrationArray
    SET N to be the set of neighbours of i with entries in filtrationValues lower than i
    IF N is empty
        #This voxel is a root.
        INCREMENT m by 1
        INITIALISE new list at index m in Segmentation containing voxel i
        SET value of e at position i to be m
        SET value of r at position i to be i
    ELSE
        #This voxel may be a root.
        SET R to be the set of roots for all voxels in N (i.e. R = r[N])
        SET potentials to be the set of all roots in R whose intensity in stack is at most p different from the intensity of i in stack
        IF potentials is empty
            #This voxel cannot connect to any existing component in Segmentation.
```

```
            IF intensity of voxel i in stack is greater than
            p
                  #This voxel is a root.
                  INCREMENT m by 1
                  INITIALISE new list at index m in
                  Segmentation containing voxel i
                  SET value of e at position i to be m
                  SET value of r at position i to be i
      ELSE
            #This voxel connects to the root of highest
            intensity.
            SET j to be the argmax of the intensities of all
            roots in R (i.e. the root with highest intensity
            in stack)
            SET n to be the entry of j in e (i.e. n = e[j])
            APPEND i to the list in Segmentation at index n
            SET value of e at position i to be n
            SET value of r at position i to be j
            #Merge all other components represented by roots
            in R.
            FOR each other root k in R
                  SET q to be the entry of k in e (i.e. q = e[k])
                  SET P to be the list in Segmentation at index
                  q
                  APPEND P to the list in Segmentation at index
                  n
                  SET the list in Segmentation at index q to be
                  empty
                  FOR each voxel z in P
                        SET value of e at position z to be n
                        SET value of r at position z to be j
#Delete empty lists.
FOR each entry k in Segmentation
      IF k is empty
            DELETE k from Segmentation
#Each entry in Segmentation now contains a list of voxels,
corresponding to each connected component.
RETURN Segmentation
```

7.2.3 tempTOBLERONE

```
#Initialise:
#stack is a three-dimensional array (x, y, frame). The first two dimensions correspond to pixel coordinates. The third dimension is frame number (time).
#p is the persistence threshold.
INPUT stack, p
INITIALISE empty list frameSegmentations
INITIALISE empty list frameRoots
#Perform 2DTOBLERONE on all frames separately.
FOR each time t
    SET image as the two-dimensional matrix in stack corresponding to the frame at time t
    CALL 2DTOBLERONE with inputs image, p
    GET outputs Segmentation and r
    APPEND Segmentation to frameSegmentations
    APPEND r to frameRoots
#Initialise union-find data structure:
#spatiotemporalSegmentation stores several lists of pixels and the times they are active, with each list corresponding to one connected component.
#spatiotemporalRoots stores the spatiotemporal root pixel of all pixels in a connected component.
#e stores the index in spatiotemporalSegmentation of all pixels in a connected component.
#m stores the total number of connected components.
INITIALISE empty list spatiotemporalSegmentation
INITIALISE empty zero vectors e and spatiotemporalRoots with length equal to number of pixels across all frames in stack
INITIALISE m to 0
FOR each time t
    #Compare each frame to the next frame in series.
    SET Segmentation as the entry in frameSegmentations at index t
    SET Roots as the entry in frameRoots at index t
    SET nextSegmentation as the entry in frameSegmentations at index t + 1 (the next frame)
    SET nextRoots as the entry in frameRoots at index t + 1 (the next frame)
    #Iterate over all components in Segmentation.
    FOR each object in Segmentation
        GET the root r of the object from Roots
        SET N to be the set of all pixels corresponding to the object
        IF entry of r in spatiotemporal roots is 0 (i.e. r has not been visited before)
            INCREMENT m by 1
            INITIALISE new list at index m in spatiotemporalSegmentation
            FOR each pixel i in N
```

```
                SET value of spatiotemporalRoots at position
                i to r
                SET value of e at position i to be m
                APPEND i to list at index m in
                spatiotemporalSegmentation
SET sr to be the spatiotemporal root of r (i.e. sr =
spatiotemporalRoots[r])
SET sm to be the entry of r in e (i.e. sm = e[r])
SET nextN to be the set of all pixels in N at the next
frame (at time t + 1)
GET the roots of all pixels in nextN from nextRoots and
SET to R
GET the spatiotemporal roots of all pixels in nextN
from spatiotemporalRoots and SET to sR
#Check whether a spatiotemporal root has already been
established.
IF sR is empty
      #Check whether number of roots has changed.
      IF R is empty
            #Do nothing, object has disappeared.
      ELSE IF there is only one root in R
            #Connect object across frames by assigning
            the same spatiotemporal root.
            FOR each pixel i connected to this root
                  SET value of spatiotemporalRoots at
                  position i to sr
                  SET value of e at position i to be sm
                  APPEND i to list at index sm in
                  spatiotemporalSegmentation
      ELSE
            #Multiple objects are present in next frame,
            object may have split.
            SET n to be the root in R with brightest
            intensity in stack
            FOR each pixel i connected to n
                  SET value of spatiotemporalRoots at
                  position i to sr
                  SET value of e at position i to be sm
                  APPEND i to list at index sm in
                  spatiotemporalSegmentation
            DELETE n from R
            #Create a new spatiotemporal root for all
            other objects.
            FOR each other root k in R
                  INCREMENT m by 1
                  INITIALISE new list at index m in
                  spatiotemporalSegmentation
                  FOR each pixel i connected to k
                        SET value of spatiotemporalRoots
                        at position i to k
```

```
                                    SET value of e at position i to be
                                    m
                                    APPEND i to list at index m in
                                    spatiotemporalSegmentation
#Each entry in spatiotemporalSegmentation now contains a list
of pixels and the times they are active, corresponding to each
connected component.
RETURN spatiotemporalSegmentation
```

7.2.4 ASMODEUS

```
#Initialise:
#Dmin - Minimum step size.
#Dmax - Maximum step size.
#ROI - A vector of two values representing the width and height
of the region of interest, respectively. The region of interest
is considered from the origin, so pointCloud coordinates must
be translated accordingly.
#times - Number of time steps or iterations of the simulation.
#pointCloud - A matrix or data frame with two columns: x and y
coordinates of each point. Used as a target distribution from
which target K function is drawn.
#rmax - Numeric value, the maximum radius for each K function
to be calculated at.
#nrval - Numeric value, the number of equally-spaced radial
values for each K function to be calculated over.
#target - Manual input for target function. Can be left as null
as long as a target pointCloud is provided.
#numberOfPoints - Number of points to use in simulation.
#initialDistribution - The initial frame of the simulation, used
to specify pre-determined spatial organisation. If left null, a
uniform random distribution will be used.
INPUT Dmin, Dmax, ROI, times, pointCloud, rmax, nrval, target,
numberOfPoints, initialDistribution
#Calculate target K function from pointCloud (if pointCloud is
given).
IF pointCloud is given (not null)
     CALL RipleyKFunction on pointCloud, with r-axis defined on
     the range of 0 to rmax spaced by nrval intervals (see
     Introduction for formula)
     SET output as target
     SET numberOfPoints as number of rows in pointCloud
IF target is not given (null)
     END: not enough information to perform simulation
#Generate initial distribution.
IF initialDistribution is given (not null)
     SET currentCloud as initialDistribution
     SET numberOfPoints as number of rows in initialDistribution
ELSE
     SET currentCloud as a completely spatially random
     distribution, defined over ROI with numberOfPoints points
#Initialise list of frames to store the point pattern at each
time frame.
INITIALISE empty list frames
#Iterate over each time frame.
FOR each time t in times
     #Calculate the K function of currentCloud.
     CALL RipleyKFunction on currentCloud, with r-axis defined on
     the range of 0 to rmax spaced by nrval intervals (see
     Introduction for formula)
```

```
        SET output as currentKFunction
        #Calculate error between currentKFunction and target for
        each point.
        #currentErrors is a list containing the average error for
        each point in currentCloud.
        CALL errorFunction on currentKFunction to calculate error
        between currentKFunction and target for each point (see
        Appendix for formula)
        SET output as currentErrors
        #Offset position of each point in currentCloud.
        FOR each point in currentCloud
            #Calculate offset for each point using stepSize.
            GET error e of point from currentErrors
            CALL stepSize on e with parameters Dmin and Dmax (see
            Appendix for formula)
            SET output as step
            SET randomAngle to be a random number between 0 and 2 *
            pi
            SET coordinates of the point in currentCloud by moving
            point along a path of length step in the direction of
            randomAngle
            #Once all points have been updated, append this new
            cloud to the list of frames.
        APPEND currentCloud to frames
#At the end of this process, frames contains a point cloud for
each time frame.
RETURN frames
```

7.2.5 P-Check

```
#Initialise:
#markedPointPattern - A marked point pattern. Matrix or data frame with first three columns corresponding to x, y and discrete mark.
#radius - Suspected domain radius.
#numberOfTrials - Number of trials to run in the permutation test.
INPUT markedPointPattern, radius, numberOfTrials
#Calculate the distance matrix of all points.
INITIALISE zero matrix distanceMatrix which will store the distance between all pairs of points
FOR each point in markedPointPattern
     FOR each other point in markedPointPattern
          SET corresponding entries in distanceMatrix equal to distance between the points
#Calculate weighted proportion of markedPointPattern.
CALL calculateWeightedProportion with distanceMatrix, radius and marks from markedPointPattern (see Appendix for formula)
SET output as P0
#Perform permutation test:
#Create an empty list to store the weighted proportion of each random trial.
INITIALISE empty list allProportions
#Iterate over all trials.
FOR trial k in numberOfTrials
     CALL sample on third column of markedPointPattern - this randomises the marks
     CALL calculateWeightedProportion with distanceMatrix, radius and marks from markedPointPattern (see Appendix for formula)
     SET output as Pk
     APPEND Pk to allProportions
#Calculate and return p-value.
SET pvalue as percentage of values in allProportions which are greater than P0
RETURN pvalue
```

7.2.6 JOSEPH

```
#Initialise:
#markedPointPattern - A marked point pattern. Matrix or data frame with first three columns corresponding to x, y and quantitative mark.
#radius - Suspected domain radius.
#deviance - Acceptable difference from mean mark value of cluster.
INPUT markedPointPattern, radius, deviance
#Calculate the distance matrix of all points.
INITIALISE zero matrix distanceMatrix which will store the distance between all pairs of points
FOR each point in markedPointPattern
     FOR each other point in markedPointPattern
          SET corresponding entries in distanceMatrix equal to distance between the points
#Determine adjacency matrix of all points.
INITIALISE zero matrix adjacencyMatrix, a logical matrix which stores whether the distance between each pair of points is less than the given radius
FOR each index in distanceMatrix
     IF the entry at the index in distanceMatrix is less than radius
          SET the entry at the same index in adjacencyMatrix to be 1
#The neighbourhood of each point is now stored in adjacencyMatrix.
INITIALISE empty vector differencesVector
FOR each point in markedPointPattern
     GET the point's neighbourhood from adjacencyMatrix
     GET average mark m1 of all points in the neighbourhood
     GET the point's mark m2 from markedPointPattern
     APPEND absolute difference between m1 and m2 to differencesVector
#Calculate filtration - order points by value in differencesVector.
CALL order on differencesVector to get filtration
#Initialise union-find data structure:
#Clusters stores several lists of points, with each list corresponding to one connected component (cluster).
#r stores the root of all points in a connected component.
#e stores the index in Clusters of all points in a connected component.
#m stores the total number of connected components.
INITIALISE empty list Clusters
INITIALISE zero vectors r and e with length equal to number of points in markedPointPattern
INITIALISE m to 0
#Perform persistent homology.
FOR each point index i in filtration
```

```
        GET neighbourhood of point i from adjacencyMatrix
        SET N to be the set of neighbours of point i with entries in
        filtrationValues lower than i
        IF N is empty
              #This point is a root.
              INCREMENT m by 1
              INITIALISE new list at index m in Clusters containing
              point i
              SET value of e at position i to be m
              SET value of r at position i to be i
        ELSE
              #This point may be a root.
              SET R to be the set of roots for all points in N (i.e.
              R = r[N])
              SET potentials to be the set of all roots k in R such
              that the absolute difference between the mark of k and
              the mark of i is at most deviance
              IF potentials is empty
                    #This point cannot connect to any existing
                    component in Clusters. This point is a root.
                    INCREMENT m by 1
                    INITIALISE new list at index m in Clusters
                    containing point i
                    SET value of e at position i to be m
                    SET value of r at position i to be i
              ELSE
                    #This point connects to the root with lowest
                    value in differencesVector.
                    SET j to be the argmin of the values of
                    differencesVector for all roots in R
                    SET n to be the entry of j in e (i.e. n = e[j])
                    APPEND i to the list in Clusters at index n
                    SET value of e at position i to be n
                    SET value of r at position i to be j
                    #Merge all other components represented by roots
                    in R.
                    FOR each other root k in R
                          SET q to be the entry of k in e (i.e. q = e[k])
                          SET P to be the list in Clusters at index q
                          APPEND P to the list in Clusters at index n
                          SET the list in Clusters at index q to be
                          empty
                          FOR each point z in P
                                SET value of e at position z to be n
                                SET value of r at position z to be j
#Each entry in Clusters now contains a list of points,
corresponding to each connected component.
RETURN Clusters
```