AUTOMATED QUANTITATIVE ANALYSIS METHODS FOR TRANSLOCATION OF BIOMOLECULES IN RELATION TO MEMBRANE STRUCTURES

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades einer Doktorin der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Master of Science in Life Science Informatics

OLGA DOMANOVA

aus Sankt Petersburg, Russische Föderation

Berichter: Universitätsprofessor Dr. rer. nat. Thomas Berlage
Universitätsprofessor Dr. rer. pol. Matthias Jarke
Universitätsprofessor Dr. med. Fabian Kießling

Tag der mündlichen Prüfung: 18. Oktober 2013

Abstract

Biological processes are complex study objects due to their dynamic nature and structural diversity of living organisms. To study dynamic processes statistically, numerous experiments with multiple observations have to be performed, and data have to be analyzed and evaluated. Owing to great technological advances, gigabytes of data are being acquired both in research and industry. Slow and subjective manual analyses are not sufficient anymore, and automated evaluation methods are required.

The distribution of biomolecules provides valuable information on a current biological state. The distribution of biomolecules depends on and is influenced by functions of biomolecules, and may thus be used to detect abnormalities. The relatively young research field *toponomics* describes the laws of spatial arrangement of molecules. Several evaluation methods have previously been developed, automatized and standardized. However, no standard evaluation methods have been reported to quantitatively analyze such an important biological process like translocation of biomolecules.

Translocation processes are vital for living organisms. For instance, substance inclusion into a cell or exclusion from it represent a translocation. Furthermore, signaling biomolecules translocate from the cytoplasm across the nuclear membrane into the nucleus to influence gene and protein expression. Investigating translocation processes may help to understand complex biological functions. It may also be used to analyze signaling events, or may even be employed for diagnostics and therapy monitoring.

Manual and case-specific methods for quantitative translocation analysis are known, but fail to be generally applicable. Therefore, I have developed a novel generic automated approach. The method is based on microscopy images of biological samples. I have defined a generic method to quantitatively express distribution of biomolecules in numeric descriptors. Herewith, changes in distribution may be analyzed using different biological samples. Thus, the samples analyzed do not necessarily have to belong to a time series.

Furthermore, not only cell cultures, but also tissue samples can be used for the analysis. Evaluations of cell cultures are simpler due to homogeneity and spatial separation of individual objects. However, structural polarity of the cells can be seen only in tissues.

I have developed two workflows based on numeric descriptors for the distribution of biomolecules. The first workflow uses structure detection in images to localize the objects for evaluation. The second workflow avoids this complex operation by a structure-independent information extraction strategy. Both workflows are generic and may be applied to quantify a wide range of translocation processes.

Kurzfassung

Biologische Prozesse sind komplex aufgrund ihrer Dynamik und struktureller Vielfalt lebender Organismen. Um dynamische Prozesse statistisch zu untersuchen, sollten zahlreiche Experimente mit mehreren Beobachtungen durchgeführt werden. Dank der großen technologischen Fortschritte werden Gigabyte von Daten erfasst, sowohl in der Forschung als auch in der Industrie. Diese müssen analysiert und ausgewertet werden. Mühsame und subjektive manuelle Analysen sind nicht mehr ausreichend, und automatisierte Auswerteverfahren sind unerlässlich.

Die Verteilung von Biomolekülen liefert wertvolle Informationen zum aktuellen biologischen Zustand von Organismen. Der biologische Zustand hängt von der Aktivität von Biomolekülen ab. Deswegen kann die Verteilung von Biomolekülen analysiert werden um Anomalien zu detektieren. Das relativ junge Forschungsfeld *Toponomics* beschreibt die Gesetze der räumlichen Anordnung von Molekülen. Mehrere in der Literatur beschriebene Auswerteverfahren wurden automatisiert und standardisiert. Es gibt jedoch keine Standardauswerteverfahren für quantitative Analyse eines wichtigen biologischen Prozesses wie die Translokation von Biomolekülen.

Translokationsprozesse sind lebenswichtig. Aufnahme von Substanzen in eine Zelle oder deren Ausschluss stellen, zum Beispiel, eine Translokation dar. Außerdem translozieren Signalmoleküle aus dem Zytoplasma durch die Kernmembran in den Zellkern, um Gen-und Proteinexpression zu beeinflussen. Untersuchung der Translokationsprozesse könnte helfen, komplexe biologische Funktionen zu verstehen. Translokationsprozesse könnten auch analysiert werden, um Signalwege zu identifizieren, oder sogar um Diagnose- und Therapieverfahren zu entwickeln.

Unterschiedliche Methoden für die quantitative Analyse von Translokation sind bekannt, aber meist manuell oder fallspezifisch und damit nicht allgemein anwendbar. Deshalb habe ich einen neuen generischen automatisierten Ansatz entwickelt. Das Verfahren basiert auf Mikroskopiebildern von biologischen Proben. Ich habe eine generische Methode definiert, um die Verteilung von Biomolekülen in numerischen Deskriptoren quantitativ auszudrücken. Hiermit können Veränderungen in der Verteilung in unterschiedlichen biologischen Proben analysiert werden. Dies ist vorteilhaft, da die Methode unabhängig davon funktioniert, ob Einzelbilder vorliegen, oder Zeitreihen.

Darüber hinaus können nicht nur Zellkulturen, sondern auch Gewebeproben zur Analyse verwendet werden. Auswertungen von Zellkulturen sind einfacher aufgrund der Homogenität und räumlicher Trennung der einzelnen Objekte. Allerdings kann strukturelle Polarität der Zellen nur in Gewebe sichtbar sein.

Ich habe zwei Workflows basierend auf numerischen Deskriptoren für die Verteilung von Biomolekülen entwickelt. Der erste Workflow setzt Strukturerkennung in Bildern ein, um Objekte für Auswertung zu lokalisieren. Der zweite Workflow vermeidet diese komplexe Operation durch eine strukturunabhängige Informationsextraktionstrategie. Beide Workflows sind generisch und könnten angewendet werden, um eine breite Palette von Translokationsprozessen zu quantifizieren.

Acknowledgements

The work on this dissertation was a great experience due to a successful combination of an interesting research topic, available facilities and a motivating working environment.

I would like to thank my colleagues at Fraunhofer FIT who helped me to get acquainted with the available technical basis and were supportive at all times. My supervisor Professor Thomas Berlage has inspired and guided my work by interesting and constructive discussions. His comments and critics have significantly influenced the whole development and this writing in particular. He was the first professor that I met in my (not very long yet) academic career, who invested so much time in reading and correcting. Stefan Borbe was a great support for the programming side, immediately resolving any arising issue. Dr. Peter Wißkirchen has not only advised me on mathematical and statistical questions but also shared his great experience in image processing.

Interdisciplinary research is strongly influenced by and even dependent on the existing collaborations. Therefore, I would like to thank our colleagues from the University Clinic Düsseldorf for their effort. Professor Dieter Häussinger and Professor Ralf Kubitz supervised the biological research and experiments, while also being very helpful in interpreting statistical results. Dr. Stefanie Kluge has standardized biological SOPs which are the basis of all our experiments. Fruitful discussions with her made it possible to translate some biological knowledge into technical descriptions, which is a great achievement.

To my teachers and professors in Saint-Petersburg, I am also very thankful. They showed me to enjoy studying, to have fun with it and to be sure that the more I learn, the more is still left to be learned.

Further, I would like to thank DAAD for the scholarship that permitted me to conduct my master studies. It was the first step towards this doctoral thesis. I am especially thankful to Doctor Markus Mathyl, the former DAAD lector in Saint-Petersburg. He was very supportive, helped me a lot in finding information and going through the Russian-German burocracy.

This work, as many other things in my life, would not have been possible without the support of my family. I am very grateful to my husband for motivating me all the time, reminding me that there's a life beyond the laptop screen, and, of course, for the technical support at home. A very special gratitude goes to my mother who was and is always there for me, sacrificing everything just to make my dreams come true. My son, my little sunshine, I thank you for the endless happiness that I feel only from looking at you :) You are the one who showed me the importance of time management better than any professor ever could :)

And last but not least, DFG (KFO-217) financially supported the project that included my doctoral thesis. I also am very thankful to the B-IT Research School for the scholarship that enabled this research.

Olga Domanova
Bonn
June, 5^{th}, 2013

Contents

1. Introduction

Living organisms are among the most complex and at the same time the most interesting subjects to study. Several common features distinguish them from the nonliving matter. They are able to grow, to reproduce and to react to external stimuli. Their underlying complexity is owing to the multi-level organization. All living organisms consist of cells, which are usually too small to be visible by a human eye. Cells are organized in tissues, which build organs. Individual organs constitute organ systems, which, in turn, compose whole organisms.

Such a complex organization is also very structured at the sub-cellular level. Sub-cellular structures, such as cytoplasm, organelles and a nucleus (for eucaryotes) precisely subdivide the matter which is essential to perform highly complex functions. Due to such a complex organization, only very small organisms, like *Escherichia coli* have been completely studied. Their genomes have been decoded, proteins have been identified and their interactions have been studied. The great majority of the living organisms, however, is still too complex to be analyzed in such a detail.

To study a biological process, molecular actors can be observed. Microscopy is one of the methods widely used for this purpose. Images provide information on the morphology of a biological sample and distribution of labeled biomolecules (DNA, RNA and proteins). Spatial distribution of biomolecules is crucially important for the normal functioning of an organism. They have to appear at the right time and the right place to be involved in the biological processes they are required for. This balance is dynamic and conditions the state of the cells. For example, transport proteins integrate into membranes to transport molecules from one side of the membrane to another. Outside the membrane, they can not perform their function anymore. Consequently, fewer molecules will be transported across the membrane. The existing balance will be altered, which may cause various dysfunctions. This mechanism is used, for instance, to regulate bile transport through the liver cells. Malfunction of this process results in cholestatic liver diseases.

This simple example illustrates the connection between distribution and function of biomolecules. *Toponomics* is the field of research describing the laws of spatial arrangement. It accounts for the temporal and spatial organization of biological molecules within the structures of the organism [113, 145, 180]. Toponomics has been applied in various areas of biological research for disease diagnosis, therapy monitoring, and evaluation of novel treatment options.

To study dynamic processes statistically, numerous experiments with multiple observations have to be performed, and data have to be analyzed and evaluated. Many of such analyses are still carried out manually. "While nothing can fully replace the expertise of a trained biologist, observing many samples by eye is time-consuming, subjective, and non-quantitative" (Lamprecht et al. 2007, p. 71). The results of such an evaluation will be error-prone simply due to human factors. The automated evaluation is a better option, as it does not only save valuable time, but also significantly speeds up the process and increases the number of measurement points. Subjective decisions can be substituted by automated classification. These characteristics are essential for any application to have a potential in high-throughput, pharmaceutical or industrial use. Automated methods have to be developed to avoid manual labor-consuming and imprecise image evaluation. This is an important prerequisite for accurate toponomics studies.

1.1. Translocation processes

Distribution of biomolecules is closely connected to and dependent on transport processes. *Transport* of biomolecules is their movement from one location to another. Transport processes

are ubiquitous in living organisms, underlying nerve conduction and muscle contraction, digestion, kidney function and the nourishment of every cell in the body [52]. If biomolecules are moving from one compartment to another, we talk of a translocation. *Translocation* processes are also widespread. First, the compartmentalization principle leads to the presence of numerous organelles. These are closed aqueous volumes that perform different functions and are separated from the cytoplasm by membranes. Concentration of biomolecules in organelles can differ from their concentration in the cytoplasm. This serves as a basis for translocation along concentration gradients (across the membranes). Second, "[...] membranes are not just containers: they are coordinators and sites of major activity" (Cardelli 2005, p. 157). Biomolecules frequently integrate into membranes, which are functional region by themselves. Third, substance exchange between the cell and the environment represents another translocation case. Inclusion of a substance into a cell (endocytosis) or exclusion (exocytosis) proceed across outer cell membranes.

According to this definition, all translocations relate to a membrane which either represents a barrier or a compartment by itself. Quantitative analysis of translocation is useful for research and has a potential for clinical applications.

1.1.1. Translocation quantification

To quantify a transport process, biological samples have to be prepared and evaluated. Microscopy images acquired at certain moments in time represent snapshots of molecule distribution in a continuous transport process. Although biological processes are three-dimensional in their nature, they are often analyzed based on two-dimensional images to reduce information volume and to speed up calculations.

Biomolecules are usually smaller than the resolution of the optical microscopy. Their distribution can be evaluated by densitometry [16, 20]. *Densitometry* generally refers to the analysis of optical density. When working with image data, densitometry extracts information on intensity distribution. Biomolecules may be labeled with fluorophores, substances that can be excited by light of a certain wavelength and that re-remit lower energy light. Densitometric analysis of such fluorescence microscopy images yields data on the intensity distribution of the fluorophores used. The concentration of the labeled biomolecules can then be inferred from the measurement of the fluorescence intensity at the respective positions in the image. The density distributions can be compared and provide the basis for the quantitative analysis of the underlying translocation process.

Translocation of biomolecules is a continuous process. Its dynamics will be influenced by certain environmental changes. For example, liver cells (hepatocytes) respond to hyper-osmolar medium by internalization of bile transporter proteins from membranes into cytoplasm. Hence, biological samples can be prepared before and after such a treatment, representing idealized *initial* and *final* states. Comparison of the bile transporter distribution in the samples before and after the exposure to hyper-osmolar medium should then quantitatively describe the translocation that took place.

Evaluation of the distribution change is straightforward for the images from one time series. They can be overlaid to quantify differences in densities. This analysis becomes much more complex when evaluating a translocation based on two separate biological samples. In this case, overlay of completely different images would not yield any information on the translocation. Thus, another quantification method is required. For instance, translocation could be assessed as the relative change in molecule densities. Therefore, distribution of biomolecules in each image needs to be normalized with respect to some structure. *Reference structure* needs to be stable and visible under various experimental conditions. Membranes may be used as reference structures for translocation quantification.

Membranes need to be localized in images to quantitatively analyze translocation in relation to them. This can be achieved by fluorescent labeling of the membranes and consecutive structure detection in images. *Structure detection* is often among the first image processing steps and represents "[...] the most challenging part of image analysis" (Ljosa and Carpenter 2009, p. 2).

Especially for biological images, it is very complex and is frequently very specific. Due to the natural variety and structural diversity, one and the same structure may look completely different in the images taken from multiple biological samples. Furthermore, biological structures can be damaged in the course of the sample preparation. Membranes, for instance, may be distorted during freezing, cutting or fixation. They will therefore exhibit a structure differing from the intact membranes. Image quality may also differ due to the settings of resolution, focus or staining quality. Structure detection is already complex for individual cells and is even more problematic for tissue sections.

Acceptably good structure detection has to account for such a natural diversity and a structural variety. A wide range of algorithms exist, ranging from a simple thresholding [64, 121] to machine learning [107, 163]. A suitable algorithm usually requires prior knowledge on the structure studied and is oftentimes very specific. An image processing workflow including sophisticated structure detection will thus be optimized for a certain biological structure and will most probably not be applicable for another one. Unlike a human eye, automated algorithms are not easily adaptable to the variation of the input data.

1.2. Aim

Translocations are widespread in living organisms, and their quantitative analysis may help to understand vital biological processes. My aim is to develop generally applicable methods for the quantitative analysis of translocation based on microscopy images of tissue samples.

> Goal 1: I will synthesize a general approach to translocation quantification via structure detection in images.

In this approach, I will employ structure detection techniques based on machine learning as the first step. A suitable classification model is built every time based on the given input data. Hence, the advantage of this approach is that it is not optimized for any specific biological structure. Machine learning for structure detection in biological images is quite popular and is used in powerful tools like Zeta [147] and CellProfiler [23]. This workflow for translocation quantification will be referred to as the *structure-based workflow*.

> Goal 2: I will develop an alternative method for quantitative translocation analysis that does not require structure detection in images.

Instead, I will develop a method to identify membranes by analysis of intensity distribution of their markers. For this purpose, I will extract 1D distribution profiles by line densitometry along random lines. Peak detection in the distribution profiles extracted will yield the central position of the membrane. It will be a general approach for densitometry-based translocation analysis that does not rely on other structural features than membrane extension. This algorithm will be more general than the first one, and will thus be applicable to numerous biological questions. Its advantage over the first algorithm is the absence of a complex structure detection, which is frequently too specific and may be slow. This workflow for translocation quantification will be referred to as the *random lines-based workflow*.

These automated workflows are instances of a novel approach to the quantitative analysis of translocation. In the next Section I will suggest a generic method for comparison of molecule distributions by calculation of numeric descriptors.

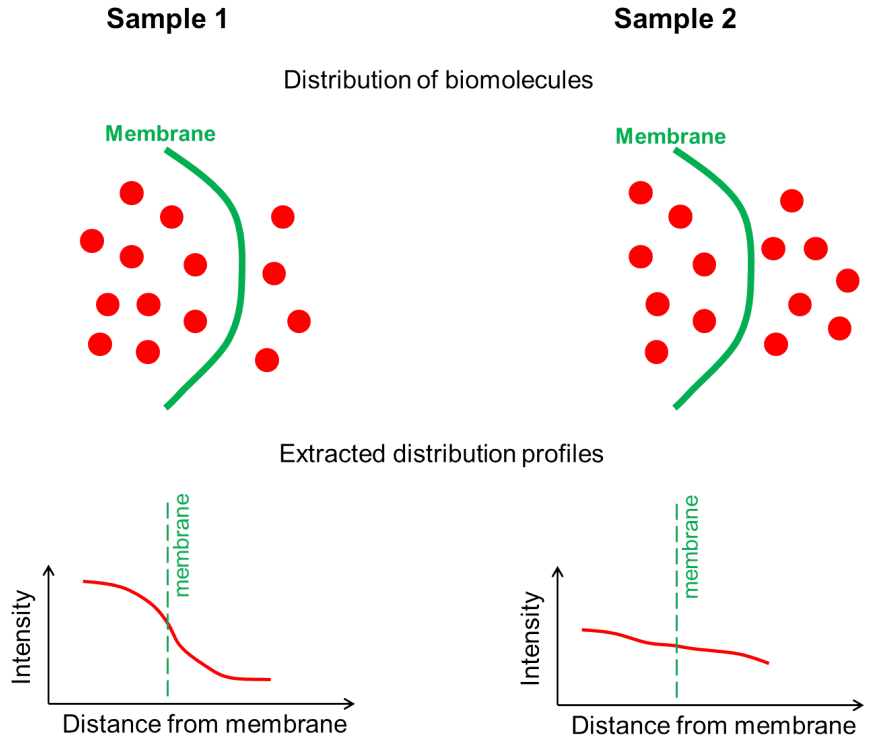


Figure 1.3.1.: Translocation of biomolecules analyzed on two biological samples (1 and 2). Distribution profiles are extracted for the biomolecule of interest (labeled red). These profiles are analyzed in relation to the membrane (labeled green).

1.3. Novel approach to translocation quantification

A translocation of biomolecules is assumed if their distribution differs in two analyzed biological samples (see Figure 1.3.1). This distribution can be described as a complex function f:

$$\textit{distribution of biomolecules} = f\ (\textit{time, initial biological conditions, biological structure, concentration, external influence,..}),$$

where most of the parameters are unknown, except for the time and external influence (experimental conditions). Neither all the parameters, nor the function f itself might be precisely determined. However, the shape of this function may be extracted by densitometric analysis. Line densitometry extracts 1D intensity vectors of fluorophores that biomolecules are labeled with. Assuming the stoichiometric binding, concentration of the biomolecules is proportional to the intensity of the fluorophores. Hence, the vectors extracted, or *profiles*, represent the unknown molecule distribution function f. I use these profiles for the quantitative analysis of translocation. This is the first information reduction step: translocation is analyzed on 1D vectors which is computationally more efficient than evaluating 2D image data.

If membranes have already been detected in images, distribution profiles may be extracted in relation to them (e.g. orthogonally). Each image, or a dataset of images, is now represented by a set of distribution profiles. If translocation of biomolecules occurs in the sample 2 relative to the sample 1, the shapes of the respective functions of molecule distribution (f_2 and f_1) will be different. Therefore, comparison of such distribution profiles should indicate a translocation.

The result of the quantitative translocation analysis depends on the quality of the profiles compared. If profiles were extracted at damaged biological structures or the noise level is high, the results will be unreliable. Therefore, those profiles need to be selected that have

been extracted at intact biological structures and that more clearly depict them. I will suggest a *profile selection* strategy for this purpose. To avoid bias, selection will be performed on the profiles of the membrane marker only, disregarding the corresponding distribution of the biomolecule of interest.

As distribution functions (f_1 and f_2) are unknown, a general method has to be developed for their comparison. This method should be able to quantify translocation of biomolecules by analysis of distribution profiles extracted from images of different biological samples. For this purpose, my approach uses numeric parameters of distribution profiles. These *descriptors* mathematically express distribution of biomolecules in relation to the membranes. For instance, the ratio of the marker intensity of biomolecules on the right and on the left side of the membrane may be such a numeric descriptor for the situation in Figure 1.3.1. There, the ratio will be larger in the sample 2 than in the sample 1.

Numeric descriptors are also used to quantitatively analyze nucleus-to-cytoplasm translocation [133, 171]. There, the ratio of the marker intensity in the nucleus relatively to its intensity in the cytoplasm is computed. This ratio describes distribution of biomolecules and may be used to indicate a translocation. Generally, construction of descriptors depends on a particular biological structure and an aspect of distribution that is assessed. I will suggest a general approach for their development.

Descriptors will be used as parameters of distribution of biomolecules in the given images. As numeric variables, descriptors can be compared statistically. Herewith, translocation of biomolecules can be quantitatively analyzed. It will even be possible for the images which are not a part of a time series, as descriptor calculation is performed for each image separately. This is the core idea of the novel approach to translocation quantification.

Calculation of descriptors also has another advantage. It represents the second information reduction step: translocation is now analyzed on single numbers which is computationally even more efficient than evaluating 1D vectors.

1.4. Contributions

The basic workflow consists of generic procedures that are either applicable to a wide range of translocations, or can easily be modified.

First, structure detection by machine learning adapts to the input data. Various biological structures will be detectable by this algorithm.

Second, a novel algorithm avoiding structure detection is even more general by its implementation.

Third, the approach to describe distribution of biomolecules using numeric descriptors is also generic. For many translocations, a set of numeric descriptors can be suggested to quantify distribution of biomolecules.

Fourth, a descriptor represents the full shape of a one-dimensional distribution function by a single numeric variable that captures any significant change in distribution as a change of value. Information reduction to single numbers will speed up calculations.

Generally, the approach is expected to be able to quantify a wide range of translocations. However, it is essential that biomolecules of interest, as well as reference structures (membranes) can be labeled. The respective markers will be detected in images and used for the evaluation. Owing to the advances in labeling techniques, these requirements should not exclude too many translocations.

1.5. Potential scientific and clinical impact

Translocation is a vital biological process. Hence, its quantitative analysis may be useful in research and in clinical practice. Translocation processes are complex and may be regulated on various levels. For instance, translocation of bile transporters may be regulated on a short-term

scale by aniso-osmolarity, while it is genetically regulated on a long-term scale. Many regulation mechanisms are unknown, although they may provide insight to vital biological processes and help to develop novel treatment options.

Translocation quantification may be applied to study regulation mechanisms. For example, Cantore et al. [20] investigated mechanisms of hyperosmolarity-induced retrieval of bile transporters from canalicular membrane in hepatocytes. Genes encoding certain proteins have been activated or deactivated, and their influence on the retrieval of bile transporters was studied. Quantitative translocation analysis was used to elucidate the underlying signaling events and to identify proteins that mediate this translocation. Furthermore, as bile acids influence glucose homeostasis and lipid metabolism, proteins mediating translocation of bile transporters may also be important molecular actors in these vital biological processes.

This example illustrates that quantitative analysis of translocation may help investigating regulatory pathways in systems biology. It may also be applied to evaluate the action of drugs influencing translocation or be used for screening. Furthermore, I expect the method to find an application in the diagnosis of diseases that are caused by or result in abnormal distribution of biomolecules. Treatment progress could also be monitored by the same algorithm. The other way round, biological and medical findings on the translocation processes may guide a further development and modification of the automated quantification workflow. This may also help to broaden the application area of the method.

This is a good example of a *translational research* which represents a bridge between the basic and applied science [37, 106]. It is defined as the "[...] process of applying ideas, insights, and discoveries generated through basic scientific inquiry to the treatment or prevention of human disease" (Fang and Casadevall 2010, p. 563). "Translational research transforms scientific discoveries arising from laboratory, clinical, or population studies into clinical applications to reduce morbidity and mortality" (Translational Research Working Group. http://www.cancer.gov/researchandfunding/trwg/TRWG-definition-and-TR-continuum, accessed on 20.05.2013).

1.6. Study design

To validate the novel approach to translocation quantification, a set of algorithms has been developed and tested on several biological translocation processes selected as models. The automated workflows will be validated against the manual evaluation, which is routinely performed by my collaboration partners. I will also compare the automated workflows to each other in terms of statistical results and speed. Significance of the translocations detected by the automated methods should be comparable to the one evaluated manually. Evaluations are performed to test the general applicability of the novel algorithms. I will also show that the results of the structure-based and the random lines-based workflows are of comparable quality, as long as enough observation points are present in the images.

1.7. Model biological processes

The work of this thesis was carried out in the framework of a DFG-funded clinical research group "Hepatobiliary transport and liver diseases" (KFO-217 [88]). Biological processes in hepatocytes (liver cells) and their connection to cholestatic liver diseases are the central research topics of this project. In particular, transport proteins and enzymes are observed that play a key role in vital biological processes. Translocation of these proteins takes place in response to certain external or internal factors. Quantitative analysis of these translocations is expected to aid investigating the underlying biological processes.

1.7.1. Diseases associated to the translocation processes studied

Liver is a multifunctional organ with a prominent role in metabolism and its dysfunctions can lead to severe diseases. Cholestatic liver diseases constitute only a subset of possible disorders but represent a major clinical and socio-economical issue [32, 74]. For instance, Progressive Familial Intrahepatic Cholestasis (PFIC) is caused by dysfunction of bile transporters. PFIC is heritable and often leads to death from liver failure [32].

Cholestatic diseases can also be caused as side effects of medication. The liver is constantly exposed to the ingested products as it performs a vital barrier function [96]. Hence, drugs directly damage hepatocytes, which may lead to hepatitis, cholestasis, cirrhosis, vascular lesions and even liver failure [96].

Pregnancy can also be a reason for a cholestatic liver state. The Intrahepatic Cholestasis of Pregnancy (ICP) is characterized by maternal pruritus and jaundice [56]. Interestingly, ICP is more probable for women carrying twins, triplets or more [117]. Although maternal effects of ICP are mild, there is a clear association between ICP and higher frequency of preterm delivery and sudden fetal death [90].

Another disease associated to the studied translocation processes is diabetes. Diabetes is a widespread disease that caused 4.6 million deaths only in 2011 [119]. It arises when blood sugar level is high due to the impairment of its metabolism. The liver plays a key role in the regulation of glucose output and uptake [79, 104]. Glucokinase is the enzyme responsible for the initial step in the metabolism of glucose. Alterations of the glucokinase function can lead to Maturity Onset Diabetes of the Young (MODY) [68]. MODY is usually detected during routine screening, as the symptoms are not always clearly recognized [68, 170].

1.7.2. Translocation processes studied

Three translocation scenarios are studied in the KFO-217. The first two processes are movements of bile transporter proteins between the plasma membranes and the cytoplasm of hepatocytes. Bile is produced in hepatocytes and is exported into a network of small channels - canaliculi. It is then directed into the intestine to aid the process of digestion. The canalicular and basolateral membranes of hepatocytes contain bile transporter proteins. For example, the bile salt export pump (Bsep) and the sodium taurocholate cotransporting polypeptide (Ntcp) are integrated into these membranes and export bile. Under certain conditions, these proteins are internalized into the cytoplasm and can not perform the transport function anymore. Up-regulated or down-regulated bile transport can lead to cholestatic or choleretic liver state.

The third translocation process investigated is a case of a nucleus-to-cytoplasm translocation. The enzyme glucokinase moves from cytoplasm into the nucleus and back depending on the glucose concentration in the medium. The more glucose is added, the more glucokinase translocates from the nuclei (inactive state) into the cytoplasm (active state). Glucokinase performs the first step in the glucose metabolism. Impairment of this process can lead to increased blood glucose and result in diabetes.

These processes represent a translocation, as biomolecules move from one compartment to another. In the first two processes, a membrane is directly involved. In the third process, biomolecules are translocated across the nuclear membrane. Hence, these three translocations are suitable test cases for the development of the automated quantification method that requires a reference structure. Furthermore, these model processes represent two widespread translocation types. First, biomolecules frequently integrate into membranes, functioning as coordinators and sites of major activity [21]. And second, nucleus-to-cytoplasm translocation is important for regulation of gene activity by integration of special biomolecules into the nucleus from the cytoplasm. Generally applicable quantification methods developed for these model processes should suit to a wide range of other translocations.

1.8. Structure of the work

As the approach is based on image data, Chapter 2 will start with an introduction to important imaging techniques. A special focus is made on microscopy as one of the most popular acquisition methods for biological images. Further, I will present an exemplary workflow for the processing of biological images. This chapter is concluded by presentation of established toponomics evaluation methods, including colocalization analysis and densitometry.

Chapter 3 describes the biological processes taken as models for my work and their importance for the normal functioning of an organism. A typical manual analysis of translocation and the established automated evaluation of the nucleus-to-cytoplasm translocation are presented. And finally, I describe biological standard operating procedures developed by my collaboration partners for the sample preparation.

The next Chapter 4 explains the novel approach to translocation quantification. I derive a generalized model for frequent translocations proceeding orthogonally to a membrane. I suggest quantitative analysis of such processes in relation to the membrane using numeric descriptors of molecule distribution. Therefore, I define a workflow based on membranes detected in biological images (structure-based workflow). Further, I suggest another workflow avoiding complex structure detection in images by a different information extraction strategy (random lines-based workflow).

To validate this approach, I have developed a set of algorithms. These are described in Chapter 5. There, I present implementation details together with the optimized algorithm parameters. First, the structure-based and the random lines-based workflows are implemented for the translocations of membrane transporters. Further, the random lines-based workflow is shown to be adaptable to a different type of a translocation process, namely to a nucleus-to-cytoplasm translocation.

My methods are then evaluated and compared to the established methods in Chapter 6. I also compare the developed algorithms to each other in terms of statistical results and speed. Achievements and encountered difficulties are discussed in Chapter 7. Individual processing steps are evaluated there, and their contribution to the good results is discussed.

Chapter 8 summarizes my findings. I will again discuss limitations regarding the covered processes. And finally, future work is presented in Chapter 9.

2. Acquisition, processing and toponomics analysis of biological images

Millions of biological images are routinely generated and evaluated owing to the great technological advances. I will therefore start this Chapter with a short overview on the most popular imaging approaches. A special focus is made on microscopy, as it is a well-established technique and is often the method of choice for numerous biological investigations. Automated microscopes are also widely used in high-throughput industrial applications.

Acquired images represent a rich information source on a biological state. To extract this information, images have to be processed and analyzed. Although a great variety of possible acquisition techniques and biological analyses exist, a short processing workflow can be generalized for them. I will present this workflow and describe important image processing operations constituting it. Segmentation of biological images, for instance, is discussed as it is among the first and crucial steps for any further analysis.

Biological images represent an especially valuable information source, because distribution of biomolecules can be studied on them. Distribution of biomolecules is directly connected to biological functions and influences them. As presented in Chapter 1, toponomics is the field of research describing the laws of spatial arrangement. I will thus continue by presentation of some prominent approaches for a toponomics analysis. Several tools will be described that have been developed to study distribution of biomolecules. Some of them are used to infer biological function or potential interactions of biomolecules from the location data. Nevertheless, an important aspect like *translocation* of biomolecules has not been sufficiently studied yet. Generally applicable translocation quantification methods are still lacking.

2.1. Biological imaging

Since ancient times, imaging has been the eyes of science. Modern imaging technologies allow for examination of complete organisms, organs, tissues, cells and even sub-cellular structures. Multi-dimensional and multi-parameter data can be acquired, visualized and analyzed. Imaging is widely used to measure physical parameters such as concentration, tissue properties, and surface area [44] and to explore a biological function [86]. The acquired data is a valuable input for building biological and mathematical models. The other way round, existing hypotheses or computational models can be verified by analyzing biological images.

"Biological material is organized in multiple layers of a structural hierarchy; thus, visualizing biological structures and processes entails a multilevel approach that combines different kinds of specimen preparation, visualization tools, and computational analysis" (Frank et al. 2002, p. 85). In the following, several widely used imaging techniques will be briefly presented with their application areas and limitations in the order of the spatial resolution increase.

2.1.1. Magnetic resonance imaging

Magnetic Resonance Imaging (MRI) is performed by applying a strong magnetic field (normally 1.5 or 3 Tesla for humans), which aligns the hydrogen nuclei spins in a direction parallel to the field [86]. A Radio Frequency (RF) pulse is applied to the sample which makes the spins acquire enough energy to tilt and precess, where an RF receiver can record the resulting signal [75]. The spins realign parallel to the main magnetic field after the RF pulse was removed. The signal strength decreases in time at a time constant T2 which is always less than the realignment

constant T1. Magnetic gradients localize spins in space and, an MR image is formed. Spin density differs among tissues in a heterogeneous specimen and enables the excellent tissue contrast [75]. Increasing magnetic field strength improves resolution, but can disturb the visual system and lead to peripheral nerve stimulation. Still, higher magnetic field strengths are applied for animal studies (up to 9 Tesla).

Functional Magnetic Resonance Imaging (fMRI) is used to image brain activity in response to specified stimuli. When a stimulus solicits a response from a certain area of the brain, metabolism in that region increases, and more oxygenated hemoglobin is found in the region [84]. The balance between oxygenated and deoxygenated hemoglobin is altered leading to a change in image contrast [123]. The obtained image is then compared with measurements under normal conditions, and a change is detected.

MRI is widely applied in both experimental and preclinical studies. Cell proliferation, bone structure, even individual cells and organelles can be imaged owing to a spatial resolution of $<$ 4 μm [86]. fMRI, in turn, is used to study the functions of the living brain in a non-invasive manner. For instance, fMRI can determine the areas of brain tissue that are functionally important. A brain tumor can thus be treated without damaging these areas [86].

MRI provides excellent tissue contrast and lacks ionizing radiation. However, high magnetic field strength is problematic for humans because of arising physiological effects. fMRI needs faster temporal resolution to acquire images of dynamic brain activity. Hence, the spatial resolution is reduced to the order of millimeters.

2.1.2. Computed tomography

Organ components (fat, water, bone and air) exhibit different X-ray absorption, leading to contrast in Computed Tomography (CT). In CT, a low energy X-ray source and a detector rotate around the subject, acquiring volumetric data [86]. For animal studies, higher energy X-rays can be applied than in the human scanners. Increased energy improves resolution, but exposes the specimen to more ionizing radiation with adverse health effects.

CT tissue-tumor contrast can be improved by introduction of iodinated contrast agents. CT can be used to image lung tumors and bone metastasis, given its fast imaging time and high spatial resolution (12-50 μm). Its spatial resolution is sufficient to visualize fine anatomical details. Still, repeated imaging for humans are not recommended due to a relatively high radiation dose of CT [105].

2.1.3. Microscopy

Microscopy provides greater spatial resolution (down to nm) allowing for imaging of individual cells, organelles and even molecules. It is a powerful tool for information extraction and analysis of the biological data. Stoffler et al. classify primary advantages of microscopes into three categories: "1) they directly produce images rather than diffraction patterns or spectras; 2) they enable us to explore biological structure at all levels from the macroscopic to the atomic scale; and 3) they allow biological matter to be imaged in its functional environment" (Stoffler et al. 1999, p. S195).

Microscopy has been extremely important for biological experiments and evaluations, and the microscope technology has evolved mightily since the first single-lens instruments [162]. It was the first imaging technique developed for studying small objects otherwise invisible for the human eye. There exist three main branches of microscopy, namely, optical, electron and scanning probe microscopy. Electron microscopy is the one that provides the best resolution ($<$ 1 nm). However, neither 3D information extraction, nor immunolocalization, nor live-cell imaging are possible [143]. Here, I will consider only the optical microscopy, as its spatial resolution is sufficient to localize biomolecules.

The simplest type of the optical microscopy is bright field microscopy. Its principle is illustrated in Figure 2.1.1. A whole specimen (object **O**) is uniformly illuminated by light which has been focused by condenser. The first inverted image of the object is created after the

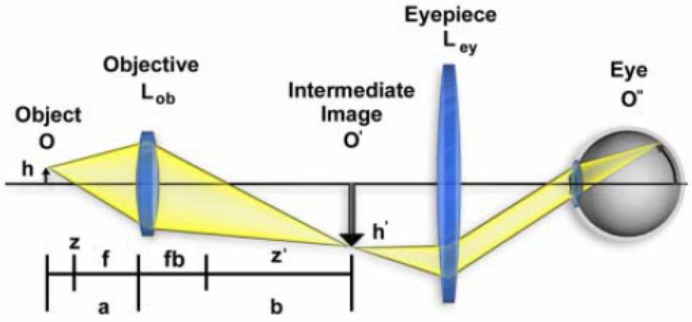
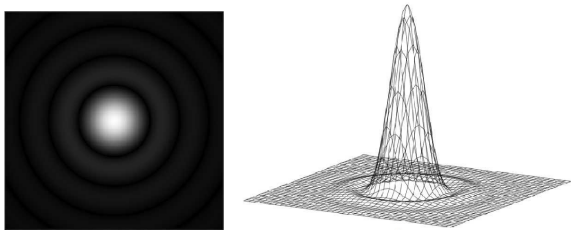


Figure 2.1.1.: Optical trains of a microscope system (from Davidson et al. [31]).



(a) (b)

Figure 2.1.2.: Airy disc (a) and point spread function (b) (from Schmolze et al. [143]).

magnification by an objective lens ($\mathbf{L}_{ob}$). This image ($\mathbf{O}^{'}$) is projected into the intermediate image plane of the microscope [31]. Further magnification of the image $\mathbf{O}^{'}$ is performed by the microscope ocular ($\mathbf{L}_{ey}$) yielding a final image of the object ($\mathbf{O}^{''}$). This image $\mathbf{O}^{''}$ can be projected either on the retina of the observer's eye or recorded by a charge-coupled device camera. The recorded value at each pixel in the obtained digital image is a digitized measurement of photon flux at a specific point. It corresponds to the voltage generated by electrons liberated by photons interacting with the detector surface [159]. The described bright-field microscopy is typically used on thinly sectioned materials due to low tissue penetration by light.

Optical resolution is known to be limited by the wave nature of light [143]. Diffraction is caused by the interference of light waves as they go through the optical system. This results in a point light source being visible as a fuzzy circle (Airy disc) after passing through a microscope (Figure 2.1.2, (a)). The point spread function (PSF) describes its corresponding intensity distribution and is shown in Figure 2.1.2, (b).

The resolution is determined by the shape of the PSF. The narrower the PSF, the smaller is the distance between the closest points that can still be distinguished as two separate objects (see Figure 2.1.3). This distance was described by Ernst Abbe [112, 124]:

$$x = \frac{\lambda}{2n \cdot sin\alpha}, \tag{2.1}$$

where λ is the wavelength of light, n is the refractive index of the medium and α is a half of the angular aperture of the lens. The numerical aperture (NA) of an optical system is defined by the denominator of the Equation 2.1. The practical limit of the resolution is approx. 200 nm in the X-Y plane, and approx. 500 nm in the Z-plane as it is described by a different formula due to its inherent shape [50]. However, recent advances in super-resolution microscopy have changed the limits of optical resolution to approx. 10 nm, allowing visualization of individual molecules as they dynamically interact [54].

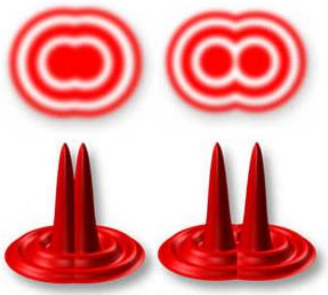


Figure 2.1.3.: Airy discs at the limit of the resolution and those which can not be resolved and appear as one spot (from Davidson et al. [31]).

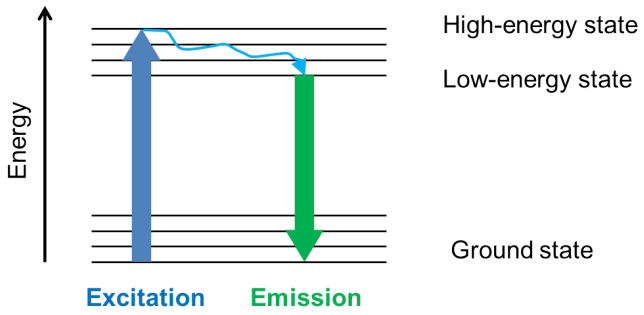


Figure 2.1.4.: Simplified Jablonski diagram.

Fluorescence microscopy

There exist numerous specialized types of light microscopy. For example, fluorescence microscopy provides contrast enhancement by labeling of biomolecules. Special substances are used which absorb and emit light. Such fluorophores enter an excited state upon photon absorption and, after a certain lifetime, can return to the ground state by emitting lower-energy light that is observed as fluorescence [25]. Figure 2.1.4 illustrates a simplified version of the Jablonski diagram describing the transition of electrons during absorption and emission of light [60].

Fluorescence microscope has to separate the excitation light irradiating the specimen from the much weaker re-radiating fluorescent light [31]. This principle is illustrated in Figure 2.1.5. A dichroic mirror is a key element of this scheme, as it is capable of reflecting light of certain wavelengths. Due to this feature the emitted light can be separated from the higher energy (and shorter wavelength) illuminating light.

There exist two general types of fluorescence: primary and secondary. Primary fluorescence, or autofluorescence, is observed when a biomolecule itself is a fluorophore. Secondary fluorescence is achieved by labeling of initially not fluorescing biomolecules. As most of the molecules are non-fluorophores, development of labeling techniques has been of a great importance. The use of fluorophores has made it possible "[...] to identify cells and sub-microscopic cellular components with a high degree of specificity amidst non-fluorescing material" (Davidson and Abramowitz 2002, p. 35). Furthermore, the presence of fluorescing molecules is detectable even if they are below the limits of the spatial resolution.

Although some fluorophores can attach themselves to organic matter, most of the specific labeling is done based on antibodies. This technique is called immunofluorescence and uses the specificity of antigen-antibody reaction to target fluorescent dyes to specific biomolecules [30]. Two general types of immunofluorescence are the direct and the indirect one. Primary, or direct,

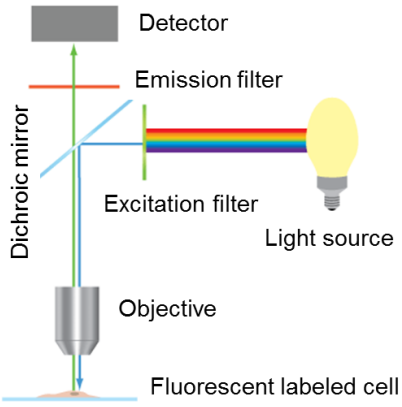


Figure 2.1.5.: Setup of a fluorescence microscope (from BIOImaging [10]).

immunofluorescence is referred to when a fluorophore is attached to one (primary) antibody. This antibody should be specific against the biomolecule that has to be labeled. Secondary, or indirect, immunofluorescence is applied when due to some reasons a direct labeling is impossible. Therefore, the primary antibody against the target biomolecule is recognized by and attached to the secondary antibody which is already labeled with a fluorophore [108]. The advantage of this method is that greater variety of fluorophores can be used for the target biomolecule labeling. However, additional preparation steps make this approach more sensitive and error-prone. Immunofluorescence is now widely used for labeling of biomolecules. "Fluorescence microscopy has become one of the most powerful tools for elucidating the cellular functions of proteins and other molecules" (Dunn et al. 2011, p. C723).

Labeling of biomolecules with different fluorophores enables sophisticated analyses, such as colocalization, for instance. There, a pixel-precise overlay of different color channels is required to deliver correct results. Multi-color fluorescence microscopes have thus to compensate shifts between the image channels that are caused by chromatic aberration.

Chromatic aberrations arise because lenses have different refractive index for different wavelengths of light. Consequently, all wavelengths can not be focused to the same convergence point. Chromatic aberration can lead to the different wavelength focused at different points of the optical axis (longitudinal) or at different positions in the focal plane (lateral). An apochromatic lens can be used for correction. Still, even the best optical microscopes exhibit chromatic aberrations to some extent [91]. Hence, for a precise analysis, an error of a particular imaging system has to be measured and mathematically corrected in 3D.

Confocal fluorescence microscopy

In wide-field fluorescence microscopy, regardless of where the microscope is focused vertically in a specimen, illumination causes the entire specimen thickness to fluoresce. In a given two-dimensional (2D) image more than 90% of fluorescence is out-of-focus light, reducing optical resolution and contrast. In comparison, "[...] confocal microscopy, a form of optical sectioning microscopy, acquires images of thin slices of a thick specimen by removing the contribution of out-of-focus light in each image plane" (Conchello and Lichtman 2005, p. 920). Therefore, a point illumination and a point detection is used producing an "image" of one point of the specimen [175]. In order to obtain an image of a finite region of the specimen it has to be scanned. A laser scanning confocal microscope enables both the point illumination and the scanning functionality [124].

A laser scanning confocal microscope focuses exciting light directed by the scanning mirror onto a small area of the specimen. The illuminated area emits light, which is then separated

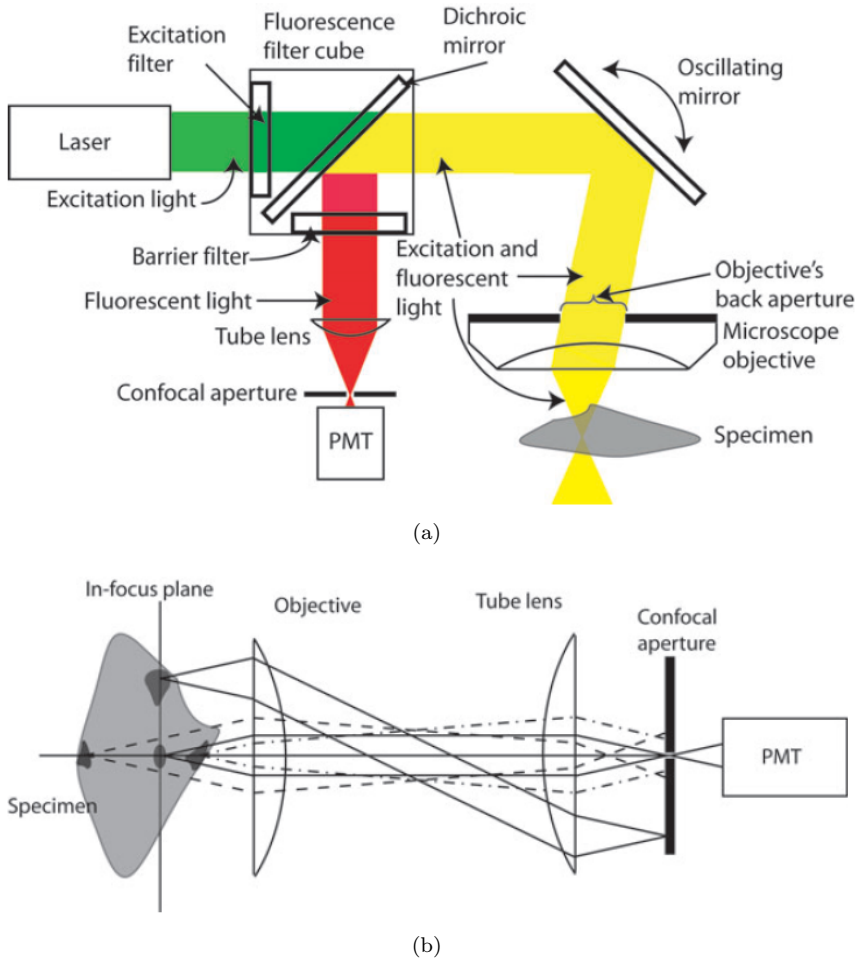


Figure 2.1.6.: The confocal principle. (a) Setup of a confocal microscope; (b) optical trains (from Conchello et al. [27]).

from the excitation light by the dichroic mirror and the barrier filter. The emitted light that passes through the pinhole of the detector is then recorded for the particular illuminated spot (see Figure 2.1.6, (a)). Figure 2.1.6, (b) illustrates the confocality principle. Solid lines show the light emitted from the focal plane. This light is then focused on the conjugated image plane where the confocal pinhole is located. In contrast, the light emitted from the points deeper or less deep than the focal plane in the specimen (the so called out-of-focus light) is focused elsewhere (behind or before the confocal pinhole). Consequently, this light does not pass through the detector pinhole and is filtered out. Scanning the focused spot across the specimen results in an optical section of the specimen at the focal plane.

Due to the elimination of the out-of-focus light confocal microscopy provides better resolution and contrast than the one achievable by a wide-field method. By changing the focal plane, optical sections of the specimen can be imaged at different depths, and a 3D representation can be reconstructed [143, 175]. Such z-stacks (series of images acquired at the same X, Y coordinates while varied Z position) are then processed, rendered and evaluated as a 3D dataset. Figure 2.1.7 illustrates an example of confocal optical sections and the resulting 3D reconstruction.

Confocal microscopy was first described in the 1950s by Minsky and is now a mature technology with many commercial implementations. Unfortunately, a relatively powerful illumination

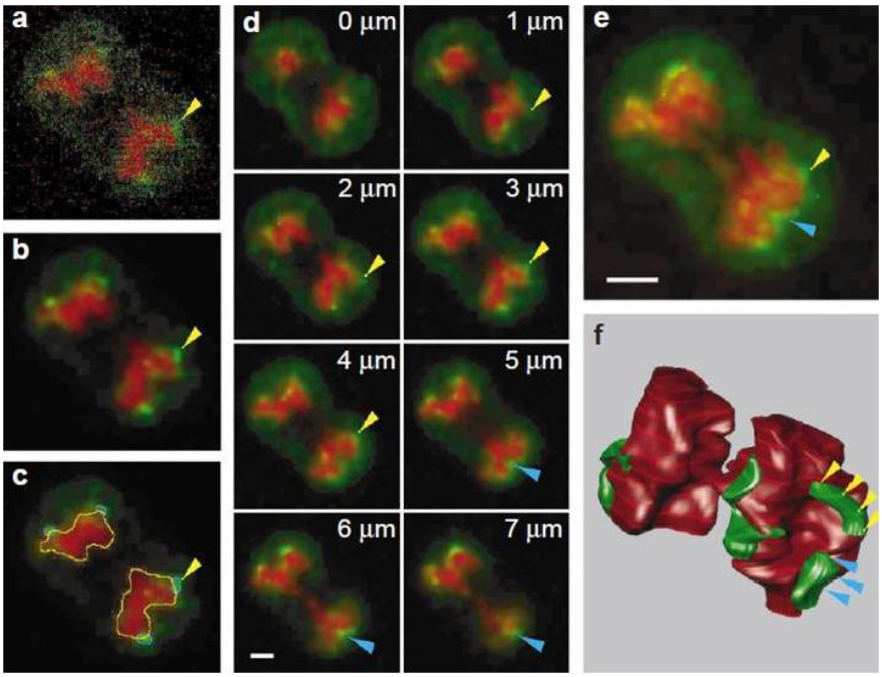


Figure 2.1.7.: 2D and 3D confocal imaging. (a) - (d) Individual optical layers with or without processing; (e) MIP of the z-stack; (f) 3D reconstruction of the z-stack (from Gerlich et al. [58]).

source is usually needed in all confocal designs, since the pinhole passes only a small amount of light, and this can be damaging to live cells [143]. Another problem is an irreversible change in the molecular structure of fluorophores, leading to a loss of fluorescence (photobleaching) [60]. However, confocal microscopy provides unique means for biological analysis and is oftentimes the only applicable technique.

Choosing among various microscopy techniques

A digital microscopy image can be considered a 2D data array. 2D bright-field images differ from confocal images in their acquisition speed. To image the whole sample by confocal microscope, it has to be scanned line by line. Resulting problems like photobleaching and time- and cost- requirements have to be weighted with the obtained benefits, namely better contrast and resolution. Especially in high-throughput applications the simplest method should be chosen which provides enough data of acceptable quality.

In most cases confocal microscopy is preferred due to its increased optical resolution. A possibility to acquire 3D data (z-stacks) is oftentimes another important argument in favor of the confocal technique. Although all biological processes are three-dimensional by nature, many of them can be studied based on 2D images. However, for numerous biological questions 3D confocal microscopy remains the only suitable analysis method. For example, a clear benefit of confocality was shown when a biological question required 3D imaging and analysis (e.g. neuronal systems [100, 178], vascularity [80, 89], intestine structures [53], etc.).

3D fluorescence imaging requires expensive equipment and is quite slow. Samples are excessively exposed to light due to the need to collect several Z-planes. This leads to problems with photobleaching of fluorophores, phototoxicity, and a slow 3D frame-rate [50]. 4D imaging is even more complicated and problematic, as 3D data has to be collected for each time point. Consequently, 2D data is preferred over 3D and 4D data if it provides sufficient information.

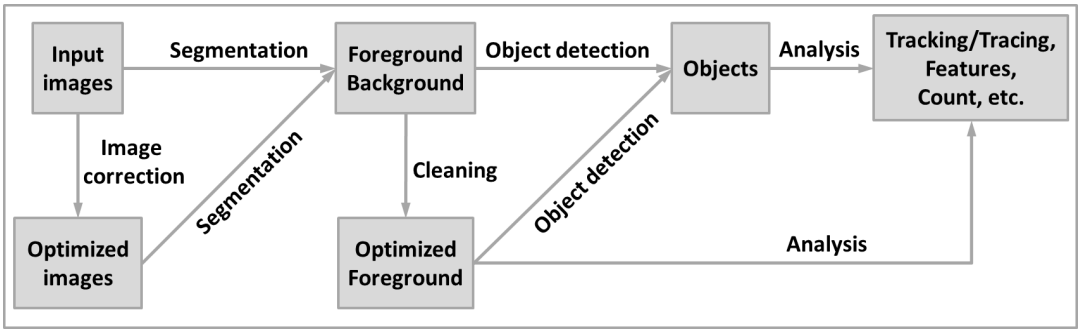


Figure 2.2.1.: Typical workflow for processing and analysis of biological images.

2.2. Image processing and analysis

"With increased imaging throughput and large-scale data acquisition, the challenge of image interpretation and information extraction has also shifted from visual inspection or interactive analysis to more automated methods" (Ruusuvuori et al. 2010, p. 1). Therefore, the development of such methods has become one of the central research topics.

Biological image processing and analysis differ from other application areas due to the natural complexity and variation of biological samples. Low signal to noise ratio, complex object morphology and variety of imaging methods make analysis of biological images especially challenging [41]. Despite countless possible image acquisition and analysis scenarios, there may be a generalized workflow. Figure 2.2.1 illustrates a short typical workflow for processing and analysis of biological image data. In the following, I will present core methods to perform these steps.

2.2.1. Segmentation, or foreground - background detection

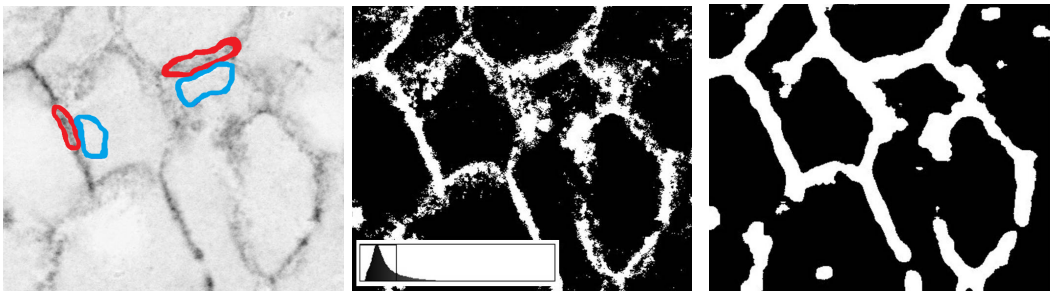
Segmentation of an image can be defined as its partition into different regions, each having certain properties [130, 153]. It is often among the first image processing steps and represents the most challenging part of image analysis [101]. Segmentation can help to find those regions in an image where a certain biological activity of interest is taking place [86]. Segmentation starts by distinguishing foreground from background. Objects can be detected based on their intensity boundaries or other, more sophisticated measurements [159].

Thresholding

Foreground pixels are those which belong to the imaged biological objects. They are therefore expected to exhibit a higher intensity than background pixels. Hence, thresholding of an image may yield the foreground and background regions (brighter and duller than the threshold, respectively). Thresholding is one of the most important approaches to image segmentation [64].

Despite the simplicity of this principle, its implementation is often quite complicated. The choice of the threshold strongly influences the segmentation results. It may be chosen manually following visual inspection. In this case, it will be subjective and will rely on the users' empirical knowledge about the input images. Otherwise, it may be calculated automatically from the intensity values. For example, Otsu developed a thresholding algorithm based on the shape of a histogram [121] (see Figure 2.2.2 (a), (b)). His algorithm chooses the threshold that minimizes the weighted sum of the intensity variance within each of the pixel classes (foreground and background) [101].

Because of variations in staining and illumination, choosing a single threshold for all locations in all images is not always effective [101]. Therefore, separate thresholds may be defined for



(a) (b) (c)

Figure 2.2.2.: Initial confocal fluorescence microscopy image (a) segmented by Otsu thresholding (b) and machine learning (c). Training examples in the initial image: red - foreground, blue - background.

each image channel in each image. A global threshold may be set for each image, or local thresholding may be performed. Local, or adaptive, thresholding algorithms calculate different thresholds for different regions in each image. Intensity values of neighboring pixels are taken into account. Such methods are advantageous if background intensity is uneven or biological samples were prepared differently.

Machine learning

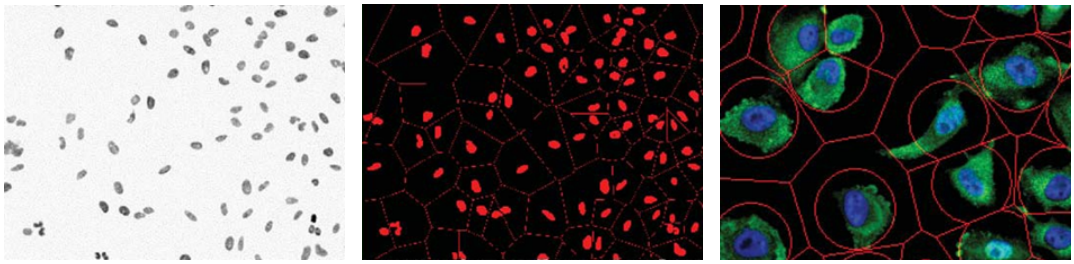
Another type of prominent segmentation algorithms is based on machine learning. Over the past 10 years, well-established machine learning tools have been successfully applied to analyze the microscope images [159]. They showed that various biological questions from single cell analysis to high-throughput screening can be solved [179]. This approach is implemented in the image processing software Zeta [177] used in my work.

The Zeta software performs foreground detection using supervised machine learning (e.g. linear discriminant analysis [176]). A user has to manually label several training examples of the representative foreground and background regions (see Figure 2.2.2 (a)). A set of texture features is calculated for the pixels belonging to these areas [67, 152]. Therefore, a square window is centered at each of the pixels from the example regions. The window size depends on the size of the analyzed structures. Features like mean intensity, its standard deviation and others are computed. Finally, a classification model discriminating between the fore- and the background is automatically learned [140, 147]. This model is then applied to all the rest pixels of the image leading to the foreground detection (see Figure 2.2.2 (c)). High intensity fragments with high contrast to surrounding pixels are discriminated from low contrast background [64]. If a dataset is homogeneous and all images have a similar intensity range, the foreground detection can be trained on one image and applied to the whole dataset. For inhomogeneous data, training examples from several images have to be used.

Segmentation of biological images represents a particularly complex task. First, the natural variety of biological structures complicates their detection based on some common physical properties. Even similar structures may be damaged in the course of sample preparation. For example, membranes may be distorted while freezing, cutting or fixing of a biological probe. Due to these and other reasons, their appearance in bioimages may significantly differ. Consequently, segmentation methods based on thresholding or shape detection may easily fail. Machine learning algorithms, in turn, build a model according to any input data given.

Image features

Machine learning algorithms are based on calculated image features. These are numerical descriptors computed directly from an image to represent its important aspects [149]. Fea-



(a) (b) (c)

Figure 2.2.3.: Nuclei and cytoplasm segmentation. From the initial image (a) nuclei are segmented (b). Corresponding cytoplasm masks are computed for each nucleus (c) (from Morelock et al. [109]).

tures may be calculated from one or more channels, require initial image segmentation or not. Field-level features can be computed when the patterns in different cells within the field are fairly homogeneous. Example features include Haralick texture features [67], Zernike moment features, morphological features, object-based features, wavelet and frequency transform coefficients, threshold adjacency statistics, features from multiresolution subspaces, and others [149]. Morphological and Haralick texture features often lead to good classification results.

However, not all features are useful for every classification. An excess of included features may inhibit performance due to their redundancy. Hence, feature subset selection has proven to be beneficial. It selects the most informative and the most relevant to the current discrimination features [107]. Alternatively, new sets of features may be created as combinations of the original ones. The initial features are projected to a lower dimensional space and its bases are computed by solving an optimization problem [149]. For example, in linear discriminant analysis features are weighted to yield a lower dimensional feature vector [176].

Seeded approaches

Numerous segmentation approaches proceed in two steps. First, so called seeds are identified. They will be used as references to create objects. Further, foreground regions (or objects) are grown around these seeds. Human operators used to define the seeds manually, but this method is not applicable for numerous images generated in a screening study [149]. Hence, automated methods for seeds segmentation are preferred.

In biological images, cells may be identified based on previously detected nuclei. For example, Morelock et al. [109] used Voronoi segmentation [81] for this purpose. First, stained nuclei were automatically segmented using histogram-based thresholding [76]. Masks of the corresponding cytoplasms were computed as intersections of circles centered at the detected nuclei and Voronoi tesselation polygons (see Figure 2.2.3). This method is frequently applied to quantify nucleus-to-cytoplasm ratio of biomolecules (see Section 3.3).

Seeded watershed segmentation [130] is also widely used for object detection in biological images, especially for cells *in vitro* [127]. Seeded watershed segmentation works with the intensity image as a landscape. The watershed by flooding proceeds by distributing water from regional minima across the relief. Barriers are calculated from intensity values. Seeds, for instance, may represent holes. Water will thus create bassins centered at the seeds. When water catchments from neighboring bassins meet, a watershed is built. A set of such watersheds represents the segmentation of an image [127].

(a) (b) (c) (d)

Figure 2.2.4.: Image processing filters. (a) Initial image (from Weltchronik [172]) is processed using median filter with radius 3.0 (b), median filter with radius 5.0 (c), or Sobel filter for edge detection (d).

2.2.2. Filtering

Filtering is widely used for image processing. Filters can be generally classified into low-pass, high-pass, band-pass and band-reject categories. Low-pass filters preserve low-frequency signals and make an image smoother. Gaussian filter and median filter are among the most widely applied smoothing techniques. In the Gaussian filter, a matrix of weight coefficients sampled from the Gaussian function is applied to an image. A median filter is a nonlinear signal processing technique developed by Tukey [130, 168] for efficient suppression of "salt and pepper" noise. Median filter on 2D data proceeds as follows. First, an odd length of a square window L is defined. Second, this window is centered at the current pixel. A median value of the covered pixels is computed. And finally, the value of the current pixel is substituted by the computed median. The larger L, the more is the image smoothed (see Figure 2.2.4 (a) - (c)).

Median filtering can also be applied to 1D data. Therefore, a sliding window with an odd L is centered at every pixel. The current pixel value is then replaced by a median computed from the pixels covered by the window [153].

High-pass filters sharpen an image by preserving high-frequency signals. For instance, a Sobel filter detects edges as high-frequency intensity variations. Figure 2.2.4, (d) illustrates a result of a Sobel operator for the initial image in Figure 2.2.4, (a).

Numerous other filters exist for different needs of image processing. These filters represent only a part of a great variety: convolution, deconvolution, blurring, deblurring, contrast enhancement [64], etc.

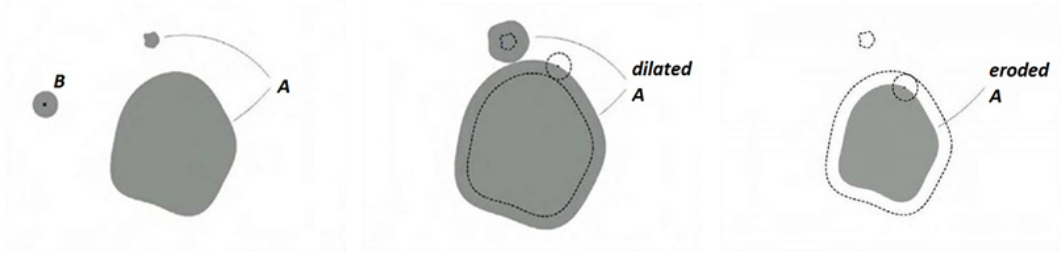
2.2.3. Noise removal

"Due to limitations in imaging technology, an accurate representation of the biological sample can be degraded by several error sources, resulting in a noisy observation of the underlying object" (Ruusuvuori et al. 2010, p. 3). There exist numerous image correction and noise removal options to improve the segmentation results.

Morphological operations

Morphological filters are frequently applied to extract some objects in images while removing the other ones. Image erosion and dilation are the basic operators which can be combined to perform most of the morphological operations [64].

Both erosion and dilation require a so called *structuring element* (SE) to process an image. The SE is a set of a known shape with which the image is probed [153]. SE must have the same dimensionality as the image under study. For 2D images, SE will also be two-dimensional. The most frequent shapes of 2D SEs are square, diamond and disk.



(a) (b) (c)

Figure 2.2.5.: Dilation (b) and erosion (c) results for the initial set of objects (a) using a disc structuring element (from What-when-how [173]).

(a) (b) (c)

Figure 2.2.6.: Morphological opening (b) and morphological closing (c) results for the initial set of objects (a). White - foreground, black - background.

Dilation of an image results in a set of locus points where the structuring element hits the set [153]. Figure 2.2.5 illustrates an initial set of objects (A) that is probed with a disk SE (B). The dilated set A includes the locus points where B touches A.

The dual operator of dilation is erosion. Erosion of an image results in a set of locus points where the structuring element fits the set [153]. These are the points which are included in A when the locus of the SE is placed at one of the points belonging to A. Figure 2.2.5, (c) shows the eroded set of points which were hit by the SE.

Erosion deletes the smallest connected components which never fit the SE (see Figure 2.2.5, (c)). Hence, erosion may be applied for the removal of small objects belonging to noise. However, erosion does not only delete small objects, it also shrinks all the other objects. Once an image has been eroded, there exists in general no inverse transformation to get the original image back [153]. This effect may be partly compensated by combination of dilation and erosion operations. In particular, morphological opening recovers most structures lost by the erosion. The eroded image is dilated using the same SE. This does not recover the smallest removed objects, but compensates for the shrinkage of the other ones. The SE rounds the set from the inside: small structures extending into the background are removed [153], and the contour of an image is smoothed [64]. Figure 2.2.6 illustrates a result of this operation.

Morphological opening has a filtering effect by removing all small structures that can not contain the SE. Its result depends on the size and the shape of the SE. Morphological closing is the dual operator of the morphological opening [39]. It aims to recover the initial shape of the dilated structures. Hence, it is performed by an erosion of a dilated image. Morphological closing rounds the set from outside [153], closing small holes in foreground objects (see an example is in Figure 2.2.6, (c)). Narrow breaks are also fused, and gaps in the contour are filled [64].

Thresholding

Morphological operators are frequently used for image filtering. Images may also be cleaned by thresholding. Thresholding of intensity values has already been described for segmentation (see Section 2.2.1). Object size can also be thresholded to delete those whose area is too small.

2.2.4. Object detection

Although thresholding enables distinguishing between background and objects, one more step is required to delineate each structure [13]. There exist numerous methods to identify objects in the foreground regions. Here, I will only shortly present two of them.

Individual objects may be found by a sudden variation of intensity values when linearly scanning an image [13]. The first derivative of such a line scan will be equal to zero inside an object or a background structure. When passing between objects and background, this derivative will be different from zero. This approach is implemented in edge-detection algorithms, such as Sobel and Laplacian filters [64, 134, 156]. If the intensity distribution is non-uniform, these filters may encounter problems.

Other methods explore the topological relationship of adjacent pixels (e.g. connexity analysis [128]). There, a processing starts with one foreground pixel (reference pixel). A number is assigned to it and its neighborhood is expected (8 pixels in 2D or 26 voxels in 3D). All adjacent foreground pixels are then assigned the same number as the reference pixel. This process is repeated with other foreground pixels, until all of them get a number assigned. Herewith, objects or connected islands of pixels are tagged with the same number. This method is applicable to any size and shape of the target structures and requires no *a-priori* knowledge of those parameters [13].

2.2.5. Skeletonization

Once the foreground is segmented, the identified objects can be analyzed. The shape of an object represents one of the important features that can be extracted. The region-based shape representing the general form of an object can be obtained by skeletonization. This process yields a skeleton, or a medial axis representing the condensed information on the objects shape while preserving its homotopy [153]. The purpose of skeletonization for the 2D case is to reduce 2D discrete objects to 1D linear representations preserving topological and geometrical information [7].

Skeletonization plays a central role in a broad range of problems in image processing [64]. For instance, it is applied for processing of scanned documents. There, to identify the characters, initial images have to be skeletonized to yield 1 pixel thick lines. Further, such lines can be vectorized and classified. In biological image processing, skeletonized images are used for object tracing (e.g. vasculature, neurites, etc. [149]). In the context of my work, skeletons will indicate directions of the membranes labeled.

The generated skeletons are expected to have several desirable properties. First, the skeleton must preserve the topology of the original image. In other words, the Euler number should be the same before and after skeletonization. Second, the skeleton should be made of one-pixel-thick objects, i.e. be as thin as possible. Third, the skeleton should lie in the middle of the shape. Fourth, skeletonization should be rotation invariant. Fifth, it should be possible to reconstruct the original image from the skeleton [7].

Various skeletonization techniques exist, differing in their complexity and computational cost. They can be classified into four types [8]. The first are thinning algorithms (e.g. wave front/grassfire transform and thinning). They iteratively remove border points, or move to the inner parts of an object in determining an object's skeleton. The second type is represented by algorithms based on the Voronoi diagram. Such methods search the locus of centers of the maximal disks contained in the polygons with vertices sampled from the boundary. The third

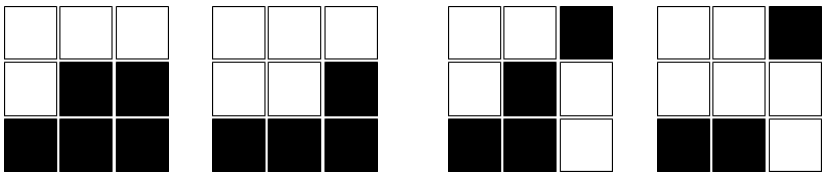


Figure 2.2.7.: Euler characteristics for topology preservation. (a) $v = 11$, $e = 15$, $f = 5$, $E = 1$; (b) $v = 10$, $e = 13$, $f = 4$, $E = 1$; (c) $v = 11$, $e = 14$, $f = 4$, $E = 1$; (d) $v = 10$, $e = 11$, $f = 3$, $E = 2$ (from Nischwitz et al. [114]).

type of algorithms is based on a distance map. And finally, the fourth type of methods uses mathematical morphology.

Different skeletonization techniques lead to different skeletons [128]. Some are more precise, but computationally expensive, others are less complicated. I will present here one of the simplest algorithms, namely, skeletonization according to the Euler characteristics [114]. It is based on the calculation of the Euler number (E) which reveals the connectivity and topology of the objects:

$$E = v - e + f, \tag{2.2}$$

where v is the number of vertices, e is the number of edges and f is the number of faces. For 2D objects $E = 1$. Two objects are considered topologically equivalent if their E is the same. Hence, a foreground pixel can be deleted (turned into background) if this operation does not change E. To test this, E is calculated for 3×3 neighborhood of the current pixel.

Figure 2.2.7 illustrates two examples for this operation. A binary image in Figure 2.2.7, (a) has $E = 1$. Deletion of the central pixel leads to the image in Figure 2.2.7, (b) which has $E = 1$. Here, E has not changed and this central pixel can be removed. Figure 2.2.7, (c) has $E = 1$. Deletion of the central pixel leads to the image in Figure 2.2.7, (d) which has $E = 2$. In this case E has changed and this central pixel can not be removed.

The skeletonization algorithm proceeds by sequentially scanning the image and examination of each pixel. Those pixels, whose deletion does not change the E, can be removed. Two remarks have to be made on this algorithm. First, line ends can be treated separately. If a full line length is to be preserved, then pixels having only one foreground neighbor will not be removed. Second, the order of pixel processing influences the skeletonization result. Therefore, the algorithm proceeds in multiple sub-iterations. Only a subset of pixels is considered in each sub-iteration. For example, pixels having both odd or both even coordinates can be processed first [114].

2.2.6. Pruning

The generated skeletons frequently contain inaccurate or redundant branches. The skeleton generating approaches suffer from the fact that a small protrusion on the boundary may result in a large skeleton branch. "This is an intrinsic problem of the definition of the skeleton, since the mapping of boundary points to the skeleton points is not continuous" (Bai et al. 2007, p. 451). Pruning of a skeleton can be performed to delete such branches.

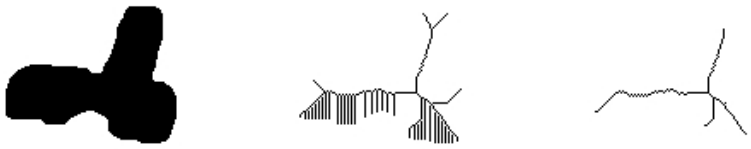
Pruning methods can be classified into two categories. First, object boundaries can be smoothed before the skeletonization. Second, skeleton points can be selectively removed, for instance, based on the assigned significance measures. Propagation velocity, maximal thickness, radius function, and axis arc length can be used as such measures [148].

Alternatively, pruning based on morphological operations can be used, for instance hit-and-miss transform [64] with two following SEs (see Figure 2.2.8). The obtained skeleton is cleaned

0	0	0
0	1	0
0		

0	0	0
0	1	0
		0

Figure 2.2.8.: Structuring elements that may be used for pruning. 1 - Foreground; 0 - background; blank - any pixel.



(a) (b) (c)

Figure 2.2.9.: (a) Initial foreground object; (b) its skeleton; (c) pruned skeleton.

by deleting short branches that are attached to the main skeleton line. Pixels of the shorter branches are deleted one by one from four directions until no further deletion is possible. Pixels of the skeleton are iteratively scanned in four directions (top-to-bottom, left-to-right, bottom-to-top, right-to-left) to detect and delete end points. Finally, only those parts of the skeleton are left that extend in the direction of the long axis of the segments (see Figure 2.2.9).

2.3. Software tools for image analysis

Bioimaging has become an integral part of any biological research. The more data are acquired and the more biological questions are investigated, the more different analysis techniques are required. There exist a great variety of image processing and analysis toolboxes, both open-source and commercial ones. They differ in available functionalities, usability, the main application focus, etc. They offer manual, assisted or completely automated methods for image processing and analysis. Here, I will briefly highlight features of some prominent examples.

Numerous programs are being developed to solve a particular biological problem. Such niche tools are very specific and frequently not applicable to different biological samples. Other generalist tools are more flexible, but may still be focused on one or another image analysis aspect. For instance, Amira [6], Imaris [78] and ImagePro Plus [77] are commercial tools which are often sold together with microscopy hardware. They offer a wide range of processing and visualization functions. However, being commercial and proprietary, they are usually black-box algorithms for the end users. Hence, they can neither be optimized nor extended by the community.

Open-source tools are often developed in a framework of some research project and so may also be quite specialized. Nevertheless, they can be extended by the users and thus are constantly growing. BioImageXD [82, 83], ImageJ [1, 144] and CellProfiler [23, 98] are among the widely used tools.

ImageJ has been used for the longest time and its functionalities have been greatly extended. Owing to its rich history and pioneering status, ImageJ can perform a wide variety of common (and many specialized) image-processing and image-analysis tasks, particularly in the life sciences [46]. Still, ImageJ is focused on 2D and 3D data which complicates multidimensional analysis. In turn, BioImageXD is focusing on processing of multidimensional data. It is also known for its 2D and 3D visualization techniques. Furthermore, ImageJ was developed for

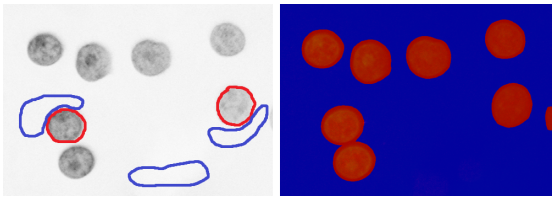


Figure 2.3.1.: Structure detection using the Zeta software. Initial microscopy image (left) with training examples for the fore- and background labeled by an expert (red and blue, respectively); detected fore- and background regions (right).

the analysis of individual images. Processing of high-throughput data is possible only with additional programming.

CellProfiler is a flexible software designed for high-throughput image analysis [23]. It is used to address several application areas, including intensity and morphology measurements, phenotype scoring by machine learning and object tracking [46]. CellProfiler is organized in a modular basis, so that workflows can be created by a user. Numerous image processing modules are available and may be combined for a specific research project. CellProfiler enables illumination correction and segmentation. The identified objects can further be quantified and classified. A large number of features is calculated for each identified cell or subcellular compartment, including area, shape, intensity, and texture [23].

2.3.1. Zeta: machine learning for image processing

A powerful software for trainable image analysis (Zeta [177]) was developed at the Fraunhofer Institute for Applied Information Technology (FIT). This software can analyze biological images (e.g. microscope images) and extract rich information on the objects depicted. An important feature of the Zeta software is application of machine learning algorithms in numerous image processing functions.

Image processing with Zeta starts with the foreground-background detection. This function is based on supervised machine learning algorithms (e.g. linear discriminant analysis [176]). A user has to manually label several training examples of the representative foreground and background regions (see Figure 2.3.1). A set of texture features is calculated for the pixels belonging to these areas [67, 152]. A classification model is learned and applied to the rest of the pixels. A model can be learned on several images and applied to a whole dataset for batch processing.

Further, Zeta can detect and count individual objects based on the foreground detection results. Moreover, it can classify the detected objects into any number of the given classes (e.g. live and dead, bright and dull, etc.), depending on the training examples given. A great flexibility of this software and its applicability to a wide spectrum of biological images is based on the underlying algorithms. For example, segmentation represents a complicated step for the image processing due to the natural biological variation. Zeta solves this problem by avoiding hard coded thresholds. Instead, a discriminating model is always learned from the actual input data. Herewith, this approach is more flexible. Numerous further Zeta functions can be applied for biological image analysis. However, in the scope of my work, only its foreground detecting functionality is used.

Available image processing and analysis toolboxes offer a wide range of functionalities. Nevertheless, quantification of molecule translocation has not yet been extensively addressed. Only nucleus-to-cytoplasm translocation can be evaluated (e.g. by CellProfiler [22]), while no standard approaches have been reported to quantify other translocation types. This is why I aim at the development of a generally applicable translocation quantification method.

2.4. High-content screening

"Imaging plays a unique role in that it can both provide insight during experiments and also be used to gather data in a high throughput fashion for later analysis" (Kherlopian et al. 2008, p. 74). Indeed, imaging techniques having been invented in research purposes have become an essential part of industrial high content screenings.

Nowadays, research and industrial applications can not be strictly separated from each other. For example, drug discovery is one of such areas. There, protein networks are often the information source to identify a drug target for a disease treatment. To uncover such networks, biological images can be acquired by automated microscopes. The images can be evaluated using, for instance, colocalization analysis. Once a network was established, high content screenings have to be performed to find a suitable drug to the target molecule. This workflow is only one possibility among numerous alternatives. Still, it reveals a strong connection between industrial and research applications. It also illustrates the necessity of advanced imaging techniques to acquire images in a high-throughput manner.

High-content screening and analysis have significantly influenced the area of drug discovery and systems biology [26]. Acquisition of a great amount of multi-parametric biological data enables sophisticated analyses that would have been impossible without such advanced techniques. High-content screening represents a type of a phenotypic assay [181]. There, cells are treated with various reagents (chemicals, RNAi, etc.), often assisted by automated liquid handling and cell culture [71, 149]. The cells are frequently fluorescently labeled or genetically engineered for further quantitative and qualitative analysis [26]. For instance, the cells can be analyzed by flow cytometry or imaged using automated microscopes.

Flow cytometers work with cells suspended in a liquid that are passed through a focused laser beam. Up to 50,000 cells per second can be analyzed and sorted by modern devices [151]. For each cell, fluorescence intensity and a light scatter are measured. Cell populations can be discriminated based on size, morphology, and other properties [151].

Automated microscopes acquire images of cells or cell cultures. The acquired images may be confocal or not, depending on a particular assay. Either way, image acquisition is only the first step. Image evaluation represents a further challenge. Image processing, computer vision, and machine learning are used to automatically process high-dimensional image data into meaningful cell biological results [149]. Image processing is useful not only in a high-throughput context, but also for biological studies, such as "[...] quantifying the amount and localization of a signaling protein, measuring changes in structures over time, tracking invading cancer cells or looking at nonspatial data such as fluorescence-lifetime data" (Eliceiri et al. 2012, p. 701).

One of the first image processing steps, segmentation, is crucially important for the evaluation of high-content screening data. If individual cells have to be segmented, various cellular components can be labeled. Labeling of the outer cell membranes enables straight-forward cell segmentation. Similar results can be achieved by labeling of some cytoplams components. Alternatively, cell nuclei can be labeled. If cells under study have only one nucleus, then the segmented nuclei could be sufficient for cell counting. If cells may have more than one nucleus or any other measurements have to be performed, individual cells have to be segmented (e.g. as regions around the segmented nuclei). Voronoi segmentation or other seeded algorithms can be applied for this task.

High-content screening enables acquisition of huge amounts of biological data. One of the possible analyses is to study the laws of spatial arrangement of biomolecules. Such toponomics approaches are presented in the next Section.

2.5. Toponomics

Biologists are increasingly interested in obtaining quantitative data by analysis of biological images [46, 61, 101]. In particular, microscopy images are most frequently analyzed, as microscopy

is claimed to be the most popular imaging technique.

Information on location of biomolecules can be extracted by image analysis. *Toponomics* is one of the relatively young research fields focusing on location of biomolecules. It describes the laws of their spatial arrangement. In the following, I will present some established toponomics analysis methods.

2.5.1. Colocalization analysis

Cellular localization and function of biomolecules are strongly connected and sometimes can be inferred from each other. It is similar with interaction between biomolecules which are colocalized. If two biomolecules (e.g. proteins) are frequently found near each other (this frequency is statistically significant), then their direct or indirect interaction can be supposed. The contrary, however, is not always true. Even if two proteins are found far away from each other, their interaction can not be excluded. There are various reasons for that. First, the interaction can be so fast that the proteins are detected already after it when they have been transferred away from each other. Or, second, a mediating factor can be present, so that the close proximity is not required for such an indirect interaction.

"[...] Direct molecular interactions crucially depend on spatial proximity [...]", or colocalization (Helmuth et al. 2010, p. 1). If structures interact, a correlation of their spatial distributions can be detected. Hence, previously unknown interactions may be discovered by colocalization analysis.

Technology

There exist two general ways to tackle the colocalization question. The first one is based on the intensity correlation. If a correlation between the intensity distributions of two biomolecules is found, then an interaction between them is inferred. To apply this method, two (or more) biomolecules of interest have to be fluorescently labeled with different markers. Microscopy images are then taken from the biological probe and the intensity distributions are analyzed for all the channels.

If two proteins bind to the same cellular compartments, then they are considered colocalized. This case has to be distinguished from a random spatial colocalization of two proteins in a cell. Statistically significant colocalization can then be detected by thorough evaluation of their localization. There exist numerous methods, such as cross-correlation or cluster analysis which are either qualitative or subjective [28]. An early attempt to quantitatively estimate the colocalization was done using the Pearson's coefficient and extended to colocalization coefficients [103]. These coefficients describe a colocalized fraction of each biomolecule, but require a manually set threshold to distinguish a specific staining.

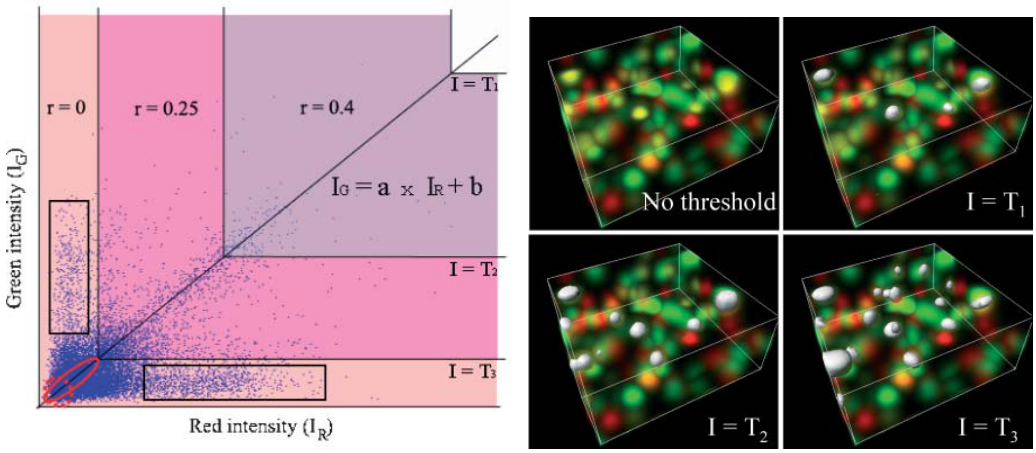
Costes et al. [28] ovecame the problem of a manually set threshold by its statistical calculation. As their whole method is based on intensities, the image quality plays a crucial role. Not only noise levels have to be low, but also the registration of the images from different channels has to be free of errors. The analysis can be performed both on 3D and 2D data, referring then to individual voxel or pixel intensities, respectively. Intensity of each biomolecule (I) is considered to be a composition of a random spatial distribution (R) and a colocalized component (C):

$$I_1 = C + R_1 \tag{2.3}$$

where a protein type 1 is taken as an example. Intensity of a protein type 2 will then also consist of a colocalized and a random component:

$$I_2 = \alpha \times C + R_2, \tag{2.4}$$

where α is a stoichiometry coefficient that takes into account that proteins 1 and 2 do not necessarily colocalize in a 1:1 ratio. Figure 2.5.1 visualizes a colocalization in a two-color image



(a) (b)

Figure 2.5.1.: Colocalization analysis by intensity correlation. (a) Colocalization in a two-color image (e.g., red and green) visualized in a two-dimensional histogram. I_G and I_R correspond to the number of pixels with red and green intensities. (b) 3D image visualizing pixels with intensity values in both channels greater than the threshold T (from Costes et al. [28]).

(e.g., red and green). A two-dimensional plot shows numbers of pixels with respective green and red intensities (I_G, I_R). The line $I_G = \alpha \times I_R + b$ represents the colocalized pixels [28] and can be approximated by least-square fit. The slope of this line is directly proportional to the Pearson correlation coefficient.

Optimal thresholds T_G and T_R are statistically identified for both channels to detect the colocalized areas. These are those pixels whose intensities are greater than the thresholds in both channels. The higher the thresholds, the less colocalizations are found, but also the smaller the number of false positives (see Figure 2.5.1, (b)). The colocalization coefficients as well as the Pearson coefficient are then computed in the colocalized areas.

This method was successfully applied on various biological datasets. Negative controls could be distinguished from the positive ones by the average colocalization of 5 and 95%, respectively. A true colocalization of 3% was also detected even in those images which would be discarded by a simple visual evaluation. The algorithm has even been commercialized by Bitplane [11]. However, intensity-based correlation analysis may be hard to interpret, as interactions have to be inferred from intensity space sensitive to imaging artefacts [72]. Furthermore, resolution of a microscope limits a spatial scale on which the interactions can be detected.

The second approach builds upon object detection methods. Initial multi-channel microscopy images have to be processed to yield discrete objects of every studied biomolecule. To achieve this, segmentation is applied by thresholding of the image in every channel (see Figure 2.5.2, (a) and (b)). The threshold can be manually selected or automatically calculated from the intensity values. Once all the pixels (or voxels for 3D data) have been classified as background noise or foreground regions, individual objects can be detected.

There exist numerous methods to identify objects in the foreground regions. Two widely used methods were presented in Section 2.2.4, namely, edge detection and connexity analysis [13, 128]. Once the objects have been identified, their colocalization can be estimated. Therefore, a concept of a centroid has to be introduced. A centroid of an object is its geometrical center, and its coordinates are obtained by averaging the coordinates of all objects' pixels (see Figure 2.5.2 (c)). Alternatively, intensity centers can be calculated taking into account individual intensities.

Centroids of the objects in the different channels are studied to compute their colocalization. A nearest neighbor for each e.g. green centroid is found among the red centroids. A distance

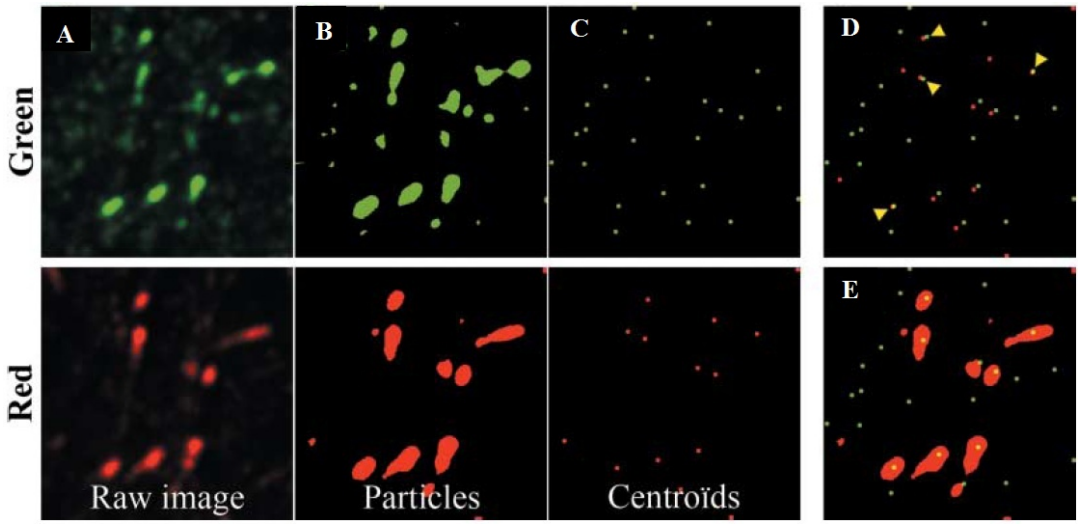


Figure 2.5.2.: Object-based colocalization analysis. (a) Raw image; (b) segmented particles; (c) detected centroids; (d) nearest-neighbor distance approach by merging green and red channel centroids; (e) merged view of centroids of the green image (c) and particles of the red image (b) illustrates the overlap (from Bolte et al. [13]).

between them is calculated and compared to the *a priori* selected threshold that describes the interaction scale [13]. If these centroids are close enough, the respective objects are considered to colocalize. Setting of a too high threshold will lead to numerous false positives, while the contrary will result in missed true interactions. Thus, the choice of the threshold is crucial for meaningful data interpretation. Instead, it can be evaluated whether centroids of one channel are overlapping with objects in the other channel (see Figure 2.5.2, (e)). This colocalization definition is more flexible as it does not punish segmentation errors as much as the first one does.

In the object-based methods potential interactions are inferred from physical space. It has various advantages over intensity space, as intuitive colocalization measures like the number of overlapping objects can be constructed [13]. Moreover, resolution of a microscope does not play a crucial role here, as the size of the smallest detected object depends on the signal-to-noise ratio [138].

Helmuth et al. [72] extended conventional object-based methods by introducing statistical measures. They built upon the non-independence of the relative positions of objects under study. This allowed a better estimation of the initial objects distribution and thus a better estimation of colocalization measures.

2.5.2. Multi-epitope-ligand cartography

"The knowledge of the proteome - the snapshot of the total protein output encoded by a genome - provides important information on: the translated genes, the relative abundance of expressed proteins and the posttranslational modification of these proteins" (Schubert 2003, p. 190). Numerous analysis techniques have been developed for it. However, information on protein localization can not be obtained by standard proteomics methods. To extract such data, intact cells or tissues have to be analyzed instead of their homogenates. Protein localization represents a valuable information source on their organizations in a tissue. Protein networks may be inferred from their locations, giving hints for previously unknown protein-protein interactions and suggesting potential drug targets.

Proteins, as any biomolecules, have to be at the right time at the right place to fulfill their function. In other words, they are topologically determined. There exist several methods to

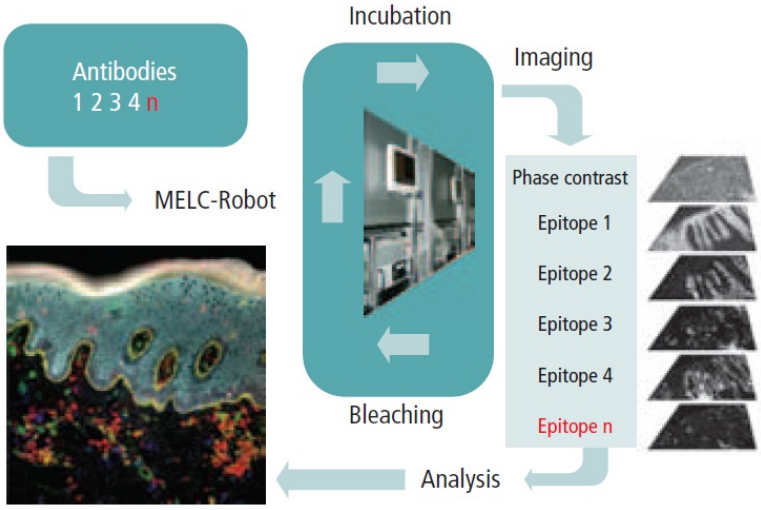


Figure 2.5.3.: Schematic illustration of the MELC process: n-fold cyclic process provides fluorescence images for n epitopes. Their point-precise overlay allows the construction of a colocation map (from Bonnekoh et al. [14]).

study protein topology. For example, closely located proteins may be identified by fluorescence resonance energy transfer (FRET) microscopy [73]. However, available immunohistochemistry and microscopy methods can image only a very limited number of proteins in a single sample. In order to reveal protein networks, a much greater number of proteins needs to be mapped simultaneously with a sufficient resolution. The Multi-Epitope-Ligand-Kartographie (MELK), further referred to as multi-epitope-ligand cartography (MELC), suggests a solution to this problem. Up to 100 proteins on the same biological sample can be localized by sequential rounds of fluorescent detection *in situ* [145, 146].

Technology

MELC technology includes a cycle of three steps [145] illustrated in Figure 2.5.3. I refer to proteins (or epitopes) here, although any other molecules whose localization is of interest can be studied, too.

1. For the i-th protein of interest, the antibody labeled with a fluorescent marker is prepared. A fixed biological probe (a blood smear, cells or a sectioned tissue) is incubated with this antibody for 30 minutes. Further, unbound markers are washed out.

2. The biological probe contains now one fluorescently labeled protein. Fluorescence and contrast phase images are taken and stored.

3. The fluorophore applied in the step (1) is bleached out to study distributions of other proteins on a *cleaned* probe. A postbleaching image is stored and later subtracted from the fluorescence image acquired in the step (2) of the next protein's cycle.

These operations are performed until localization of all molecules of interest was recorded. The maximum number of successfully performed additive rounds has already reached 100 [125].

After image acquisition, protein localization patterns are detected and analyzed. All fluorescence images are aligned based on the corresponding phase contrast images, so that the accuracy of $\pm$ 1 pixel is reached. Flat-field correction is applied to compensate for the illumination artifacts. The obtained images are binarized in order to simplify further processing and enable easier comparison of different experiments. Gray-scale images are thresholded, while

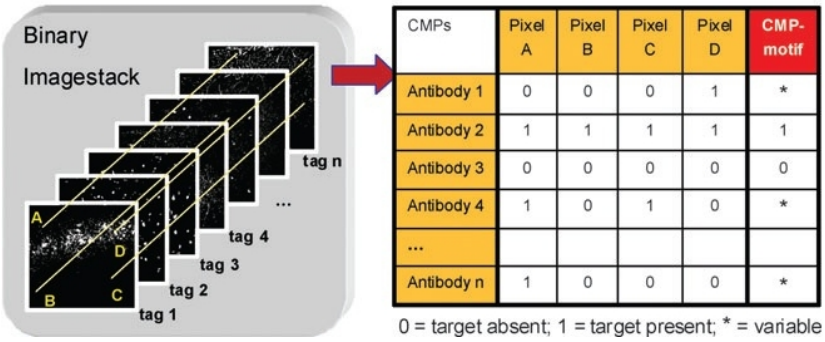


Figure 2.5.4.: Generation of combinatorial molecular phenotypes (CMP) from the fluorescence images (from Pierre et al. [125]).

one and the same threshold is applied for a specific marker in all MELC runs. Herewith, for every pixel and every fluorescent tag in all the images a value of 0 or 1 is assigned, depending on whether sufficient protein intensity was detected on this position. Superimposed binarized images compose a matrix of localization patterns for each pixel which is described as combinatorial molecular phenotype, or CMP [125]. Each primary data set of the fluorescence intensity images produces a collection of 0.1-vectors (CMPs) [146].

If n different tags are used in a MELC run, then a theoretical maximum of 2^n different CMPs can be detected. Only a part of them will be biologically meaningful and can really occur in nature. However, the information extraction and analysis are still challenging due to a huge amount of the generated data.

CMPs are summarized in expanded CMP-motifs which are defined as pixel-related code for the presence or absence of a protein at a current position in all images. Zero is assigned when a protein is absent, one - if it is always present, and a $*$ if it varies among the dataset (see Figure 2.5.4). Some CMP motifs denote functional regions of the cell or tissue in question [146]. Toponome and molecular colocalization maps of cells and tissues can be produced if CMP motifs are color-coded (see Figure 2.5.3). Depending on the dataset, a 2D or a 3D toponome map can be obtained.

Overall frequencies of motifs are studied and can vary significantly between different experiments. Therefore, relative CMP frequencies are preferred for the quantitative comparison [126, 145].

Applications

The MELC technology was the first to enable systemic investigations regarding the topology of protein networks. “The ability to detect and quantify cellular protein networks is an important step toward the understanding how the proteome determines different cellular functions” (Pierre and Scholich 2010, p. 646). For example, color-coded toponome maps are a powerful visualization method which can be applied to distinguish healthy and diseased states. If a protein organization is affected by a disease, then the changes will be detectable by MELC. Schubert et al. [146] analyzed skin biopsies from healthy patients and those affected by psoriasis, confirming MELC potential. Protein distributions were in fact found to be significantly different (see Figure 2.5.5) and disease-specific CMP motifs were identified.

Toponome maps can also be used to identify subcellular protein networks. Herewith, enough information can be extracted to model dynamics of molecular networks *in silico*. Specific protein-protein interactions can be characterized and even quantified by MELC and selective analysis of CMPs. This toponomics approach allows studying even those processes which take place in subpopulations of cells while earlier only massive effects could be detected.

Furthermore, drug effects can now be examined on a great level of detail. Western blots and

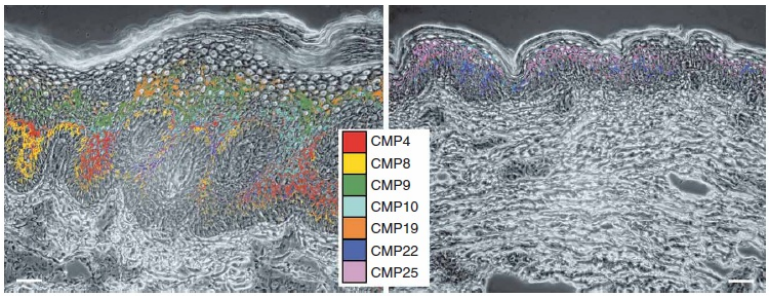


Figure 2.5.5.: Toponome maps of skin biopsies of psoriatic (right) and healthy (left) skin (from Schubert et al. [146]).

immunohistochemistry are frequently used to study drugs actions, missing however those effects which do not result in protein expression changes. Topological proteomics, in contrast, is able to detect these as well as off-target effects resulting in the toponome changes. MELC enabled analysis of protein organization that may help to find target proteins in disease [146]. Toponomics analyses have already proven their value in basic and biomedical research by predicting novel protein-protein interactions and by identifying new drug targets [125].

Limitations and potential for further development

The toponome is a dynamic structure in any cell and varies depending on a time point (life cycle phase), applied treatment or any other factor. Furthermore, toponomes of genetically identical cells are likely to differ, not to mention different tissue samples. Therefore, only thorough statistical evaluations can show which CMP-motifs are representative for a current state of a cell and which CMPs frequencies fluctuate simply due to the biological variability. One study considered this question by evaluating toponomes of three genetically homogeneous animals. Surprisingly, only 45% of common CMP-motifs were identified in these animals [12].

As numerous proteins can be detected by MELC, they have to be tagged by respective antibodies. The order in which these labeled antibodies are applied can have an impact on their binding efficiency. Until now, only one case of a reduced affinity was documented (30% [125]), when two epitopes of one and the same protein had to be localized. In other performed MELC runs the order of applied tags seems not to influence the result significantly [146].

Image processing may also be a source of errors. Thresholds for the image binarization can be either manually selected or automatically generated. Although they could have a strong impact on MELC results, the contrary was shown. MELC was found to produce robust results at various thresholds set either by human experts or generated by a system. Still, binarization process causes problems in images with low signal-to-noise ratio [125].

"A main problem of the conventional proteome analysis is the missing correlation of cellular protein expression with protein localization and function" (Pierre and Scholich 2010, p. 645). Numerous preclinical drug developments fail because of lacking knowledge on protein networks changes caused by disease. Thus, reliable tools are required that could study topology and its changes in signaling pathways [125].

Although MELC provides valuable information on protein toponomics, only qualitative samples comparison has been possible so far. Distinct protein localizations can be detected in biological samples, but respective quantification methods are still lacking.

2.5.3. Densitometry

Toponomy of biomolecule distribution can also be assessed by density analysis, or densitometry. *Densitometry* is widely applied to measure bone loss [24]. To analyze it, bones are scanned by

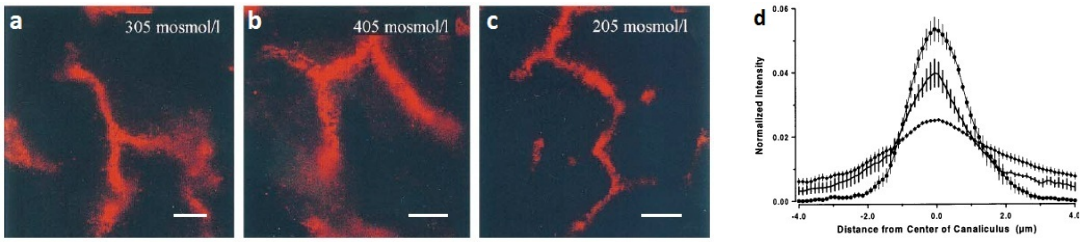


Figure 2.5.6.: (a) - (c) Fluorescence images with a labeled membrane protein (red) under various experimental conditions; (d) corresponding extracted average 1D density vectors (from Kubitz et al. [92]).

X-ray and their density is measured to identify potential osteoporosis (a disease characterized by reduced bone mass, deterioration of bone tissue leading to bone fragility [29]).

Generally, densitometry is based on a measurement of optical density of a sample (e.g. solution of a biomolecule of interest, western blots [16, 19, 55]). The light intensity is measured before and after going through the sample, and the difference yields the absorbance (or the optical density) of the sample. Densitometry can also be performed on 2D images, e.g. fluorescence images. In this case, pixel intensity will provide information on the "amount" of the marker at this particular position. As optical density is dependent on a concentration of biomolecules, the latter one can be determined in a linear measurement range. However, immunofluorescence methods do not allow absolute quantity measurements between images, due to biological variation and unstable binding of labels to the target molecules. Therefore, no absolute measurements should be performed by densitometry on fluorescence images. The extracted density should thus be normalized and only then used for samples comparison.

A density can be measured at a particular point, along a line (1D), or in some limited region (2D). For example, the densitometric analysis can be performed along a concentration gradient to study a directed transport process. This will provide information on the distribution of biomolecules along the translocation axis, which is required to assess the translocation progress. Figure 2.5.6 illustrates a case of such 1D densitometric analysis in fluorescence images. Here, a membrane protein is labeled. Its fluorescence intensity is monitored along a concentration gradient, yielding a 1D density vector. These vectors represent the current distribution of the labeled protein orthogonally to the membrane. Figure 2.5.6, (a) - (c) illustrates images of biological samples prepared with different osmolarity. Distribution of the labeled protein is dependent on the osmolarity, and differs in the given images (Figure 2.5.6, (d)). Hence, respective biological samples may be compared based on the extracted distribution vectors.

2.5.4. Probabilistic density maps

Apart from distribution of biomolecules, cellular morphology can be studied to evaluate changes after a treatment. "Despite the central role of morphology comparisons in cell biological approaches, few statistical tools are available that allow biological scientists without a high level of statistical training to quantify the similarity or difference of fluorescent images containing multifactorial information" (Duong et al. 2012, p. 8382). Schauer et al. developed a method based on probabilistic density maps to solve this problem [139].

In particular, Schauer et al. studied the steady state organization of endomembranes. This includes such structures, as the Golgi apparatus, early endosomes, lysosomes, etc. Endomembranes are linked to cell cytoskeleton, which, in turn, is essential for membranous transport. Hence, spatial organization of endomembranes influencing their function is also crucial for the transport processes. Vice versa, transport events also alter the cytoskeleton structure. Furthermore, dynamic cellular organization changes over time (or its life cycle). All these factors result in constantly varying spatial organization.

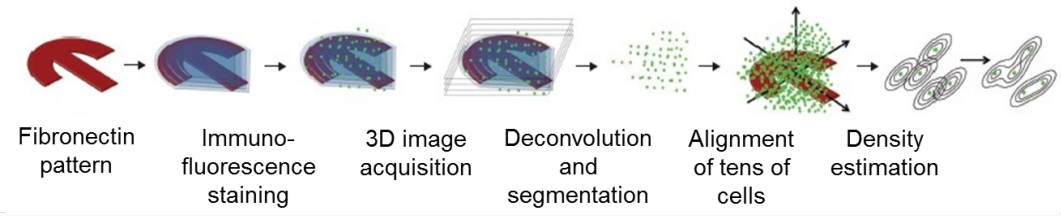


Figure 2.5.7.: Generation of probability density maps. Cells are grown on micropatterns. Marker proteins are labeled and fluorescent images are acquired. Deconvolved images are segmented to yield marker positions. These are aligned and processed to result in probability density maps (from Schauer et al. [139]).

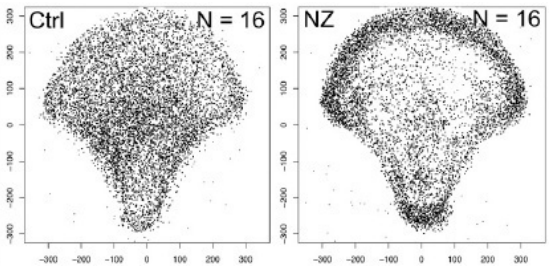


Figure 2.5.8.: Samples comparison based on probability density maps. Representative 2D scatter plots of a control (Ctrl) and treated (NZ) biological sample (from Duong et al. [43]).

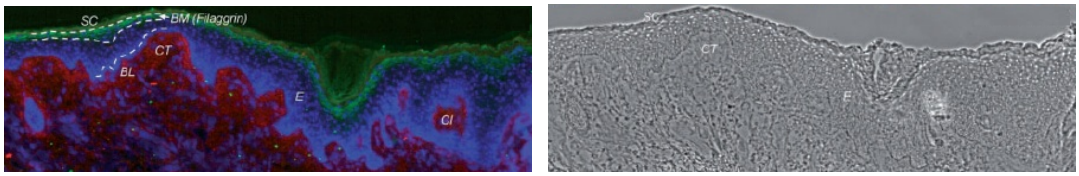
Technology

To study 3D endomembrane structures, their well known marker proteins were fluorescently labeled. Cells for the analysis were produced using microfabricated patterns (see Figure 2.5.7). This resulted in a certain shape, normal cell cycle and thus their functional integrity. Quantitative maps of the spatial organization of intracellular membranous compartments could be created owing to the usage of the patterned cells. Moreover, even slightest changes in endomembrane organization induced by cytoskeleton disruption could be detected [139].

3D image stacks of the grown cells were acquired and deconvolved. Following segmentation yielded positions of the fluorescently labeled marker proteins. These positions were aligned for numerous cells using characteristic landmarks of micropatterns. Further, the probability density function was calculated using Gaussian kernels. Calculated probability density maps of intracellular structures can be used to compare different biological samples. Duong et al. developed a test statistic allowing a density-based comparison of multivariate data [43]. Figure 2.5.8 illustrates microtubules morphology for a control condition (Ctrl) and after the treatment with a drug (NZ). Representative 2D scatter plots show the difference in the marker protein distributions. The quantitative assessment is performed based on the developed test statistic.

Application

"The structural features of cells and the topological relationships between the numerous intracellular compartments give rise to multivariate data whose unbiased, automatic comparison is a major challenge" (Duong et al. 2012, p. 8382). Schauer et al. developed a method which can automatically compare steady state morphological cell organizations. Their method directly assesses the spatial organization avoiding extraction of numerical features and their classification. Intracellular structures are transformed into three-dimensional kernels, enabling their direct mathematical comparison.



(a) (b)

Figure 2.5.9.: (a) Fluorescent image of tri-color stained histological section. Blue - cell nuclei, red - collagen, green - biomarker; (b) phase-contrast image of this section (from Grabe et al. [66]).

The suggested test compares complex data from fluorescent microscopy without reducing the provided information into simple summary statistics [43]. It was also shown to be valid for continuous intracellular structures. And finally, even unconstrained cells morphology could be compared by this approach.

2.5.5. Quantitative spatial profiles

An interesting approach to discover protein localization and networks was suggested by Grabe et al. [66]. They were extracting intensity profiles of epithelial biomarkers and creating a protein network from these data. This was the first quantitative assessment of the epithelial differentiation process while only purely qualitative descriptions have been published before.

As stratified layers of the epithelium differ from each other, the biomarker protein expression levels are also different. To detect and analyze this phenomenon, spatiotemporal protein networks can be constructed and studied. Tissue samples conserving the protein localization information represent a suitable data source. Widely used microarray data is not appropriate here, as the localization information is lost during the sample preparation.

A quantitative characterization of the protein distribution is required for the network construction. Therefore, intensity profiles of the biomarkers are extracted and analyzed. Protein networks can be constructed from these data and give insight into the coregulation of the participating proteins. Even completely novel interaction hints can be provided. Having this in mind, Grabe et al. developed a method for the quantitative spatial analysis of epithelial tissue sections. Automated image acquisition and processing facilities enable analysis of large datasets and provide a basis for a potential high-throughput application.

Technology

Histological sections with immunohistochemically labeled proteins were studied. A tri-color staining was performed labeling cell nuclei of the connecting tissue, collagen and the biomarker of interest (see Figure 2.5.9, (a)). Slides were scanned using an automated microscope, and resulting images were cut into smaller subimages.

The image processing started by detection of the epithelium and its segmentation. Nuclei and collagen staining together with the phase-contrast images (see Figure 2.5.9, (b)) were used for it. The upper border of the epithelium was identified based on all three stainings, and a normalized distance image was constructed. A border line at beginning of the epithelium (E) was described as 0% distance, and the end of the epithelium was set to 100% distance (see Figure 2.5.10).

Intensity profiles of the biomarker (green channel in Figure 2.5.9, (a)) were extracted for all sliding distance intervals in the epithelium. Respective marker intensities were averaged and corrected for the background signal. Obtained profiles were normalized to a maximum of the fluorescence intensity thus enabling comparison between different experiments. These are the resulting quantitative spatial profiles (QSPs) which are further analyzed.

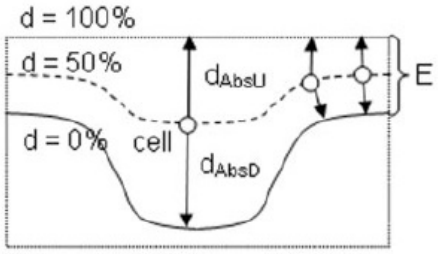


Figure 2.5.10.: An example of normalized distance image (from Grabe et al. [66]).

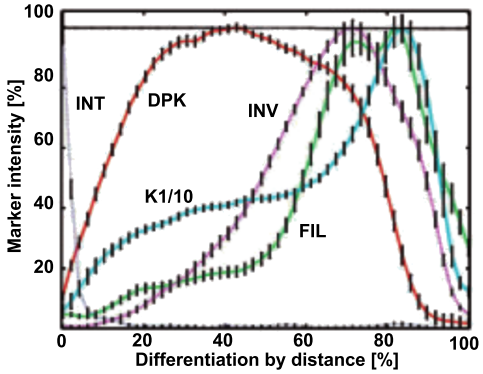


Figure 2.5.11.: An example of QSPs: X axis - relative distance of a pixel to the beginning of the epithelium; Y axis - the normalized fluorescence intensity for a respective marker (from Grabe et al. [66]).

A spatiotemporal protein network was constructed for each histological section. Background signals were ignored as only abundant proteins are considered for this analysis. First, correlating QSPs were detected. If QSPs of two biomarkers correlate, then their coregulation during the differentiation process can be supposed. It is not necessary that some two proteins correlate during the whole differentiation process. Even a part of this time is sufficient for them to be coregulated. If such a period with 100% QSPs correlation can be detected, then the proteins are considered to be connected in a network. A required minimum length of this period constitutes 10% of the total QSP.

Second, a so called protein leadership was identified for each pair of QSPs. A protein is considered to lead if its concentration grows or falls faster than of the other protein. Based on all the individual correlation and leadership networks, a consensus network was constructed. An arrow between any two proteins in the consensus network appeared only if there were at least 2 arrows in the individual networks with the correlation above 50%.

Applications and potential

Grabe et al. studied spatiotemporal expression patterns of five protein biomarkers of epidermal differentiation [66]. An example of the average created QSPs is presented in Figure 2.5.11. Detected intensity of the biomarkers changes with the distance (or the differentiation %). The obtained results for the individual biomarkers correlated well to the numerous qualitative descriptions from the literature. QSPs appeared to be the first method to describe the protein expression in the differentiation process quantitatively thus allowing numerical comparison between experiments.

It was, for example, confirmed that the biomarker integrin $\alpha 6$ (INT) was expressed first in the epithelium (a narrow grey peak in Figure 2.5.11). QSPs of other studied biomarkers

also complied with the previous qualitative knowledge on their expression. Furthermore, the constructed spatiotemporal consensus network (not shown) corresponded well to the known interactions.

The method suggested by Grabe et al. is the first quantitative space- and time-resolving measurement of epithelial biomarkers. Purely qualitative descriptions do not provide a sufficient basis for any network reconstruction or modeling. QSPs contain valuable information on the protein localization and interaction which should still be analyzed. Authors also presented a general workflow of the data processing. In particular, combination of the data from tissue sections of varying width and distinct intensity levels is enabled due to the normalization procedures.

Generally, a spatiotemporal protein network can be created for other biological processes. It does not only describe QSPs, but also illustrate regulatory relationships between the proteins. However, if two proteins are connected in such a network, it does not necessarily mean their direct interaction. It can be a sign of their coregulation or a concerted action mediated by other biomolecules. Thus, previously unknown protein interactions or cascades can be discovered by thorough analysis of the consensus networks. Authors suggest their method to be a valuable quantitative extension to current qualitative conventional approaches of mapping the proteome expression data [66, 129]. This method was developed on images of immunofluorescent tissue sections, but may also be applied to other image types providing sufficient spatial resolution.

3. Biological and medical background

Biological data is crucially important for the development of new analysis methods. I will thus describe biological processes used as models for my work. Owing to the collaboration in the DFG-funded clinical research group "Hepatobiliary transport and liver diseases" (KFO-217 [88]), I had the access to rich data on different biological processes in the liver cells (hepatocytes). Such processes connected to cholestatic liver diseases are the central research topics of this project. In particular, transporter proteins in hepatocyte membranes are observed. Translocation of these proteins takes place in response to certain external or internal factors. Quantitative analysis of these translocations is expected to illustrate the underlying biological processes and their connection to liver diseases. Manual translocation quantification was routinely performed to assess these translocations. I will present this manual method, as it is used to validate the novel implemented automated quantification.

3.1. Biological model: liver and its diseases

The liver is a multifunctional organ with a prominent role in metabolism and its dysfunctions can lead to severe diseases. These are unfortunately widespread and represent a major clinical and socio-economical issue. Cholestatic liver diseases constitute only a subset of possible disorders. Nevertheless, they are very frequent (more than 3 million patients in Germany) and thus are actively researched. Cholestasis may result from dysregulation of transporter proteins in the membranes of hepatocytes, the most abundant liver cell population. For instance, Progressive Familial Intrahepatic Cholestasis (PFIC) is caused by dysfunction of bile transporters. In PFIC patients, certain mutations occur in genes encoding transport proteins involved into bile salt homeostasis [18]. This heritable disease is usually diagnosed in childhood. PFIC often leads to death from liver failure at ages usually ranging from infancy to adolescence [32]. Its exact prevalence is still unknown, but the incidence is estimated to approx. 1/50,000 births. PFIC represents 10 to 15% of causes of cholestasis and 10 to 15% of liver transplantation indications in children [32, 74].

Cholestatic diseases can also be caused as side effects of medication. The liver plays a predominant role in drug biotransformation and disposition from the body [96]. Liver is the first to face and sense intestinally absorbed nutrients. Being constantly exposed to the ingested products, it performs a vital barrier function. Drug-induced liver injury accounts for up to 7% of all reports of adverse drug effects voluntarily reported to pharmacovigilance registries. Drugs directly damage hepatocytes, as they may interact with bile ducts or may change bile flow. The phenotypes commonly encountered thus include hepatitis, cholestasis, cirrhosis, vascular lesions and even fulminant hepatic failure [96].

Moreover, pregnancy can also be a reason for a cholestatic liver state. The Intrahepatic Cholestasis of Pregnancy (ICP) is characterized by maternal pruritus and jaundice in the third trimester [56]. The incidence of ICP varies widely with geographical location and ethnicity. It is most common in South America, where early reports described an overall incidence of 10%. The current incidence of ICP in Europe stays nearly constant around 1%. Interestingly, ICP is more probable for women carrying twins, triplets or more. Female relatives of affected women have a higher than average risk of also being affected [117]. Maternal effects of ICP are mild; however, there is a clear association between ICP and higher frequency of fetal distress, preterm delivery, and sudden fetal death [90].

Liver is also the major regulator of blood glucose homeostasis and amino acid catabolism. It is involved in the synthesis and catabolism of physiologic molecules, the degradation of

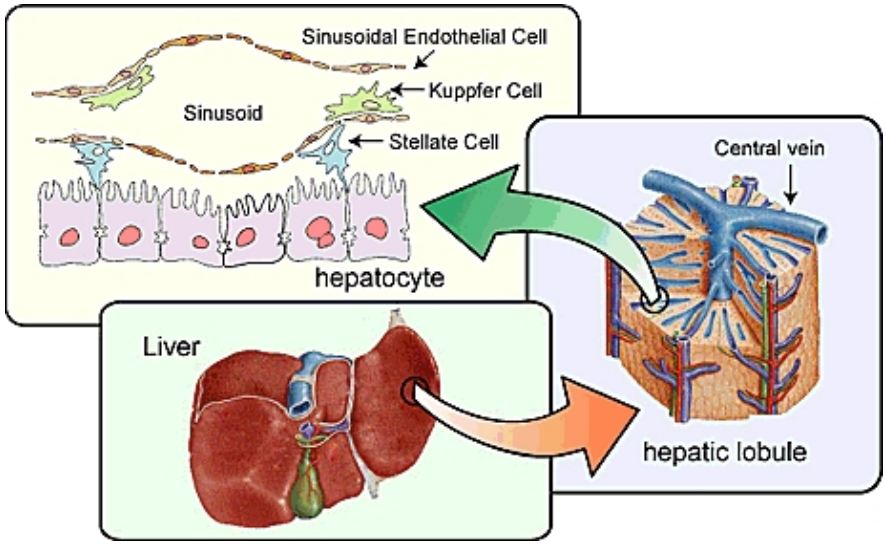


Figure 3.1.1.: Liver and hepatocyte structure (from Akaike Lab [3]).

xenobiotics, the excretion of waste products via the biliary route and the formation of bile. In the following, structural features of hepatocytes will be described together with the functional impairments leading to cholestasis.

3.1.1. Hepatocytes

Hepatocytes constitute the most abundant (60-65%) and functionally diverse cell population of the liver, which exhibit structural and functional polarity [17, 47, 70, 164]. Polarity of hepatocytes is essential to perform vital biological functions (e.g. bile transport). However, different membranes of hepatocytes are only shown in tissue, not in cell culture. The basolateral surface covers nearly 90% of the total surface of the hepatocytes (see Figure 3.1.1). It comprises the sinusoidal membrane and the lateral membrane. The latter one contains tight-junctions to form cellular contacts between adjacent hepatocytes (see Figure 3.1.2). The tight-junctions determine the exchange of fluids and electrolytes between Disse's space and the canalicular space. The *canaliculi* are tiny ducts delimited by the canalicular (apical) membrane of two adjacent hepatocytes [47].

An important function of hepatocytes is the formation of bile. Hepatocytes generate bile flow within bile canaliculi by continuous vectorial secretion of bile salts and other solutes across their canalicular membrane [59]. Bile acids are synthesized in the hepatocytes, secreted into the bile, released into the proximal intestine to facilitate lipid absorption, and reabsorbed in the distal intestine [57]. Bile acids in the portal blood are taken up by the hepatocytes, and are resecreted to continue cycling between the intestine and the liver establishing their enterohepatic circulation [5, 47, 87]. Maintaining a balance between bile acid synthesis and secretion is vital [5], as increased bile salt concentrations may alter cholesterol synthesis, immune functions and even lead to apoptosis [9, 45]. The formation of bile depends on the structural and functional integrity of the bile-secretory apparatus and its impairment leads to cholestatic liver diseases.

Hepatocytes generate the primary bile in their canaliculi, blind tubular structures with a very high surface/volume ratio. The primary bile is then modified by secretory and reabsorptive processes as it passes through the bile ducts. Cholestasis results from dysregulation of bile transporter proteins in the sinusoidal [161] and the canalicular membranes [85]. Long term down-regulation of these transport systems involves changes in mRNA and protein levels [167] or can be explained genetically. The short term regulation, in turn, is determined by covalent modifications of transport proteins (e.g. phosphorylation) [65, 115], substrate availability or

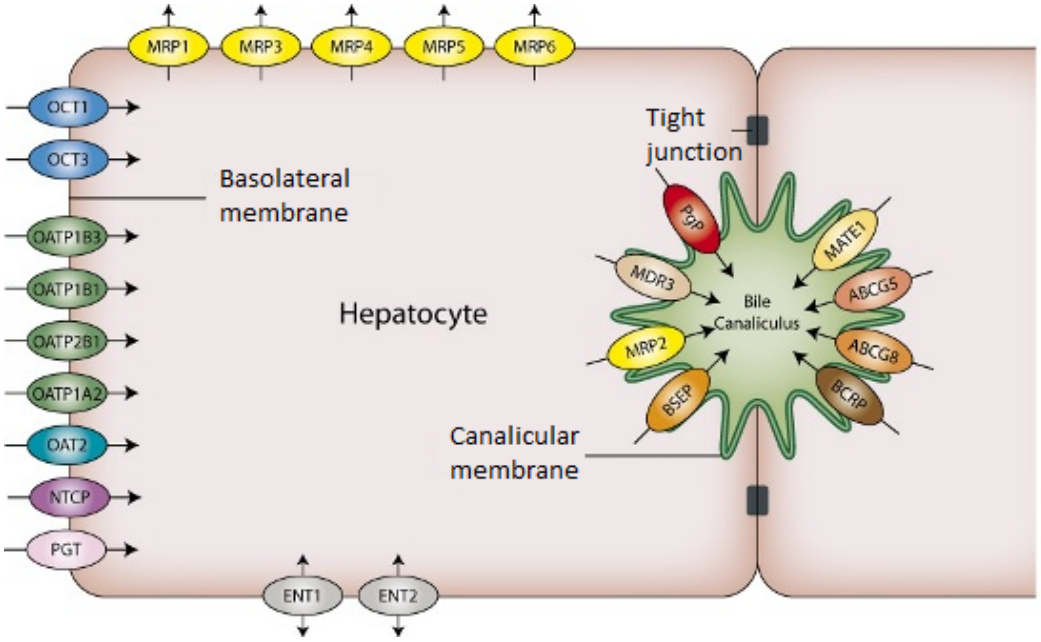


Figure 3.1.2.: Transport proteins in hepatocyte membranes (from Solvo Biotechnology [155]).

competition [69]. Subcellular transporter localization can also influence their function by rapid endo- and exocytosis of transporter-bearing vesicles from and into the respective cell membrane [95].

3.1.2. Canalicular membrane

The conjugate export pump multidrug resistance protein 2 (Mrp2) as well as the bile salt export pump (Bsep) are regulated in rat liver on a short term scale by retrieval from and insertion into the canalicular membrane in response to e.g. anisoosmolarity or oxidative stress [36, 92–94, 97, 135, 141, 142]. Schmitt et al. [142] were able to show in rat livers perfused under normoosmotic conditions (305 mosmol/L) that Bsep was mainly localized in the canaliculi. Canalicular positions were determined by localization of the Zonula occludens 1 protein (Zo-1), a protein of the tight junction complex. Zo-1 delineates the bile canaliculi (see Figure 3.1.3), in relation to which Bsep displacement was observed. Figure 3.1.3 illustrates a similar example where Mrp2 is localized in the bile canaliculi.

Consecutive hyperosmotic exposure (405 mosmol/L) leads to an increase of immunoreactive Bsep inside the hepatocytes. Intracellular Bsep immunostaining exhibits a punctate pattern suggestive for the localization of Bsep in putative vesicles. Hyperosmolarity leads to Bsep retrieval from the canalicular membrane (see Figure 3.1.4), reduces bile acid secretion and results in cholestasis. In contrast, hypoosmolarity favors the insertion of intracellular Bsep into the canalicular membrane, the secretion of bile acids and leads to choleresis. These osmodependent shifts in Bsep location occur within less than 30 minutes and are reversible [142]. I will further refer to the translocation of Bsep between canalicular membranes and cytoplasm of hepatocytes as to the *1D-Can process*.

3.1.3. Basolateral membrane

The basolateral membrane includes other transport proteins which constitute a part of the whole complex bile acids exchange system (see Figure 3.1.2). For example, Ntcp (sodium taurocholate cotransporting polypeptide) is the major transporter for the bile salt uptake at the basolateral

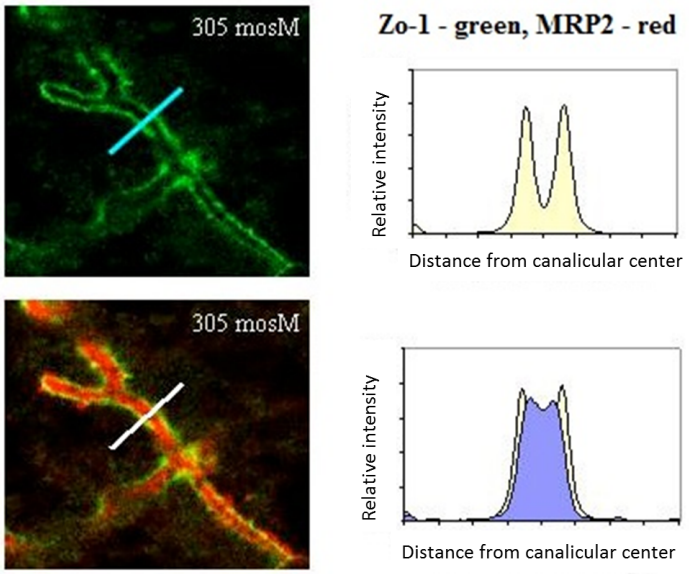


Figure 3.1.3.: Transporter Mrp2 (labeled red) is mostly localized in the canalicular membrane which is detected by Zo-1 (labeled green) (from Kubitz et al. [92]).

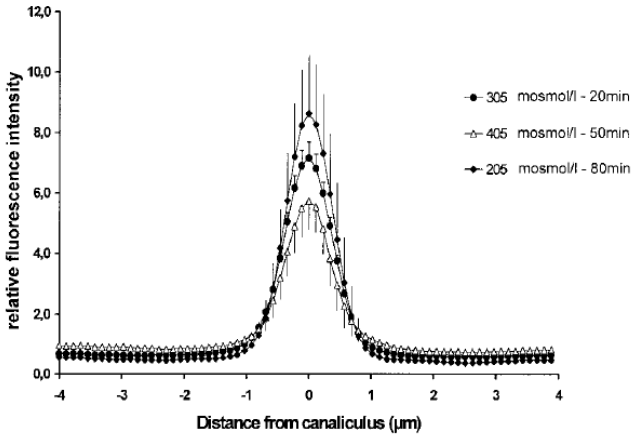


Figure 3.1.4.: Densitometric intensity profiles of Bsep across canalicular membranes. Hyperosmolarity (405 mosmol/L) leads to decreased Bsep concentration in the canalicular membranes, while hypoosmolarity(205 mosmol/L) increases its content (from Schmitt et al. [142]).

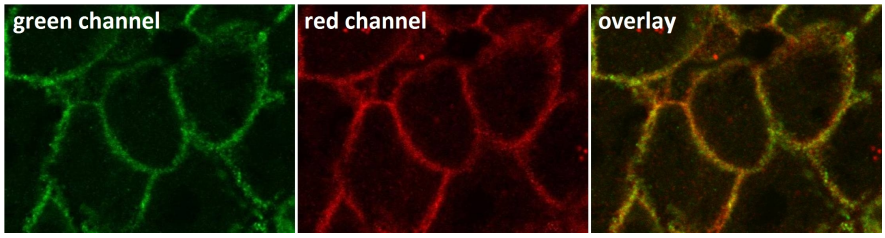


Figure 3.1.5.: Basolateral membrane stained for Na^+-K^+-ATPase (green) and Ntcp (red).

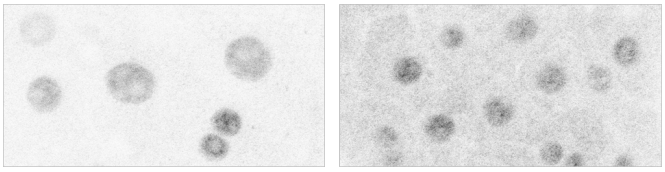


Figure 3.1.6.: Glucokinase distribution depends on the glucose concentration. At low glucose concentration most of the glucokinase is localized in the hepatocyte nuclei (left). Higher glucose concentration leads to the increased glucokinase content in the cytoplasm (right).

membrane. Hence, its localization is of interest as it directly influences its function. It can also be regulated both on the long-term and on the short-term basis. Similar to the transporters at the canalicular membrane, Ntcp is regulated on a short-term basis by insertion into the plasma membrane and internalization into the cytoplasm of neighboring hepatocytes [40]. I will refer to the translocation of Ntcp between basolateral membranes and cytoplasm of hepatocytes as to the *1D-Bas process*.

A standard structural marker used for the analysis of basolateral membranes is the Na^+-K^+-ATPase. Its localization does not change under experimental conditions, and other protein distributions are frequently studied in relation to it. An example of a basolateral membrane stained for the Na^+-K^+-ATPase and Ntcp is shown in Figure 3.1.5.

3.1.4. Glucokinase translocation

A prime function of the liver is the maintenance of blood-glucose homeostasis by rapid clearance of the glucose that reaches the liver via the portal vein after a meal [2]. The enzyme glucokinase (EC 2.7.1.2) [62] catalyzes the initial step in utilization of glucose, the primary cellular substrate [63]. The liver regulates glucose output and uptake during fasting and feeding [104]. Its activity is modulated in a coordinated manner via a complex set of mechanisms, including gene expression, changes in cellular location, and interaction with regulatory proteins [104].

Cellular compartmentalization of glucokinase (GK) is influenced by changing the glucose concentration (see Figure 3.1.6). GKRP (glucokinase regulatory protein) is mainly located in the hepatocyte nucleus, while the cellular location of GK depends on the cells metabolic status [104]. When glucose concentrations in the medium decreases, GKRP bind to GK and transfers it into hepatocyte nuclei [171]. If increased glucose concentration is detected, GKRP releases the GK from the nuclei into cytoplasm, so that it is activated to take part in the glucose metabolism (see Figure 3.1.7). GKRP was shown to be essential to maintain the balance of cellular GK distribution [33]. I will refer to the translocation of glucokinase between nuclei and cytoplasm of hepatocytes as to the *2D-Nuc process*.

The liver is equipped with an insulin-independent system allowing extremely rapid equilibration of glucose concentration [79]. Localization of glucokinase condition the first step of the glucose metabolism. Alterations of the glucokinase function can lead to Maturity Onset Diabetes of the Young (MODY) [68]. MODY is defined as a familial form of early-onset type

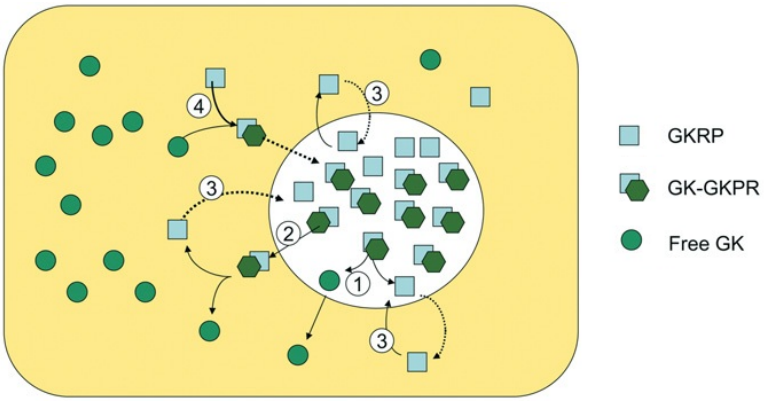


Figure 3.1.7.: Subcellular location of glucokinase (GK). At low glucose concentrations, GK is sequestered in the nucleus, bound to glucokinase regulatory protein (GKRP) as an inactive complex. Addition of glucose causes translocation of GK to the cytoplasm either in a free form (1) or in a GK−GKRP complex (2). The released GKRP may either remain in the cytoplasm or reshuttle back to the nucleus (3). When glucose concentration declines to basal levels, GK binds to cytoplasmic GKRP and is imported into the nucleus (4) (from Agius [2]).

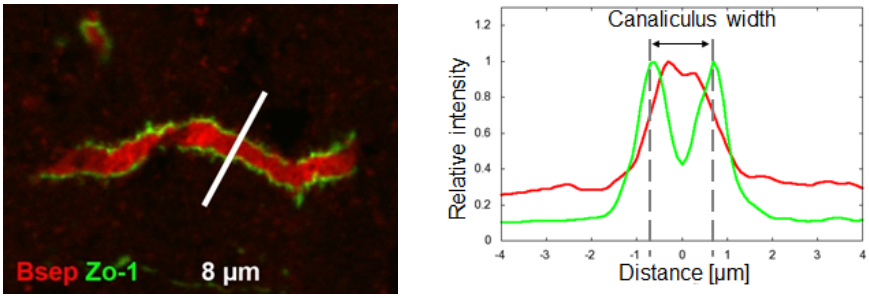
2 diabetes, which usually develops in childhood, adolescence, or young adulthood [170]. Only in 2011 diabetes caused 4.6 million deaths, while the total number of patients was estimated to 366 million people [119]. MODY is estimated to be responsible for 2 to 5% of the diabetes type 2 cases [120]. Patients with MODY are usually detected during routine screening, as the symptoms are not always clearly recognized [68].

3.2. Manual analysis of membrane protein translocation

Several toponomic localization studies of Bsep and Mrp2 were published so far [92, 93, 97, 110, 141, 142], comparing transporter protein distribution by manual processing of microscopic images. By means of immunohistochemistry, proteins of interest (Bsep or Mrp2 and Zo-1) were labeled with fluorescent markers. As described in Section 3.1.2, Zo-1 is used to localize the canaliculi, because the tight junctions stretch as a line where a canalicular membrane meets adjacent hepatocytes. Two roughly parallel Zo-1 lines signal the presence of a canaliculus running parallel to the image plane of the microscope. An example of such a canaliculus is shown in Figure 3.2.1, (a). Microscopic images of stained canalicular membranes were manually assessed and processed by human experts.

The *manual translocation quantification workflow* (*manual workflow*) described in Schmitt et al. [142] proceeds as follows. Confocal microscopy images of rat liver tissue sections were studied to detect Bsep translocation from the canalicular membranes under anisoosmolar conditions. Therefore, densitometric analysis was performed to observe its distribution. The areas for the analysis were chosen by assessing the apparent integrity of the canaliculi. Regions of interest (ROI) were found where the immunostaining of Zo-1 delineating the canaliculi were undisrupted and parallel lines (see example in Figure 3.2.1, (a)). Canalicular segments were expelled when they were bent, small or non uniform. The ROI selection was performed only based on Zo-1 distribution ignoring the distribution of Bsep in order to avoid bias.

A confocal z-stack was recorded on each of 10 positions for the cryosection. Following densitometric analysis was performed using the software Image-Pro Plus [77] as described by Kubitz et al. [92]. The profile of the fluorescence intensity was measured over a thick line (length of 8 μm) orthogonally to the canaliculus. To each of the pixels the mean fluorescence intensity over the line perpendicular to the length was calculated, yielding a mean fluorescence intensity



(a) (b)

Figure 3.2.1.: An example of a canaliculus suitable for the densitometric analysis (a). Respective extracted fluorescence intensity profiles (b).

profile (separately for the green and red channel). The thickness of the line depended on the individual canaliculus. Only straight canalicular segments were chosen while the straightness was normally given over 4 to 6 μm [142].

The statistical variances were calculated from the obtained normalized Bsep fluorescence intensity profiles. Different experiments were compared by statistical tests on these variances in order to test whether observations reflect a pattern and not a chance [122]. A generally applicable unpaired test was selected that does not make any assumtions on the variable distribution. One of the tests meeting all these requirements is the *Wilcoxon Rank Sum Test* (WRST) [174]. It is a widely used nonparametric test [99] which compares observations from independent sets of arbitrary sample sizes. This test estimates equality or inequality of two distributions [131]. A statistically significant difference was detected at p-value ≤ 0.05.

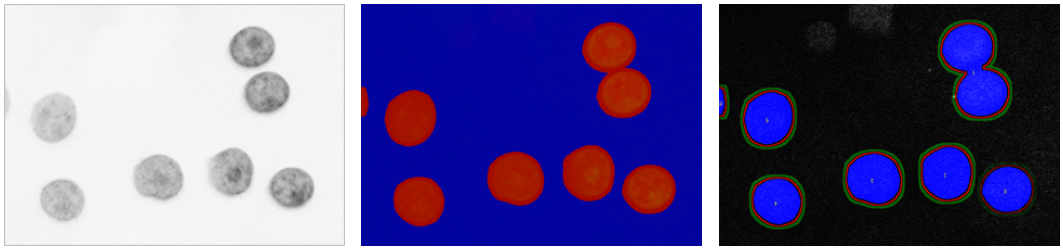
Manual workflow has been recently automatized in part using the in-house software Profilizer (developed by Martin Becker) for the intensity profile extraction [35]. Profilizer users can view images and interactively select positions of intensity profiles. A user has to click two points symmetrically to a membrane fragment in a microscopic image. Profilizer will then draw a line between these points indicating an intensity profile direction. It will also cut the line to 4 μm on both sides of the membrane and output diagrams of pixel intensities for each of the channels (the intensity profiles). These manually extracted intensity profiles are exported for further evaluation.

The length of the extracted intensity profiles was 8 μm (Figure 3.2.1, (a)), corresponding to 81 pixels with a pixel size of 100 nm. Intensity profiles (Figure 3.2.1, (b)) were accepted according to the appearance of Zo-1 fluorescence. Acceptable intensity profiles had two maxima of similar size and a minimum between these peaks close to baseline level. An empirical image analysis showed that canalicular diameter varies in the range of 0.8 – 2.5 μm. Therefore, the profiles were excluded if the distance between the two maximal intensities was < 1.0 or > 2.5 μm. Ten intensity profiles are usually selected per image and evaluated. The immunostaining of Bsep was disregarded for profile selection.

Manual profile extraction can be performed on biological images of any structures. For example, apart from the presented above canalicular membranes, also basolateral structures can be analyzed. A disadvantage of this method, as of any other manual analysis, is its subjectivity. An expert makes the decision where to extract the profiles based on experience, making this method also time consuming and error-prone.

3.3. Automated analysis of nucleus-to-cytoplasm translocation

Cell compartmentalization results in numerous biological processes involving two or more compartments. The nucleus does not only include the genetic information, but also participates in



(a) (b) (c)

Figure 3.3.1.: Nuclei and cytoplasm segmentation using the software Zeta. From the initial image (a) nuclei are segmented (red objects in (b)). (c) Corresponding cytoplasm masks (green) are computed for each nucleus (blue) separated from it (skipped area marked red).

various regulatory cascades. Therefore, a nucleus-cytoplasm pair is widely studied for translocation of regulatory biomolecules, e.g. in cancer research [133].

Early works were based on a manual analysis of microscopic images to identify cell cytoplasms and nuclei [34, 102]. Further analysis consisted of area measurement or intensity-based evaluations. As more and more scientific image processing software became available, detection of the nuclei and cytoplasms could be performed automatically. For example, Morelock et al. [109] automatically segmented stained nuclei [76] and computed masks of the corresponding cytoplasms. For this purpose, intersections of circles centered at the detected nuclei and Voronoi tesselation polygons were found (see Figure 2.2.3). Further, the ratio of the fluorescence intensity in the nucleus to the total intensity was calculated as characteristics for the translocation state.

Later works included staining of the cytoplasm, which enabled its automated segmentation [116]. Similar methods were also applied to study the 2D-Nuc process (glucokinase translocation between nucleus and cytoplasm) [171]. At the Fraunhofer Insitute for Applied Information Technology (FIT) the in-house image analysis software Zeta was extended to tackle this problem.

As described in Section 2.3.1, Zeta can perform image segmentation. This function was applied to the images of hepatocytes (Figure 3.3.1, (a)) with the nuclei stained with DAPI (4',6-diamidino-2-phenylindole, a fluorescent stain that binds to DNA). Nuclear regions were identified (Figure 3.3.1, (b)). As no additional marker was used for the cytoplasm, it was detected as con-centric regions centered at the detected nuclei (green rings in Figure 3.3.1, (c)). To ensure a separation between the nuclei and the cytoplasm, the cytoplasmatic regions were segmented in a certain distance (1.0 μm) from the nuclei (red rings in Figure 3.3.1, (c)).

Once the regions corresponding to nuclei and cytoplasm were identified, the average fluorescence intensity of the glucokinase marker is calculated. The ratio of these average intensities is characteristic for the current biological liver state. This ratio can be used to study influence of the glucose concentration in the medium on the 2D-Nuc process (glucokinase translocation between the nuclei and the cytoplasm).

3.4. Biological SOPs

Data quality is crucial for any analysis and development. Especially in biological research, standards in sample preparation are of a great importance. My colleagues at the University Clinic Düsseldorf have developed and standardized the biological sample preparation workflow.

In the following, I will briefly present some important steps. Detailed protocols can be found in the referenced publications (Kubitz et al [92, 93], Mühlfeld et al. [111], Cantore et al. [20] and Donner and Keppler [38]). For all the treatments, animals received care according to the "Guide for the Care and Use of Laboratory Animals" [118]. The experiments were approved

by the responsible local authorities.

3.4.1. Rat liver perfusion

Rat liver tissue is used to develop automated translocation quantification on the example of the cholestatic liver diseases. Cholestasis occurs, when bile salt transport by the hepatocytes is impaired. One of the possible reasons is the translocation of bile transporter proteins from the hepatocyte membranes (active state) into cytoplasm (inactive state). Cholestasis can be induced experimentally to generate suitable data for the research. For this purpose, rat livers can be perfused in hyper-osmolar buffers or undergo a common bile duct ligature.

Livers of male Wistar rats were perfused *in situ* as described in Kubitz et al. [92, 93]. The main component of perfusion medium was the bicarbonate-buffered Krebs-Henseleit to maintain the liver tissue. In normoosmotic perfusions, the osmolarity was 305 mosmol/L. Hyperosmotic exposure (405 mosmol/L) was performed by raising the NaCl concentration in the perfusion medium.

In other experiments, bile salts were added to the perfusion medium to examine their influence on the translocation of bile transporters. Two studied bile salts are taurocholate (TC) and taurochenodeoxycholate (TCDC). Distribution of bile transporter proteins was analyzed in the tissue samples before and after addition of TCDC or TC.

3.4.2. Bile duct ligature in rat liver

Following general anesthesia, male Sprague Dawley rats underwent double ligature of the proximal common bile duct. This treatment leads to the accumulation of bile and increase of the bile duct diameter. Control animals underwent a sham operation as described in Donner and Keppler [38]. Livers were removed 7 days after bile duct ligature (BDL) or sham operation.

3.4.3. Cryosectioning and immunostaining

Sample preparation and immunostaining were performed according to a SOP to assure reproducibility [35]. The tissue samples were cut in 5 μm sections and fixed with methanol. Further, slides were washed and immunohistochemically labeled using indirect fluorescence. The slides were first incubated with a combination of the primary antibodies (rabbit anti-Bsep and mouse anti-Zo-1, or rabbit anti-Ntcp and mouse anti-Na^+-K^+-ATPase, respectively). Subsequently, a combination of the secondary antibodies (Alexa Fluor 488-conjugated goat anti-mouse, green; Alexa Fluor 546-conjugated goat anti-rabbit, red) was added [20, 111].

3.4.4. Image acquisition

Immunostained rat liver tissue samples were analyzed using a LSM 510 confocal laser scanning system with a 63 × Plan-Apochromat objective (NA 1.4), Zeiss, Jena, Germany. The excitation wavelength was 488 nm for Alexa Fluor 488 and 543 nm for Alexa Fluor 546. Emission was detected by a 505 – 530 nm (green) and a 560 – 615 nm (red) bandpass filter. Image acquisition was adjusted to a final pixel size of 100 nm.

For each cryosection, images from 10 different regions were taken in a randomized fashion. Only samples that were prepared in parallel in all steps were compared, using the same adjustments for all parameters (i.e. laser power, filter settings, pinhole, lens, voltages at the photo multiplier tubes, format size and zoom, scan speed, and z-step size when whole thickness of the tissue samples) were analyzed [20].

3.4.5. Flow cytometry

Flow cytometry was performed in order to independently validate results of the densitometric analysis of microscopy images. The experiments were performed according to the protocol

described in Mühlfeld et al. [111]. Briefly, a FLAG-Ntcp-EGFP plasmid was used to assess Ntcp plasma membrane localization by the analysis of intra- and extracellular fluorescence. This plasmid was transfected into HepG2 cells [158], an immortal hepatocyte cell line. Three stable independent clones were established.

Further, FLAG-Ntcp-EGFP expressing HepG2 cells were treated with different bile salts [111] to examine their influence on the distribution of Ntcp. Cells were centrifuged, resuspended and filtered through a gauze. The cells were immunohistochemically labeled with Alexa Fluor R-647-PE. Fluorescence intensities of EGFP and Alexa Fluor-647-PE were measured in 20.000 single cells in a flow cytometer, using an excitation wavelength of 488 nm. The cells were gated for their characteristic forward and sideward scatter. Fluorescent measurements were acquired at 530 $\pm$ 30 nm (EGFP) and $>$ 670 nm (Alexa Fluor 647-PE).

3.4.6. Preparation of datasets for the analysis of glucokinase translocation

Primary hepatocytes were isolated from rat liver and cultivated. The cells were incubated with various glucose concentrations (0 – 4.5 g/L). In addition, the cells were incubated with other reagents (e.g. with Insulin, Sorbitol, etc.) to assess their influence on glucokinase translocation.

DNA was labeled with Hoechst 33258 to localize the nuclei. Its excitation wavelength is approx. 350 nm, while the emission is detected around 461 nm. Glucokinase was immunohistochemically labeled with Cy3. The excitation and emission wavelengths are 550 nm and 570 nm, respectively. Confocal images were acquired using a LSM 510 confocal laser scanning system with a 63 $\times$ Plan-Apochromat objective (NA 1.4). The final pixel size was adjusted to 110 nm/pixel.

4. Novel approach to translocation quantification

4.1. Introduction to transport processes

Transport processes play a vital role in living organisms. These continuously operating complex systems fully depend on biomolecules appearing at the right time and the right place. Only in this case, the biomolecules can be involved into the processes they are required for.

Cellular organization is complex, but highly structured due to the *compartmentalization principle*. It permits a precise subdivision of the cell lumen providing it with functionally specialized aqueous spaces [4]. For instance, the endoplasmic reticulum produces proteins, mitochondria generate energy for the cell, lysosomes hydrolyze enzymes, etc. Biomolecules produced in the respective organelles have then to be delivered to the locations they are required at.

Cell compartmentalization also leads to the presence of numerous membrane structures. Most organelles are surrounded by membranes, so that their content is separated from the cytoplasm. Thus, if a biomolecule has to be delivered into or exported from such an organelle, a membrane needs to be crossed. The same applies to all inter-cellular processes. Membranes are those bounding surfaces (natural barriers) which separate volumes (cytoplasm, organelles, vacuoles, etc.). Hence, numerous closed volumes in the cell lumen enable different concentrations of biomolecules within them. Otherwise, all biomolecules would be homogeneously distributed in the cell lumen due to diffusion.

Living organisms depend on distribution of biomolecules and on transport processes that maintain the balance. There exist several types of transport processes, ranging from passive diffusion to active transport mediated by a carrier molecule (e.g. transport protein). They can also be subdivided into directed ones and a simple thermal movement (Brownian motion). The directed transport processes are in the focus of this thesis, as their quantitative assessment can be useful in research purposes.

As mentioned before, cell compartmentalization allows for different concentrations of biomolecules in different cellular compartments or a cytoplasm. Hence, concentration gradients run across the membranes. If a trajectory is orthogonal to the membrane, it represents the *optimal transport trajectory* (see Figure 4.1.1). Optimal transport trajectory for the movement of biomolecules is found in most directed transport processes due to its highest efficiency. For example, active transport requires energy to move biomolecules from the regions of their lower into the regions of their higher concentration. The shorter the way, the less energy will be spent. Passive transport (e.g. diffusion) also proceeds along the concentration gradient [150].

In principle, the exact transport mechanism must be known to model and track biomolecules (e.g. within biological images). To simplify this analysis, we restrict ourselves to the transport across a membrane structure that operates along a straight line orthogonal to the membrane (optimal transport trajectory). Most physical processes will show this behavior if some gradient is established across the membrane. Therefore, distribution of biomolecules will be further observed along the optimal transport trajectory.

4.2. Observation of transport processes

To quantify a transport process, biological samples have to be prepared and evaluated. For example, distribution of biomolecules can be analyzed biochemically. Alternatively, imaging could

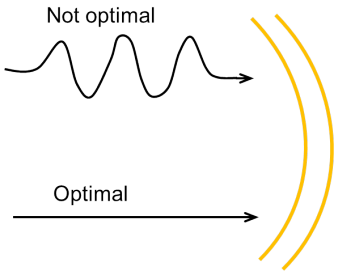


Figure 4.1.1.: Optimal trajectory for directed transport of biomolecules.

be applied for this purpose. Optical microscopy is often the method of choice for biological analysis. It is a well-established technique to image samples on different scales, from macroscopic to atomic. Microscopy images acquired at certain moments in time represent snapshots of molecule distribution in a continuous transport process. Frequently confocal laser scanning microscopy is preferred as it acquires images of the highest resolution. Not only 2D images can be produced, but also 3D analysis is enabled by acquisition of z-stacks. Biological processes which are three-dimensional by nature, can then be studied as such. Still, high-throughput applications generally use less computationally expensive 2D data. 3D or 4D (time series of 3D data) are mostly acquired in research purposes, providing valuable information on the dynamics of the cellular processes.

Although distribution of biomolecules is frequently studied by imaging, molecules of interest are usually smaller than the resolution of the imaging device. The problem of insufficient image resolution is overcome by observation of molecules in a set and not individually. To detect molecules concentration optically, a densitometric analysis can be applied (see Section 2.5.3). Densitometry generally refers to the measurement of the optical density in the sample and inferring the concentration from these data.

Distribution of biomolecules can be measured in fluorescence images. Proteins, for instance, can be tagged with fluorophores using antibodies. The recorded pixel intensity in the images is proportional to the fluorophore concentration. Assuming the stoichiometric binding between the fluorophore and the molecules of interest, the concentration of the latter one can be assessed. I will further refer to *densitometry* or *densitometric analysis* performed on fluorescence microscopy images, as they are the information source used in my work.

As discussed earlier (see Section 2.5.3), densitometry can be performed either at points, along a line (1D) or in some closed region (2D). Line densitometry results in vectors representing distribution of biomolecules along some line. For example, Grabe et al. applied 1D densitometry to study protein distribution across the stratified epithelial layers [66]. When taken along the concentration gradient, 1D distribution vectors will contain the information essential to monitor a directed transport process exhibiting the *optimal transport trajectory*.

4.2.1. 3D vs. 2D data: z-stacks vs. individual images

Cellular processes are three-dimensional and need to be studied in all 3 dimensions in principle. However, they are frequently analyzed on the basis of 2D images due to several reasons. First, there are no widespread native and cost efficient 3D optical devices. Acquisition of confocal z-stacks allows an approximation to the depiction of the 3D nature of the processes. Second, the smallest sufficient information source should be used in high-throughput applications to meet the compromise between the valuable data and time costs.

Biological structures normally do not exhibit uniform spatial organization like crystals. Membranes and other biological structural elements (small vessels, etc.) are usually found in all angular orientations. Hence, costly 3D analysis may be in principle substituted by acquisition and evaluation of numerous 2D images. For example, a homogeneous tissue sample (e.g.

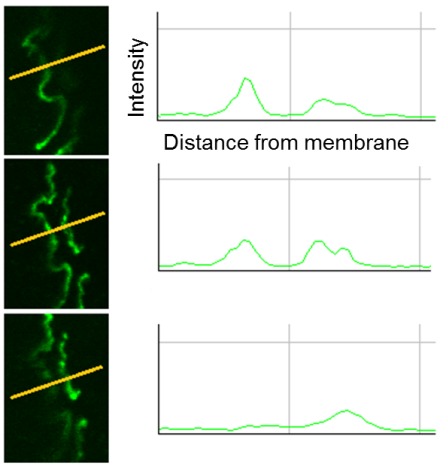


Figure 4.2.1.: Example of a canaliculus depicted in 3 consecutive images of a z-stack with respective densitometric profiles of a structural marker (Zo-1). This membrane fragment is cut in its principal axis only by one optical plane (middle).

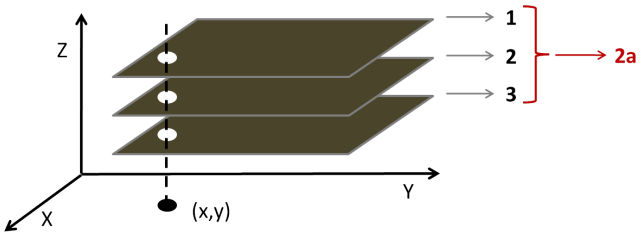


Figure 4.2.2.: 3D information extraction from a z-stack.

liver biopsy) can be well analyzed by 2D microscopy images acquired under any angle. As membranes and other biological structures can be found at all spatial orientations, a 2D image acquired under a random angle is expected to depict at least some of them in its image plane. Analysis of a sufficiently large number of 2D images is thus expected to be an approximation to computationally costly 3D evaluation.

Individual 2D images from a confocal z-stack may contain sufficient information to reliably evaluate a 3D biological process. Nevertheless, a combination of such 2D images may be beneficial. Figure 4.2.1 illustrates a small canalicular fragment which is not parallel to all of the depicted focal image planes. Not all of the extracted densitometric profiles are symmetrical. This canaliculus can only be optimally imaged in one optical layer (e.g. the middle layer in Figure 4.2.1). However, this optical layer is not necessarily also the optimal one for the rest of the biological structures.

To overcome this problem, I suggest 3D information extraction by combination of 2D images from confocal z-stacks (see Figure 4.2.2). Every n neighboring layers (e.g. $n = 3$) are averaged in order to better depict those membranes, which are not exactly parallel to the focal plane. This operation can improve suboptimal intensity profiles if the information from the neighboring layers is of a better quality. Initial confocal layers 1, 2, and 3 are averaged to produce a new layer $2a$ (see Figure 4.2.2):

$$I(2a, x, y) = \frac{I(1, x, y) + I(2, x, y) + I(3, x, y)}{3}, \tag{4.1}$$

where $I(j, x, y)$ corresponds to the pixel intensity in the confocal layer j at the position (x, y). The number of the averaged layers n depends on the size and properties of the biological

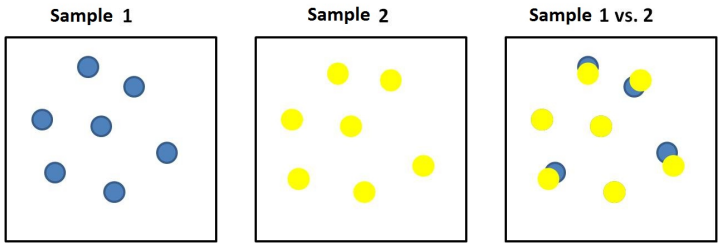


Figure 4.3.1.: Translocation of biomolecules detected in two images of one time series by their overlay.

structure studied. n can vary between 2 and the largest required number of neighboring confocal layers to cover the whole biological structure.

4.3. Generalization of translocation quantification and optimal 1D transport model

Transport of a biomolecule refers to its movement (either itself or by a carrier molecule) from one location to another. If biomolecules are moving from one functional region to another, we talk of a *translocation*. Translocation processes are widespread in the living organisms. The compartmentalization principle leads to the presence of numerous organelles representing closed aqueous volumes. Organelles are separated from the cytoplasm by membranes which allows for a different concentration of molecules in organelles and in the cytoplasm. This serves as a basis for translocations proceeding along the concentration gradients (across the membranes). Further, biomolecules frequently integrate into membranes, functioning as coordinators and sites of major activity [21]. Hence, according to my definition, all translocations include a membrane. It either represents a barrier or a functional region by itself.

To assess translocation of a biomolecule of interest (further called a *functional marker*) from static images, at least two biological samples have to be analyzed and compared (e.g. the treated sample (1) and the untreated sample (2)). If functional marker distribution differs from the sample (1) to the sample (2), then the translocation is assumed.

Evaluation of functional marker translocation is straightforward if samples (1) and (2) constitute a time series. With a suitably low density of the functional marker, individual points can be tracked. In this case, respective images can be overlaid and functional marker distributions can be compared (see Figure 4.3.1). Overlapping regions or degree of localization change could be calculated from these images. Still, neither general direction of the translocation, nor its reliable quantitative measure can be obtained if only the functional marker was labeled in the biological samples.

This method can not be applied to assess translocation of biomolecules in images of two different biological experiments which are not a part of one time series. Overlay of completely different images would not yield any information on the translocation. To overcome this problem, the functional marker translocation can be quantified in relation to some biological structure. One of the stable components of this biological structure has to be labeled, too. This *structural marker* is a biomolecule whose localization is known and is visible under various experimental conditions. According to my definition, translocation processes involve one or another membrane. Thus, a membrane protein meeting these requirements can be selected as the structural marker. For instance, housekeeping or structural membrane proteins are frequently used for this purpose [49, 160].

Figure 4.3.2 illustrates a model of the directed transport adopted in the scope of my work. It represents the functional marker translocation along the optimal transport trajectory (concentration gradient) across the membrane and orthogonally to it (*1D translocation*). This model

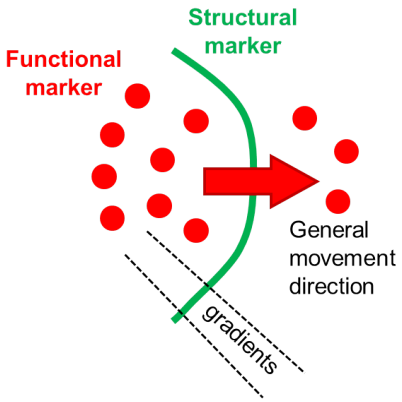


Figure 4.3.2.: Optimal 1D transport model for directed translocation of the functional marker (red) across a membrane (structural marker, green) along the optimal trajectory.

is further referred to as *optimal 1D transport model*. Figure 4.3.3 illustrates an example of a translocation corresponding to the optimal 1D transport model. The functional marker (red) is translocated across the membrane. One of the stable membrane proteins is selected as a structural marker and is immunohistochemically tagged (green).

Biological images with at least two labeled markers contain sufficient information for the translocation quantification, and will be referred to as *multi-marker images*. Such multi-marker images of biological samples prepared under different experimental conditions can be compared. Information on the marker distribution can be extracted by densitometry. For instance, line densitometry results in 1D vectors (or *profiles*) of numbers describing individual pixel intensities corresponding to concentrations of biomolecules. Such a *profile extraction* is performed for the processes corresponding to the optimal 1D transport model. The profiles are extracted orthogonally to the structural marker (along the concentration gradients).

When these data have been collected, the profiles can be evaluated and statistically compared as density vectors. First, they have to be normalized to exclude the variability of any factors influencing fluorescence intensity of the tags used (e.g. binding affinity). This can be done by setting either the total profile intensity or the maximum intensity value to 1.

Example profiles for the biological samples 1 and 2 (e.g. initial and final state) corresponding to the translocation process from Figure 4.3.3 are shown below. Green curves represent fluorescence intensity distribution of the structural marker. It is the same in both samples, as a stable membrane protein was selected as the structural marker. Red curves show distribution of a functional marker, and it changes from sample (1) to sample (2). The profile of sample (2) exhibits an elevated intensity (or concentration) on the right and decreased concentration of a functional marker on the left side of the membrane relative to sample (1). This suggests that functional marker translocation took place in sample (2). This translocation will then be quantitatively assessed based on the extracted profiles.

The upcoming chapters take 2D fluorescence images as an example of the data for translocation assessment. However, general ideas are also applicable to the data from other sources. 2D images are analyzed as a simplified representation of complex 3D biological processes. Following projection of the directed transport processes onto 1 dimension along the concentration gradient and evaluation of the extracted densitometric profiles leads to further significant information reduction. Consequently, much faster, simpler and more efficient analysis can be performed with 1D vectors than with 3D data.

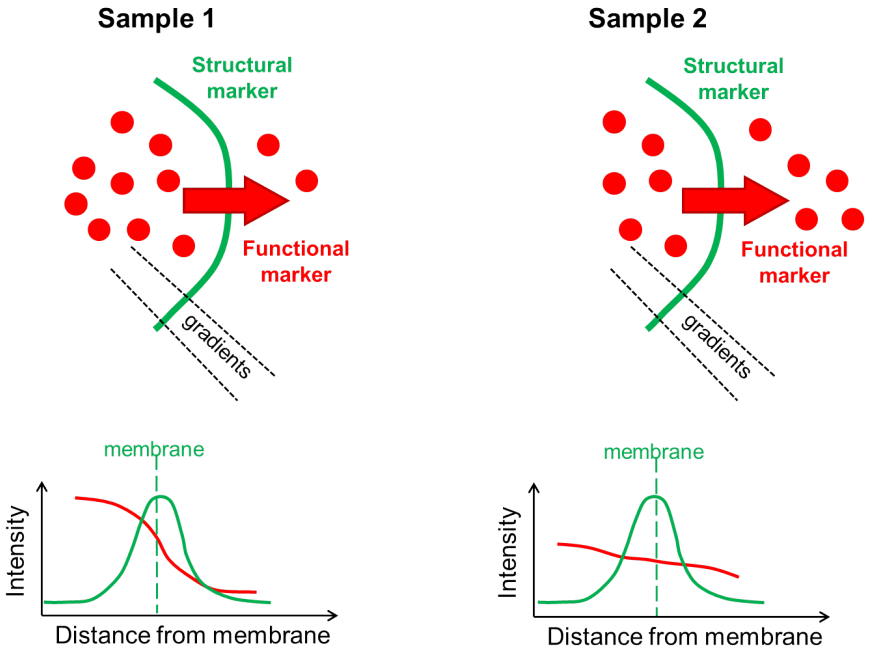


Figure 4.3.3.: Optimal 1D transport model for translocation of biomolecules across a membrane. The sample 1 is compared to the sample 2 based on the corresponding extracted intensity profiles.

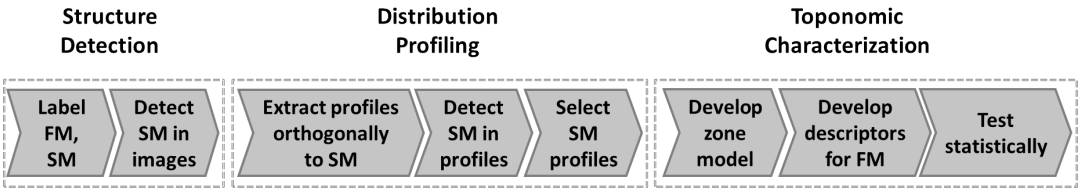


Figure 4.4.1.: Structure-based workflow. FM - functional marker, SM - structural marker.

4.4. General steps for translocation quantification

In the following, I will define general processing steps required to quantify translocation of biomolecules corresponding to the optimal 1D transport model. Figure 4.4.1 summarizes the workflow presented below. Briefly, structure detection is performed first to localize biological structures in the images. Further, line densitometry extracts information on the distribution of the functional marker perpendicular to the membranes (distribution profiling). And finally, functional marker distribution in relation to the membranes is toponomically characterized. Numeric descriptors are developed for this purpose. They enable statistical evaluations of biomolecule distribution. These steps constitute the defined *structure-based translocation quantification workflow (structure-based workflow)*.

4.4.1. Structure detection

First, a functional marker has to be tagged to be detectable in multi-marker images. This can be achieved, for instance, by immunohistochemical labeling with a fluorescent dye.

Second, a taggable structural marker that delineates borders of biological (cellular or subcellular) structures serving as a reference for the translocation has to be selected. Oftentimes a

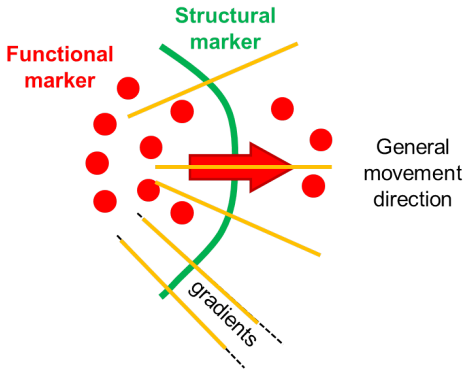


Figure 4.4.2.: Possible positions to extract intensity profiles (yellow lines). The profiles run along concentration gradients and are centered on the structural marker positions.

membrane protein is chosen as such, as its localization is stable and independent of experimental conditions. This is a great advantage, as numerous membranes are present in cellular organisms due to the compartmentalization principle. A membrane protein suitable as a structural marker can be found for almost any translocation process. The selected structural marker should be labeled, too.

Third, the structural marker has to be detected in the multi-marker images. Image processing can be applied for this purpose. Structure detection could be performed by thresholding of intensity values. A threshold has to be selected to distinguish between background and foreground signal. Alternatively, structure recognition or pattern detection can be performed by more sophisticated algorithms. For example, the earlier presented trainable Zeta software (see Section 2.3.1) uses machine learning to detect foreground regions.

Structure detection in biological images is often problematic due to natural variation and irregularities in biological structures. Hence, it should be avoided if acceptable results can be achieved without this computationally expensive step. In Section 4.4.3 I will suggest an alternative approach based on peak detection in the extracted 1D distribution profiles.

4.4.2. Distribution profiling

Once the structural marker was detected in the images, trajectories of the functional marker translocation can be "found". They run along the concentration gradients orthogonally to the detected structural marker objects (see Figure 4.3.3). Intensity profiles for the structural and the functional marker can then be extracted along these trajectories. Such profiles are further called *structure-based profiles* and the respective profile extraction is further referred to as the *structure-based profile extraction*.

Profile extraction

One of the possible 1D translocation scenarios is that the functional marker is *moving* from one side of a biological structure to the other across the membrane (structural marker). Otherwise, the functional marker can be included into/excluded from the membrane. Anyhow, the profiles have to sufficiently cover both sides of a biological structure around the membrane (or any other boundary surface taken as a reference for the translocation). The profiles could be centered at the structural marker positions (see Figure 4.4.2) if the membrane separates symmetrical objects (e.g. two cells of equal size). Optimal profile length is selected for every studied case individually as it depends on the particular biological process, cells or structure sizes and other factors. The extracted profiles are further statistically analyzed.

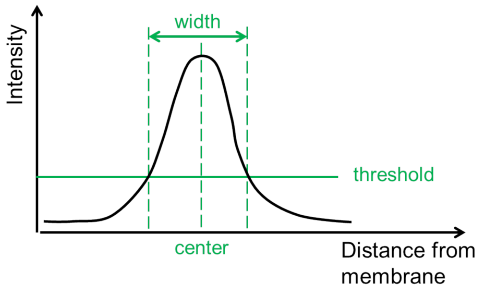


Figure 4.4.3.: Signal detection in distribution profiles.

Profile selection

Only a subset of the extracted intensity profiles represents the translocation phenomenon. Particularly in confocal imaging, the image itself and the extracted profiles are not representative if the focal image plane does not cut the biological structure in the principal axis (see top and bottom images in Figure 4.2.1). In manual image analysis, humans select membrane fragments that are depicted symmetrically, contrasty and cleanly. In order to implement such a strategy in the automated algorithm, intensity profiles suitable for translocation quantification should be selected. This selection is carried out based only on the structural marker profiles. The corresponding distribution of the functional marker is ignored to avoid bias.

To analyze the structural marker profiles, positions of the structural marker objects (membranes) have to be identified in each of them. Peak detection is performed on these profiles to yield central positions of the objects (see Figure 4.4.3). Widths of the structural marker objects can also be identified, for example, by thresholding of the intensity values. These operations represent a kind of a structure detection. However, such a structure detection on 1D profiles is simpler and much less computationally expensive than on 2D images.

Selection of the structural marker profiles is performed based on these data. First, membrane segments non-parallel to the image plane are filtered out. Profiles extracted there can be identified if properties of the studied biological structure are known. A suitable example are the profiles in Figure 4.2.1. There, only one profile (in the middle) is symmetrical, as the biological structure where it is extracted is parallel to the image focal plane. Two other profiles will be rejected due to the lack of symmetry. Further selection conditions rely on the statistical knowledge on the structure parameters. For example, profiles extracted at extremely narrow or wide membranes can be filtered out. Noisy profiles are also eliminated, as they are likely to exhibit numerous local maxima and minima corresponding to background signal. Figure 4.4.4 illustrates some examples of accepted and rejected profiles of the structural marker at the canalicular membranes (Zo-1).

An arbitrary number of such criteria can be applied, depending on the available biological knowledge and the required strictness of the selection. The selected profiles *illustrate* my idea on the perfect structural marker object and the corresponding profile. Positions on the biological structures where structural marker profiles were accepted are thus considered valid for the evaluation of the translocation. Respective functional marker profiles are extracted only there and used for the quantification. The profile selection has a great impact on the quantification. It should be minimally strict to still account for the natural biological variation, but specific enough to eliminate damaged regions and noise.

Profile extraction and selection yield a set of profiles that clearly depict the biological structure studied. Such profiles are representative for the translocation quantification. Profile extraction and selection constitute the *distribution profiling*.

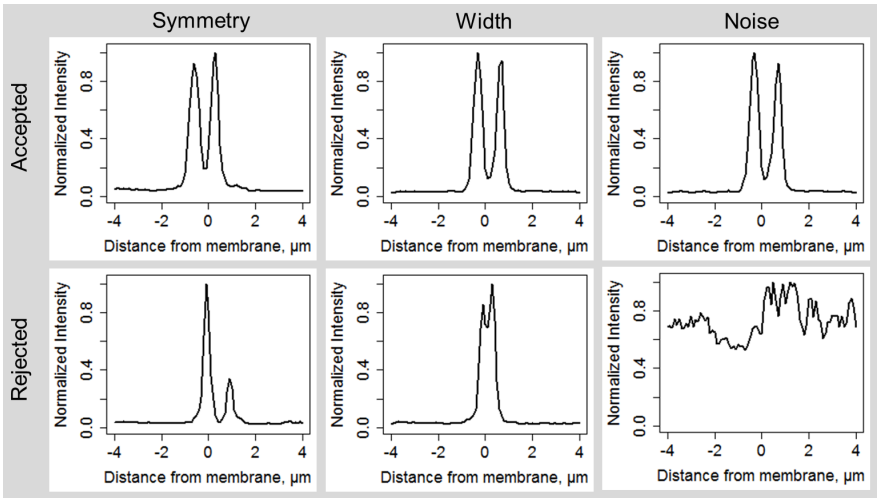


Figure 4.4.4.: Examples of selected and rejected structural marker (Zo-1) profiles at canalicular membranes.

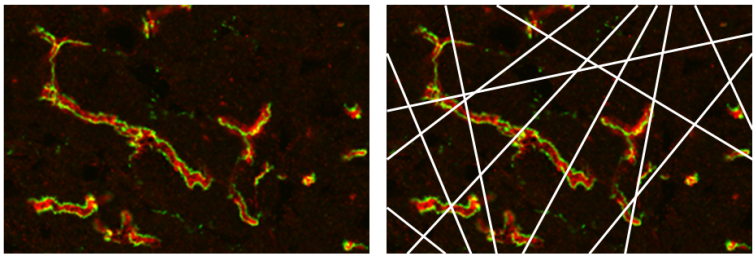


Figure 4.4.5.: Random profile extraction along random lines that are drawn in initial images.

4.4.3. Avoiding structure detection

As discussed above, structure detection in biological images is complicated due to the natural variation. Avoiding this processing step could not only speed up the computation, but also make the algorithm more generally applicable.

The underlying idea of the suggested translocation quantification is to extract the information on distribution of biomolecules from multi-marker images and evaluate it statistically. According to the defined structure-based workflow, intensity profiles represent sufficient information source for such an evaluation. Hence, the aim of the image processing is the extraction of the distribution profiles. Following this logic, any other operations yielding such profiles may be applied instead.

Time consuming structure detection in biological images may be avoided by drawing numerous random lines in these images and evaluating profiles extracted along them (see Figure 4.4.5). I will further refer to such profiles as *random profiles*. Such a profile extraction is further referred to as the *random profile extraction*.

To access the information on the distribution of biomolecules, the extracted 1D random profiles have to be analyzed exactly like the profiles obtained after structure detection in the images (see Section 4.4.2). In particular, peak detection on the random profiles will localize the membrane objects. Objects' parameters (e.g. width) will be calculated from the random profiles. Further, profile selection will be performed based on these parameters. The same selection criteria will be used that were defined for the profiles extracted after structure detection in the images. Herewith, complex and time consuming structure detection in multi-marker

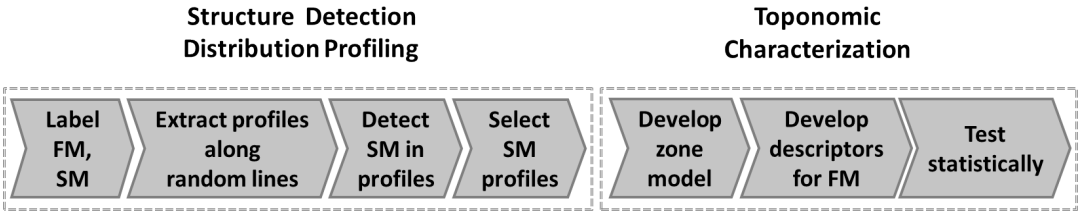


Figure 4.4.6.: Random lines-based workflow. FM - functional marker, SM - structural marker.

images is avoided. Instead, structure (membrane) detection is done by peak detection on 1D profiles. This is the part of the analysis that has to be performed anyway for the profile selection.

The random profile extraction is more efficient than the structure-based profile extraction. Moreover, it is expected to be applicable to an even broader range of biological structures due to the absence of the specific and complex structure detection step. If the number of random lines is sufficient, random profiles will contain enough information on the distribution of biomolecules to evaluate the translocation.

The workflow including the *random profile extraction* is further referred to as the *random lines-based translocation quantification workflow* (random lines-based workflow). Figures 4.4.1 and 4.4.6 illustrate again the difference between the structure-based and the random lines-based workflows.

4.4.4. Toponomic characterization

We assume that translocation took place if the distribution of biomolecules differs in two analyzed biological samples. I presented ways of information extraction from 2D multi-marker images leading to 1D vectors describing the distribution of biomolecules along the concentration gradients. Further ideas for the translocation quantification and analysis rely on the optimal 1D transport model and are based on evaluation of such extracted profiles. These profiles represent distribution of biomolecules in the studied biological samples. This distribution can also be described as a complex function f:

$$\textit{distribution of biomolecules} = f\ (\textit{time, initial biological conditions, biological structure, concentration, external influence,..}),$$

where most of the parameters are unknown, except for the time and external influence (experimental conditions). Neither all the parameters, nor the function f itself might be precisely determined. The only available information about this function is its *shape* in the analyzed samples (see profiles in Figure 4.3.3). Such distribution functions of different samples should be compared to assess the translocation of biomolecules. As these functions are unknown, a general method has to be developed for their comparison. This method should be able to quantify the translocation of biomolecules by analysis of distribution profiles extracted from different biological samples.

Parameter values could be estimated for a known function either directly or by fitting. For example, Gaussian curves are characterized by μ (mean) and σ (standard deviation). If distribution of biomolecules could be described by a Gaussian function, then comparison of the biological samples would be possible by comparison of the μ and σ calculated from the respective intensity profiles. A similar logic applies to any function whose equation is known. Parameters of the function f are determined for the samples (1) and (2) and serve as a basis for their comparison. In many cases, however, the function f is and stays unknown. It can not be assumed, that any distribution could be described by a Gaussian function. Consequently, there is no mathematical reasoning to use μ and σ as parameters for the comparison; and a more general analysis method needs to be found.

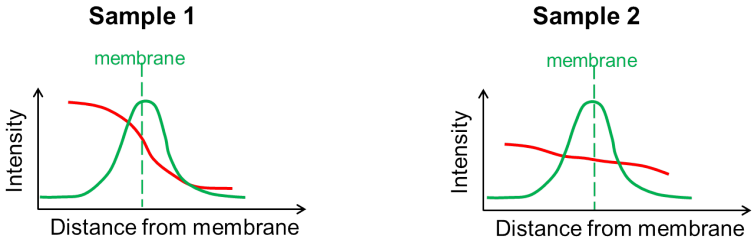


Figure 4.4.7.: Extracted intensity profiles for the optimal 1D transport model from Figure 4.3.3.

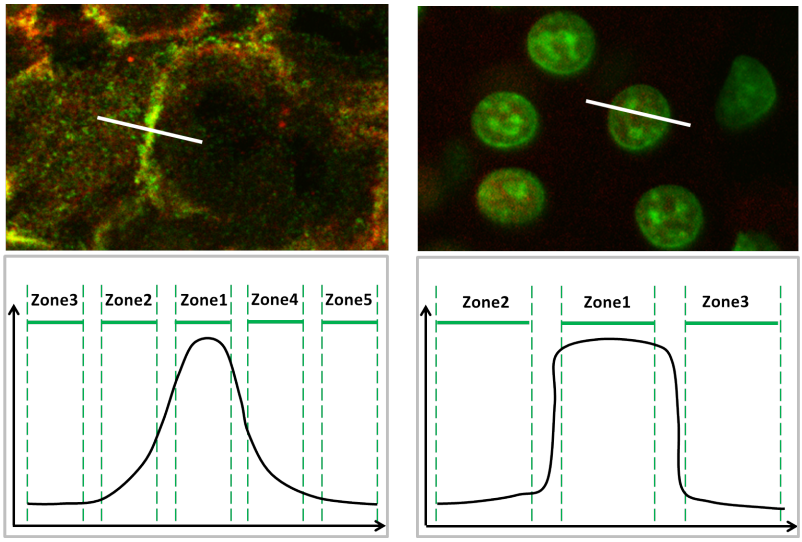
Similar to μ and σ for Gaussian functions, parameters for 1D profiles may be calculated for their mathematical comparison. These *descriptors* are numeric variables calculated from combinations of individual intensity values from the profiles. Specific descriptors can be developed for any translocation depending on the underlying biological process. Let us illustrate this on the profiles from Figure 4.3.3, repeated here for the convenience (see Figure 4.4.7).

Numerous different combinations of the intensity values can be created from 1D profiles. For instance, one could compare the summarized intensities of the functional marker on the left and on the right sides of the membrane. For sample (1), summarized intensity on the left side will be much larger than the one on the right side, while for sample (2) the difference will be less significant. Herewith, the difference between the functional marker distribution in sample (1) and sample (2) can be shown and quantified. However, biological experiments do not provide stability in the achieved intensity values of fluorescent markers. The comparison of absolute intensity values should then be avoided. This can be achieved, for instance, by normalization of these values to the range 0 - 1. Another option would be to develop dimensionless descriptors (e.g. ratios of intensity values). They will allow for reliable statistical comparison of different experiments. In the considered example from Figure 4.4.7, such a dimensionless descriptor could be the ratio of the summarized intensities on the left and on the right sides of the membrane. Unlike the absolute intensity values, this descriptor is dimensionless and can be used to compare samples prepared under different experimental conditions.

This simple example illustrates that for every biological process and distribution function, specific numerical descriptors for the distribution profiles can be developed. Descriptors represent a further information reduction step from 1D vector to a single number. Numeric descriptors, developed individually for each case, characterize the unknown distribution function f. They can thus be used for dataset comparison, similar to μ and σ for Gaussian functions. Calculation of the numeric descriptors is thus a *toponomic characterization* of the molecule distribution.

To develop numeric descriptors best fitting to a particular biological process, distribution profiles can be subdivided into zones. A *zone model* is a set of biologically relevant zones (regions) defined in molecule distribution profiles. I consider every profile as a rich information source on a current biological state. Figure 4.4.8 (a) illustrates an example of an image with a structural marker (labeled green) and a suitable zone model created after examination of numerous extracted profiles. An example position of an intensity profile is marked by a white line in the initial image, and the respective extracted profile is shown below. The cell regions close to the structural marker (Zones 2 and 4) may contain different concentrations of the functional marker (labeled red) and may have different physical and biological properties than the regions further away from the boundary surface (Zones 3 and 5). All these zones are also completely different from the membrane itself (Zone 1). Therefore, it seems intuitive to subdivide a complete intensity profile into such small zones.

To develop a zone model for a particular biological process, extracted profiles have to be analyzed taking the biological background into account. For example, the smaller the structures studied, the smaller should be the created zones. Similarly, the stronger the changes in the functional marker concentration depending on a distance from the membrane, the shorter should



(a) (b)

Figure 4.4.8.: Multi-marker images and corresponding zone models for structural marker profiles. (a) Membrane protein; (b) nucleus marker (labeled green).

be the zones.

The position of the structural marker in a distribution profile indicates the membrane to be crossed by the functional marker. It has to be identified in every profile. Peak detection on a profile yields a position of the membrane, or its center. This position is also a starting point to count the distance from the membrane. After this position is identified (e.g. as the center of the profile), a specific zone model can be applied.

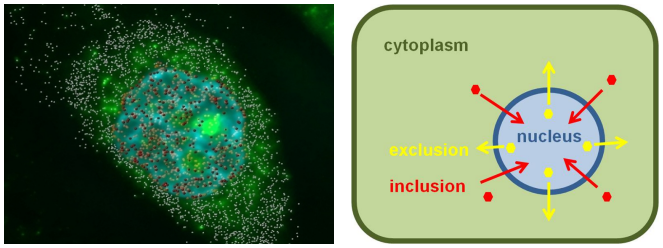
After having created a specific zone model, suitable descriptors can be developed by combining intensity values from different zones. Let us remember that the zone model as well the descriptor formulas have been elaborated based on the structural marker profiles. However, all the following evaluations will be performed on the functional marker profiles, as they are the information source on the potential translocation of biomolecules. Descriptor values are calculated from the functional marker profiles and are further evaluated.

Distributions of the descriptor values from different samples can be compared by a statistical test to detect changes. The selected test has to be unpaired and should make no assumptions on the *a-priori* distributions of the descriptor variables. Resulting p-values will show whether any statistically significant difference is detected or not. As numeric descriptors represent the extracted profiles of the functional marker, one concludes that the detected changes apply to the studied distribution of a functional marker. The translocation can thus be assessed quantitatively.

The calculated significance may also be used to identify which descriptor is more relevant for the biological process studied. The upcoming chapters will show that the introduction of structure-specific descriptors can increase significance levels of the detected translocation.

4.5. Quantification of other translocation scenarios

The optimal 1D transport model is used as the basis to define the structure-based and the random lines-based workflows. Numerous biological transport processes correspond to this scheme. However, another translocation type is also frequently found in living organisms, namely, inclusion/exclusion processes. There, small particles are absorbed by a cell or an organelle, or these particles are exported from it.



(a) (b)

Figure 4.5.1.: (a) Nucleus (from Bitplane [11]) depicted in 2D to illustrate 2D translocation (b).

Inclusion/exclusion processes also involve a membrane structure that can serve as a reference structure for translocation assessment. It can be either a membrane, or a complete biological structure itself. For example, a nucleus plays this role in the earlier presented glucokinase translocation (2D-Nuc process, see Section 3.1.4). There, the functional marker (glucokinase) is translocated into closed regions (nuclei) delineated by the structural marker (DNA). I will further refer to such processes as to *2D translocation*. 2D translocation exhibits 3D properties, as any biological event. However, when captured in 2D multi-marker images, the translocation may be depicted as two-dimensional (see Figure 4.5.1).

Quantification of a 2D translocation can be performed according to the defined structure-based and the random lines-based workflows, similar to the processes corresponding to the 1D optimal transport model. Only structure detection in profiles will proceed somewhat differently. In the case of 2D translocation, a structural marker does not delineate a membrane, but a complete structure within a cell. Hence, peak detection on such a structural marker profile will not yield one central point, but two borders of the structure. Figure 4.4.8, (b) illustrates a case where DNA is a structural marker and a functional marker is translocated into/from the nucleus. One of the possible profile positions is marked by a white line. Peak detection on the respective structural marker profile will yield a region corresponding to the labeled nucleus. Optimal transport trajectories will also correspond to the concentration gradients and be orthogonal to the boundary surfaces (membranes). Therefore, both translocation quantification workflows are also applicable here. A suitable zone model and numeric descriptors will be developed keeping in mind a different profile structure.

4.6. Limitations of the approach

The suggested approach to the translocation quantification covers a great number of transport processes. However, there are some process types which are not likely to be analyzed by this workflow.

First, we are limited to directed transportation. Stochastic molecule motion is thus not covered. Second, both a functional and a structural marker need to be taggable and detectable in biological images. If this is impossible, my workflow can not be applied. Third, if an actual translocation trajectory is very different from the optimal one (the concentration gradient), then the extracted intensity profiles might not be representative for the changes in the molecule distribution. Consequently, calculated descriptors might not be representative for the translocation progress, and no reliable statistical measures can be obtained.

I believe that the novel translocation quantification should be applicable to a wide range of biological transport processes.

5. Implementation of automated translocation quantification workflows

The preceding chapter summarizes ideas on translocation quantification. It includes general considerations about biological transport processes, in particular those corresponding to the *optimal 1D transport model* (directed transport orthogonally to a membrane). I have defined two workflows for the automated quantification of translocation processes corresponding to this model.

One quantification workflow is based on structure detection in images (*structure-based workflow*). There, distribution profiles of biomolecules are extracted by line densitometry orthogonally to the structures detected. Extracted profiles undergo a selection procedure in order to filter out noisy profiles and those depicting damaged biological structures. Distribution of biomolecules is further *toponomically characterized* in relation to the structures detected. Process-specific numeric descriptors are developed for this.

Due to the complexity of structure detection in biological images, I propose another automated workflow that avoids this step. This *random lines-based workflow* starts with profile extraction along random lines. Further, structure detection is performed on the extracted *random profiles* and not in the image. The random profiles also undergo a selection procedure. And finally, distribution of biomolecules is *toponomically characterized* as in the *structure-based workflow*.

The *random lines-based workflow* can be more easily applied to different biological structures than the *structure-based workflow*, owing to the absence of structure detection in images. In this Section, I will present implementation of the *structure-based* and the *random lines-based workflows* for the model biological processes.

Three translocation processes are used as validation models. Two processes correspond to the *optimal 1D transport model* defined in Chapter 4. The first process (1D-Can) is the translocation of bile salt export pump (Bsep) between the canalicular membrane and the cytoplasm of hepatocytes. The second process (1D-Bas) is the translocation of sodium taurocholate cotransporting polypeptide (Ntcp) between basolateral membrane and the cytoplasm of hepatocytes. The third process is a 2D translocation of glucokinase between the nuclei and the cytoplasm of hepatocytes (2D-Nuc process). These biological processes are investigated by my collaboration partners in the DFG-funded clinical research group "Hepatobiliary transport and liver diseases". Owing to this cooperation, I had examples of translocations corresponding to the defined 1D and 2D models and could test my theoretical considerations.

In order to test the validity of my theory, the automated workflows are implemented in a set of algorithms. **Application I** describes the implementation of the *structure-based workflow* for the 1D-Can process (see Table 5.0.1). **Application II** is the implementation of the *structure-based workflow* for the 1D-Bas process. **Application I** and **II** are validated against the manual translocation quantification (*manual workflow*) taken as the standard method (Section 3.2). Statistical tests on the automatically and manually extracted data are expected to indicate a significant functional marker translocation between positive and negative controls (p-value ≤ 0.05) but not between negative controls. Descriptor values, as parameters of the molecule distribution function, are expected to exhibit similar distribution in the manually and automatically extracted data.

Further, the *random lines-based workflow* is implemented for the 1D-Can and 1D-Bas processes (**Application III** and **IV**). **Application III** and **IV** are validated against the *structure-based workflow*. These applications are not validated against the *manual workflow*, because not

Model Process	Translocation quantification		
	Standard method	Structure-based	Random lines-based
1D-Can	Manual	Application I	Application III
1D-Bas	Manual	Application II	Application VI
2D-Nuc	Structure-based		Application V

Table 5.0.1.: Experimental design.

all datasets have been manually processed by my colleagues. Significance levels of the detected translocation and distribution of descriptor values are considered.

And finally, the *random lines-based workflow* is adapted to quantify a 2D translocation on the example of the 2D-Nuc process (**Application V**). There, nucleus-to-cytoplasm intensity ratios are calculated to describe the functional marker distribution. **Application V** is validated against the standard *structure-based* quantification of the nucleus-to-cytoplasm translocation (Section 3.3). Distribution of the nucleus-to-cytoplasm intensity ratios calculated by these two automated workflows are compared.

The upcoming sections include detailed descriptions of the developed algorithms. Statistical results of the developed automated workflows are presented in Chapter 6 together with their validation.

5.1. General steps for the structure-based workflow

I will first focus on 1D cases illustrated by protein translocations at the canalicular and the basolateral membranes of hepatocytes (1D-Can and 1D-Bas processes). The common processing steps will be presented in the upcoming sections, followed by case specific implementations described in more detail. Later, quantification of a 2D translocation on the example of glucokinase nucleus-to-cytoplasm translocation (2D-Nuc process) will be discussed.

5.1.1. Structure detection

Figure 5.1.1 illustrates the first processing steps on the example of a confocal fluorescence microscopy image of the canalicular membranes. According to the structure-based workflow, structure detection is performed first. Numerous algorithms exist and can be applied for this purpose (see Section 2.2.1). Some of them are generally applicable, some are optimized to a certain type of biological images. I perform the structure detection (Figure 5.1.1, (b)) using the software Zeta, presented in Section 2.3.1. The foreground-background detection is implemented there using machine learning trainable by an expert.

The yellow circles in Figure 5.1.1 point out small foreground fragments which are deleted in a consecutive step. The detected foreground regions (white) correspond to membranes in the input multi-marker images. They can be refined and cleaned from noise by simple image processing. Morphological opening [153] deletes tiny objects which might have been created by the foreground-background detection (Figure 5.1.1, (c)). Small objects are further deleted from the foreground mask to exclude potentially damaged or incomplete structures (Figure 5.1.1, (d)). Subsequent morphological closing (Figure 5.1.1, (e)) eliminates gaps which were possibly introduced by morphological opening. The final cleaning eliminates left over small objects (Figure 5.1.1, (f)).

The next step of the structure-based workflow is the profile extraction in relation to the structures detected (*structure-based profile extraction*). According to the optimal 1D transport model, molecules are translocated orthogonally to a membrane. To find these directions, the orientation of the membranes has to be analyzed. This can be done by skeletonization of foreground regions (Figure 5.1.2, (a)). Among various algorithms, skeletonization according to the Euler characteristics [114] is one of the least complicated implementations. I have selected

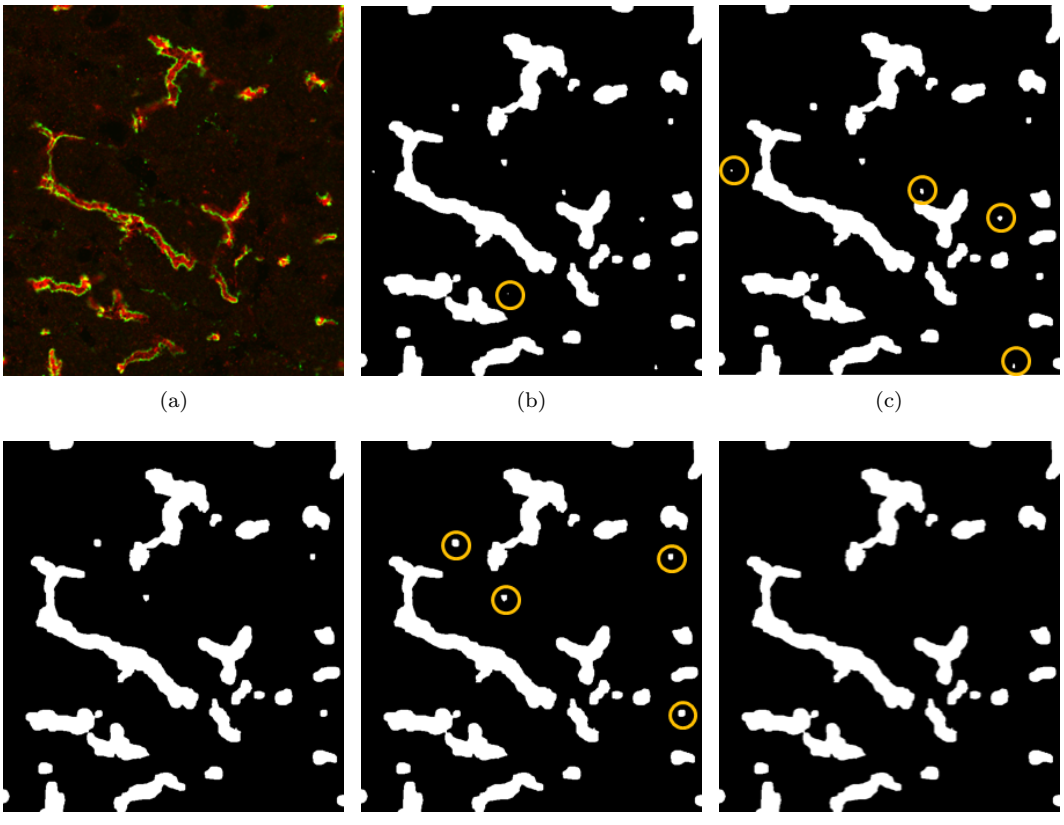


Figure 5.1.1.: Canalicular membranes: structure detection steps I. (a) Initial image; (b) detected foreground regions (white); (c) morphological opening; (d) thresholding; (e) morphological closing; (f) thresholding.

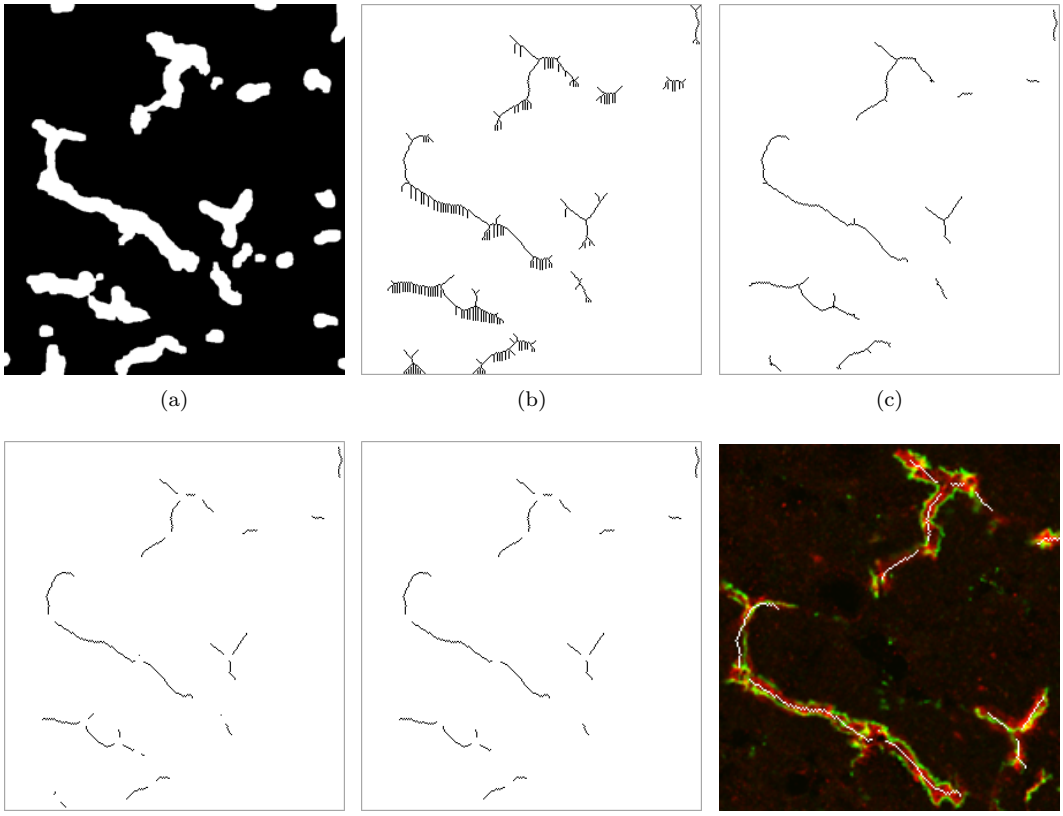
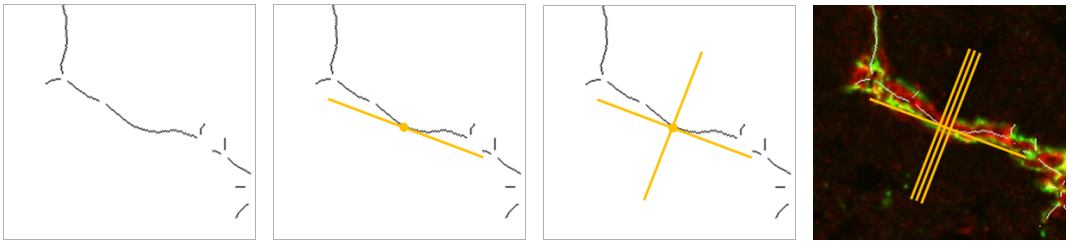


Figure 5.1.2.: Canalicular membranes: structure detection steps II. (a) Cleaned foreground regions; (b) skeletonization; (c) pruning; (d) deletion of branching points and their neighbors; (e) thresholding; (f) resulting skeleton fragments in the initial image.

this algorithm, as the obtained skeletons (Figure 5.1.2, (b)) can be optimized later. For example, pruning deletes short branches that are attached to the main skeleton line. Pixels of the shorter branches are deleted one by one from four directions until no further deletion is possible. Only those parts of the skeleton are left that extend in the direction of the long axis of the membrane segments (Figure 5.1.2, (c)). Further cleaning of the skeleton is motivated by the manual strategy of selecting membrane segments for densitometric analysis. Only long, clean and unbranched membrane segments are considered suitable. Therefore, small fragments and branching points with several their neighbors are deleted from the skeleton (Figure 5.1.2, (d) - (e)). These procedures result in big and straight skeleton fragments, as Figure 5.1.2, (f) illustrates. These skeleton fragments indicate directions of the membranes, orthogonally to which distribution profiles will be extracted.

5.1.2. Distribution profiling

Structure-based profiles are extracted both for the functional and structural marker orthogonally to the membranes (see Section 4.3). Therefore, at every pixel of the skeleton (Figure 5.1.3, (a)) the following operations are performed. First, a tangent is fitted to the skeleton line at this pixel (Figure 5.1.3, (b)), so that the direction of this particular membrane segment is identified. Then, an orthogonal line is drawn through this pixel indicating the optimal translocation trajectory. This line is centered at the membrane, as both the canalicular and the basolateral



(a) (b) (c) (d)

Figure 5.1.3.: Distribution profiling illustrated on canalicular membranes. (a) Resulting skeleton fragments; (b) fitted tangent; (c) line orthogonal to the skeleton, along which intensity profiles are extracted; (d) initial image with the lines allowing extraction of a wider profile.

membranes separate cells of similar sizes. The extracted profiles should thus be symmetrical and spread equally to both sides of the membrane (Figure 5.1.3, (c)). Along this optimal translocation trajectory, pixel intensities are extracted and recorded.

Similar to Schmitt et al. [142], the width of the profile is increased by adding intensity from neighboring parallel lines (Figure 5.1.3, (d)). The average of several such intensity profiles represents a wider profile extracted at the given pixel.

Only a subset of structure-based profiles cleanly represents the translocation phenomenon. Profile selection allows elimination of noisy profiles and those taken at damaged regions of the biological structures. Selection criteria depend on the knowledge about the biological structure and will be therefore presented in the upcoming case specific sections.

5.1.3. Toponomic characterization

The usual evaluation scenario includes 2 datasets of n images (e.g. $n = 10$) acquired from biological samples. These might be, for example, the rat livers perfused in a normo-osmolar or hyper-osmolar buffer. Based on the available multi-marker images, functional marker distribution has to be evaluated, and its translocation has to be quantified. Therefore, selected distribution profiles are used to calculate process-specific numeric descriptors. The upcoming sections include details for the model 1D-Can and 1D-Bas processes. Here, only common steps for the toponomic characterization are presented.

Numeric descriptors represent the extracted distribution profiles, and thus the unknown molecule distribution function. Descriptor values calculated for different datasets are compared by Wilcoxon rank sum test (WRST) (using the software R [132]). The resulting p-values indicate a statistical significance of the detected translocation of the functional marker. The newly defined numeric descriptors are also compared to the statistical variance which was used for the translocation quantification in previous publications. Therefore, the datasets of images are compared by WRSTs on the statistical variance calculated for the selected profiles. Significance of the translocation detected on the statistical variance is compared to that calculated using the numeric descriptors. The higher significance levels suggest the better relevance to the translocation phenomenon.

WRST is chosen because its properties suit well to the data to be evaluated. First, it does not make any assumption on variables distribution. Second, it can be unpaired. As compared datasets and the respective extracted profiles are independent, the applied statistical test has to be unpaired. Furthermore, the number of the profiles extracted and/or selected from the dataset is unknown beforehand. WRST is advantageous, as it can compare populations of different sizes. However, it is sensitive to the population size. Namely, the larger the compared groups, the more significant differences can be detected. I introduce a following scheme to permit a consistent evaluation:

1. Extract and select profiles from the datasets to be compared.
2. Randomly choose 100 profiles for each of the datasets.
3. Calculate descriptor values for these 100 profiles.
4. Perform a WRST on 100 vs. 100 descriptor values.
5. Save the resulting p-value.
6. Repeat steps (2)-(5) 100 times.
7. Report the median p-value from the 100 performed evaluations.

This evaluation scheme enables a consistent comparison of the structure-based and the manual workflows. First, the group size of 100 profiles is selected according to the established manual workflow. Experts manually extract 10 profiles/image. A standard dataset consists of 10 images, resulting in 100 manually extracted profiles. Second, WRST is sensitive to the population sizes. Thus, for the statistically correct comparison of the manual and the automated workflows, the group sizes of the compared profiles have to be equal (e.g. 100 profiles/dataset). Third, the automated profile extraction normally yields much more than 10 profiles/image. Thus, random sampling of 100 profiles from all the profiles extracted from a particular dataset is always possible.

All the statistical results presented in the upcoming sections are generated according to this scheme.

5.2. Validation of the novel automated workflows

The novel automated workflows have to be validated against the established approaches. The structure-based workflow is validated against the manual workflow, while the random lines-based workflow is validated against the structure-based workflow.

For the validation of the structure-based workflow, datasets of images are automatically processed, profiles are extracted, selected and ranked. In parallel, the same images undergo the manual profile extraction, as presented in Section 3.2.

Further, biologically relevant zones are defined for a particular translocation process. Numeric descriptors are developed based on these zones. Numeric descriptors are calculated both from the automatically and manually extracted profiles and used for the comparison of the datasets by WRSTs. Resulting p-values indicate the significance of the detected translocation phenomenon.

The manual workflow is performed as a standard evaluation. It should neither indicate a false positive effect when comparing negative controls, nor a significant translocation may be missed between positive and negative controls. The results of the structure-based workflow are considered valid if a significant translocation is detected between positive and negative controls and its absence is indicated in the negative controls.

The structure-based workflow is also validated against the manual workflow by descriptor values. Descriptor values are numeric parameters of the extracted profiles. The extracted profiles, in turn, represent the unknown molecule distribution function (see Section 4.4.4). If descriptors calculated from the manually and automatically extracted profiles exhibit comparable distributions, the original molecule distribution functions will also be comparable. Herewith, the automatically extracted distribution profiles may be considered a valid alternative to the manually extracted profiles to represent the molecule distribution. To test this, means (μ) and standard deviations (σ) of the descriptor values are compared.

The random lines-based workflow is validated the structure-based workflow similarly.

5.3. Application I: structure-based workflow for 1D-Can process

Two biological processes corresponding to the defined 1D optimal transport model are studied in the scope of this thesis. The first one concerns the bile salt export pump (Bsep) internalization into hepatocytes from the canalicular membranes (1D-Can process, for biological background see Section 3.1.2). Bsep is a functional marker whose translocation is to be quantified. Zonula occludens 1 protein (Zo-1) is the structural marker, as it delineates canalicular tight junctions and is visible under various experimental conditions. These proteins are stained in the liver tissue sections. Confocal fluorescence microscopy images are acquired and evaluated according to the structure-based workflow.

5.3.1. Structure detection

Structure detection is performed according to the presented scheme (see Section 5.1.1). Here, only specific details are given.

Foreground regions detected by the Zeta software are first cleaned by morphological opening using a round structuring element with the radius of 3 pixels. Resulting objects are thresholded to eliminate those whose area is too small (see Table 5.3.1). These regions most probably correspond to wrongly segmented noise. The respective threshold as well as all other referenced later, is selected based on detailed examination of numerous confocal images and tests of various thresholds. The images used for the development and tests of the algorithms have a resolution of 100 nm/pixel.

Subsequently, morphological closing is performed with the same structuring element. Areas of the resulting foreground regions are again thresholded. This time, bigger objects are eliminated which may correspond to very small canaliculi fragments. After skeletonization, branching points and their 2 neighbors are deleted. The rest of the skeleton fragments is further thresholded.

Processing of images with different resolution is enabled by scaling of thresholds and other parameters (e.g. profile length). A scaling factor is calculated and is further used to scale the parameters:

$$factor = \frac{100.0}{pixelSize}. \tag{5.3.1}$$

5.3.2. Profile extraction

Profile extraction is performed for both the functional and the structural marker at every position of the selected skeleton fragments (see Section 5.1.2). The length of the extracted profiles depends on the size of the biological structures studied. An empirical image analysis shows that canalicular diameter varies in the range of 0.8 – 2.5 μm. Apart from the canaliculus, extracted profiles should also cover significant part of the cytoplasm of the adjacent hepatocytes to illustrate the Bsep translocation between them. Therefore, the selected profile length should be > 2.5 μm, but small enough so that no further hepatocytes are covered by one profile.

Operation	Threshold, pixel
Cleaning after morphological opening	125
Cleaning after morphological closing	250
Delete number of branching points neighbors	2
Cleaning after branching points deletion	7

Table 5.3.1.: Algorithm parameters for the 1D-Can process.

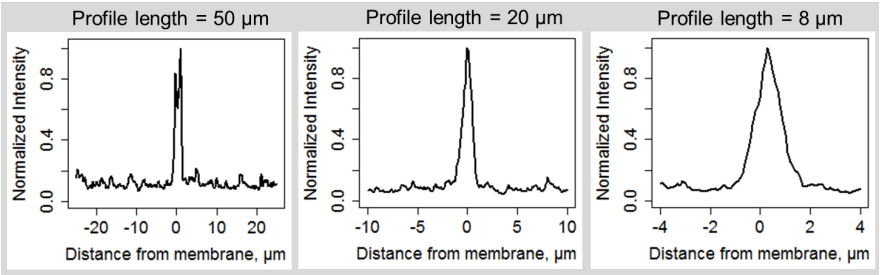
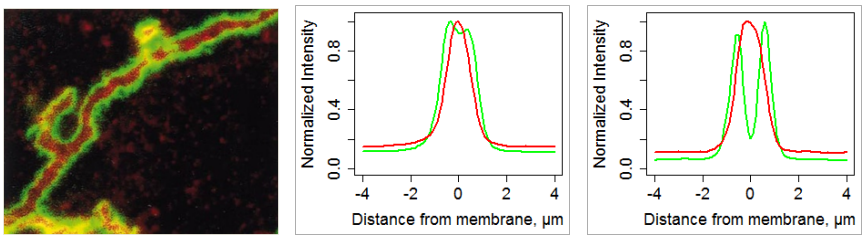


Figure 5.3.1.: Selection of the optimal profile length for quantification of the 1D-Can process. Profile length 8.0 μm is found to be sufficient to cover Bsep distribution changes.



(a) (b) (c)

Figure 5.3.2.: Motivation of profile selection. (a) Typical fragment of a canalicular membrane; (b) plots of the averages over all extracted profiles (red - Bsep, green - Zo-1); (c) expected plots of the average profiles.

Rat hepatocytes have a diameter of 20 – 30 μm [169], so that a profile of 50 μm covers approximately the fluorescence intensity distribution of one hepatocyte on each side of the canaliculus. A range of profile lengths from 6.0 to 50.0 μm was evaluated. It was found that the intensity of Bsep stays nearly constant for distances of more than 4.0 μm from the central point of the canalicular membrane. Consequently, the smallest length is selected which is sufficient to illustrate the Bsep concentration variation in the cytoplasm (see Figure 5.3.1). It is the length of 8.0 μm, corresponding to 81 pixels (nearest odd value) for the image resolution 100 nm/pixel.

5.3.3. Profile selection

Human experts evaluate only intact and representative membrane fragments in the manual analysis. Automated profile extraction results in a large number of profiles possibly extracted on such positions which would be discarded by an expert. Hence, only a subset of the extracted intensity profiles represents the translocation phenomenon. Therefore, the extracted profiles have to undergo a selection procedure to be considered for further translocation quantification. This is explained on the following example. Numerous profiles can be extracted from every image (Figure 5.3.2, (a)). Their average is calculated to illustrate the molecule distribution in a particular biological sample. However, the average over all extracted Zo-1 profiles does not exhibit two clear intensity peaks corresponding to the tight junctions (green line in Figure 5.3.2, (b)). The expected shape of a profile is illustrated in Figure 5.3.2, (c), exhibiting two symmetrical peaks of Zo-1 intensity. This suggests that some profiles included in the initial averaging might have been extracted at damaged canalicular membranes. Hence, the profile selection is required to filter them out.

Knowledge of the structures studied (e.g. size, symmetry) is required to implement a selection strategy. As presented before, profile selection is performed on the structural marker profiles.

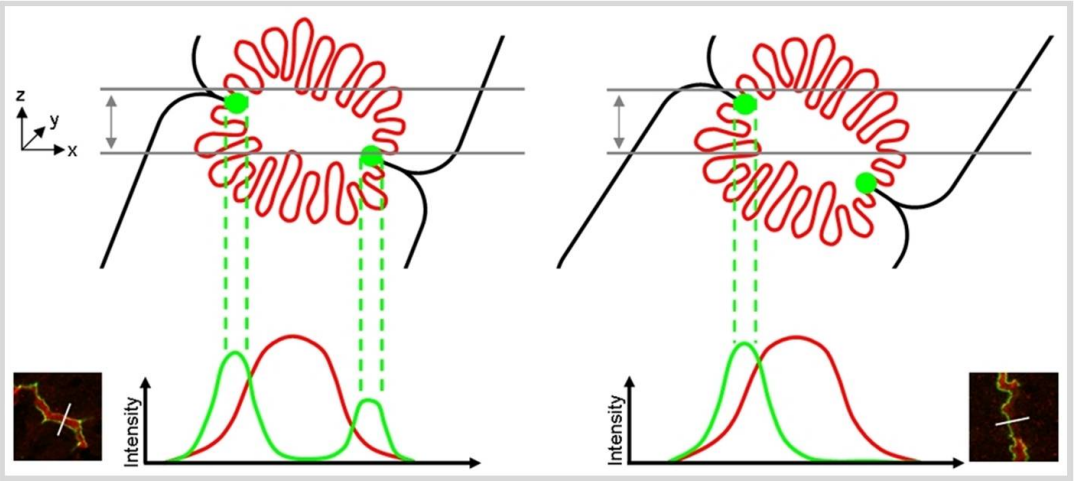


Figure 5.3.3.: Optical section of a canaliculus. If a canaliculus is not cut by an optical image plane in its principal axis, then its tight junctions will not be imaged symmetrically, leading to unsymmetrical Zo-1 intensity profiles.

Zo-1 intensity profiles are tested according to the four conditions described below.

First, membrane segments non-parallel to the focal image plane are filtered out. Such profiles exhibit Zo-1 intensity peaks of unequal height, indicating that the image plane does not match the orientation of the canaliculus (see Figure 5.3.3). The first condition restricts the height difference between the two largest local maxima. The relative difference between the peaks (5.3.2) should not exceed an empirically identified threshold of 15%:

$$\frac{maxValue_1 - maxValue_2}{maxValue_2} < 0.15, \tag{5.3.2}$$

where $maxValue_1$ is the absolute maximum of the Zo-1 intensity profile, $maxValue_2$ is a height of the second local maximum. The second condition (5.3.3) constrains the distance between the two intensity peaks (tight junctions) corresponding to the canalicular diameter. It is known to vary in the range 0.8 – 2.5 μm:

$$0.7 < \|(positionMax_1 - positionMax_2)\| < 2.6, \tag{5.3.3}$$

where $positionMax_1$ and $positionMax_2$ denote positions of the peaks (their coordinates μm) on the profile. The third condition eliminates intensity profiles where the local minimum between the two peaks is not low enough (low contrast). A small intensity difference between Zo-1 peaks and the valley might be caused by a damaged tissue region. It may also indicate that the focal plane is above or below the tight junctions, resulting in a low Zo-1 intensity. Thus, the third criterion (5.3.4) filters out profiles where a valley is not deep enough:

$$\frac{minValue}{maxValue_2} < 0.2, \tag{5.3.4}$$

where *minValue* defines the intensity of the local minimum between the considered two local maxima. And finally, high quality Zo-1 profiles are selected and noisy profiles are eliminated. The former ones exhibit two prominent intensity peaks separated by a low local minimum and flat tails on the sides. The latter ones, in contrast, show further local maxima:

$$\frac{P_{60}(localmaxima)}{maxValue_2} < 0.4, \tag{5.3.5}$$

where P_{60} *(local maxima)* is the 60^{th} percentile of the local maxima values. Only if all four

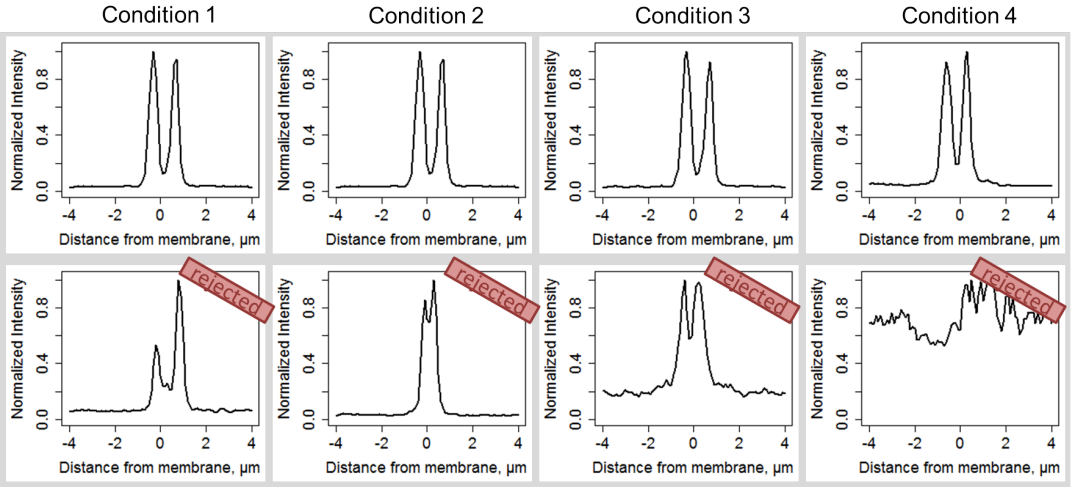


Figure 5.3.4.: Examples of selected and rejected structural marker (Zo-1) profiles at canalicular membranes.

conditions are fulfilled (illustrated again in Figure 5.3.4), a particular Zo-1 intensity profile is considered valid. The respective Bsep intensity profile is extracted at the same position of the skeleton and is passed to further analysis.

5.3.4. Profile ranking

Not all the selected profiles are equally relevant to the quantified translocation. Some of them were extracted at more symmetrical membranes, or those with brighter staining. Furthermore, the number of selected profiles is usually much larger than the number of the profiles obtained manually. The quality of the automatically extracted profiles can be ranked, and only the best profiles can be used for the statistical evaluation. The ranking works on the structural marker only, so it does not introduce a bias to the functional descriptor.

The parameters of the selection process are also used for the ranking procedure. The first selection criterion (difference of peak intensities, see Equation 5.3.2) is used for the ranking r_1. The third parameter (peak - valley contrast, see Equation 5.3.4) leads to the ranking r_2, defined as the ratio of the central valley height to the peaks height. The smaller the values of these criteria, the higher the ranks assigned. An unweighted combination of r_1 and r_2 is calculated as the final ranking. According to this final ranking, the top n profiles (n = 10, 20, 30, etc.) are taken for further analysis.

5.3.5. Profile normalization

Functional marker profiles can be normalized for the quantitative analysis in several ways. First, protein distribution profiles can be normalized to a standard distance between peaks. However, in the samples I have analyzed so far, variations of the canalicular width are sufficiently small to not require such a scaling. Second, absolute intensity values of the profiles can be scaled to a defined range (e.g. 0 - 1). Relative descriptors, presented in the next section, obviate the need for such a normalization. And finally, the lateral coordinate systems of profiles can be centered, compensating for skeleton lines not centered in the canaliculus. This might be caused by the skeletonization algorithm, as it uses only approximate foreground regions whose borders are not necessarily symmetric with the Zo-1 maxima. However, zones are defined for the descriptor calculation individually in each profile. Therefore, centering of the lateral coordinates will not influence the defined zones. Thus, no profile normalization is required in my workflows.

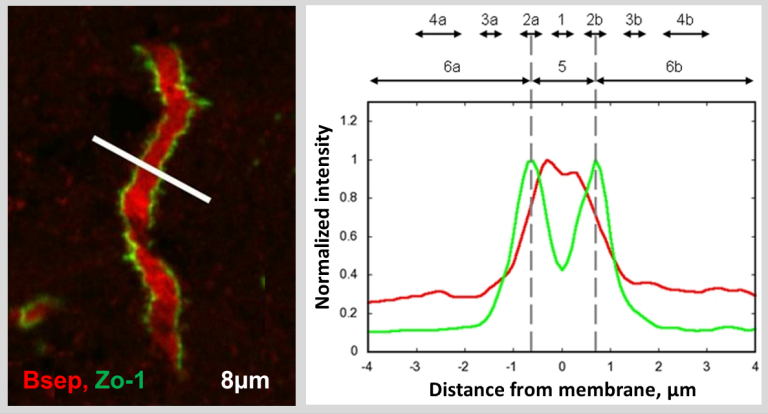


Figure 5.3.5.: Typical canalicular fragment and a zone model for the 1D-Can process.

5.3.6. Toponomic characterization

Profile zones

The selected intensity profiles are used to assess the translocation of the functional marker (Bsep). According to the structure-based workflow (see Section 4.4.4), biologically relevant zones have to be identified. Figure 5.3.5 illustrates a typical canaliculus fragment and the suggested zone model. The zone model is developed using structural marker (Zo-1) profiles.

Every canaliculus is delineated by tight junctions, corresponding to the intensity peaks in the Zo-1 profiles. First, these two peaks are identified, and the center of the profile is found. The part of the profile between the Zo-1 peaks (tight junctions) is considered to be the interior of the canaliculus, while the parts outside the tight junctions correspond to cytoplasm of the adjacent hepatocytes.

Following zones are defined for the accepted Zo-1 profiles. Zone 1 has a length of 0.5 μm, and is centered between the peaks. The integral Bsep intensity of zone 1 represents the amount of Bsep in the center of the canaliculus. Zones 2a and 2b (0.5 μm each) are centered on the peaks corresponding to tight junctions. Their integral intensity indicates protein concentration in the canalicular membrane. The length of 0.5 μm is chosen according to the empirical determination of canalicular width (0.8 – 2.5 μm). It allows covering of the whole canaliculus even if it is very narrow (at the lowest possible width border of 0.8 μm). Even in this case zones 1, 2a and 2b will not overlap. Zone lengths are fixed, so that descriptors calculated from zonal integral intensities can be compared between different images.

Intracellular zones are defined to analyze the internalization of the functional marker. Zones 3a and 3b (0.5 μm each) describe intracellular fluorescence intensities close to the canalicular membrane. Zones 3a and 3b are spatially separated from the tight junctions and start 0.5 μm away from the respective ends of the zones 2a and 2b. Thus, they do not cover regions with a high potential variation. Zones 4a and 4b are situated closer to the centers of hepatocytes and measure 1.0 μm each. They are also spatially separated by 0.5 μm from the zones 3a and 3b. Given the image resolution of 100 nm/pixel, 0.5 μm translates to 5 pixels. This length is scaled respectively if images with different pixel size are processed.

The last zones (5, 6a and 6b) do not have a fixed length. Instead, their length is defined for each profile individually. Zone 5 covers the fluorescence intensities in the interior of the canaliculus between the peaks. Zones 6a and 6b combine the intracellular intensities of the profile from the maximum values of the peaks to the ends. These zones are introduced to calculate a relative amount of the functional marker in the canaliculus to its content in the region covered by the whole profile.

Descriptors

Numerical descriptors are developed to quantify the 1D-Can process using the zones identified in Zo-1 profiles. Relative descriptors of the protein distribution are preferred, as they avoid absolute quantity measurements between immunofluorescence images. The following descriptors are defined:

$$X = \frac{sum(Zone1)}{sum(Zone2a + Zone2b)}, \tag{5.3.6}$$

$$Y = \frac{sum(Zone1)}{sum(Zone3a + Zone3b)}, \tag{5.3.7}$$

$$Z = \frac{sum(Zone1)}{sum(Zone4a + Zone4b)}, \tag{5.3.8}$$

$$A = \frac{sum(Zone1)}{sum(Zone2a + Zone2b) + sum(Zone3a + Zone3b)}, \tag{5.3.9}$$

$$B = \frac{sum(Zone1)}{sum(Zone3a + Zone3b) + sum(Zone4a + Zone4b)}, \tag{5.3.10}$$

$$C = \frac{sum(Zone2a + Zone2b)}{sum(Zone3a + Zone3b)}, \tag{5.3.11}$$

$$D = \frac{sum(Zone2a + Zone2b)}{sum(Zone4a + Zone4b)}, \tag{5.3.12}$$

$$E = \frac{sum(Zone2a + Zone2b)}{sum(Zone3a + Zone3b) + sum(Zone4a + Zone4b)}, \tag{5.3.13}$$

$$F = \frac{sum(Zone5)}{sum(Zone5) + sum(Zone6a + Zone6b)}, \tag{5.3.14}$$

where *sum(Zone 1)* corresponds to the integral fluorescence intensity in the zone 1, *sum(Zone 2a)* - in the zone 2a, etc. For example, descriptor X is calculated as a ratio of Bsep fluorescence intensity in the center of the canaliculus relative to Bsep intensity at the tight junctions. Descriptor Y shows the ratio of the central Bsep intensity relative to its intensity in the cytoplasm of the adjacent hepatocytes immediately close to the tight junctions. If Bsep translocation from the canalicular membrane took place, then these descriptor values would decrease.

Descriptors C, D and E relate the Bsep intensity at the tight junctions to various regions in the cytosol. Their values should also decrease under cholestatic conditions. In particular, Bsep internalization affects *sum(Zone2a + Zone2b)* by broadening of the Bsep intensity profile. It also leads to an elevated Bsep fluorescence intensity in the cytoplasm due to its internalization from the canalicular membrane, increasing *sum(Zone3a), sum(Zone3b), sum(Zone4a), and sum(Zone4b)*. As zone lengths are kept constant for all profiles, values of the descriptors can be compared even between the images of different samples.

The last descriptor F (internalization degree) represents a ratio of the Bsep intensity inside the canaliculus relative to the total Bsep intensity covered by the profile. It is thus the only descriptor that does not use zones of the fixed length. Descriptor F will be compared to the other descriptors in order to evaluate the impact of the fixed-length zones on the translocation quantification.

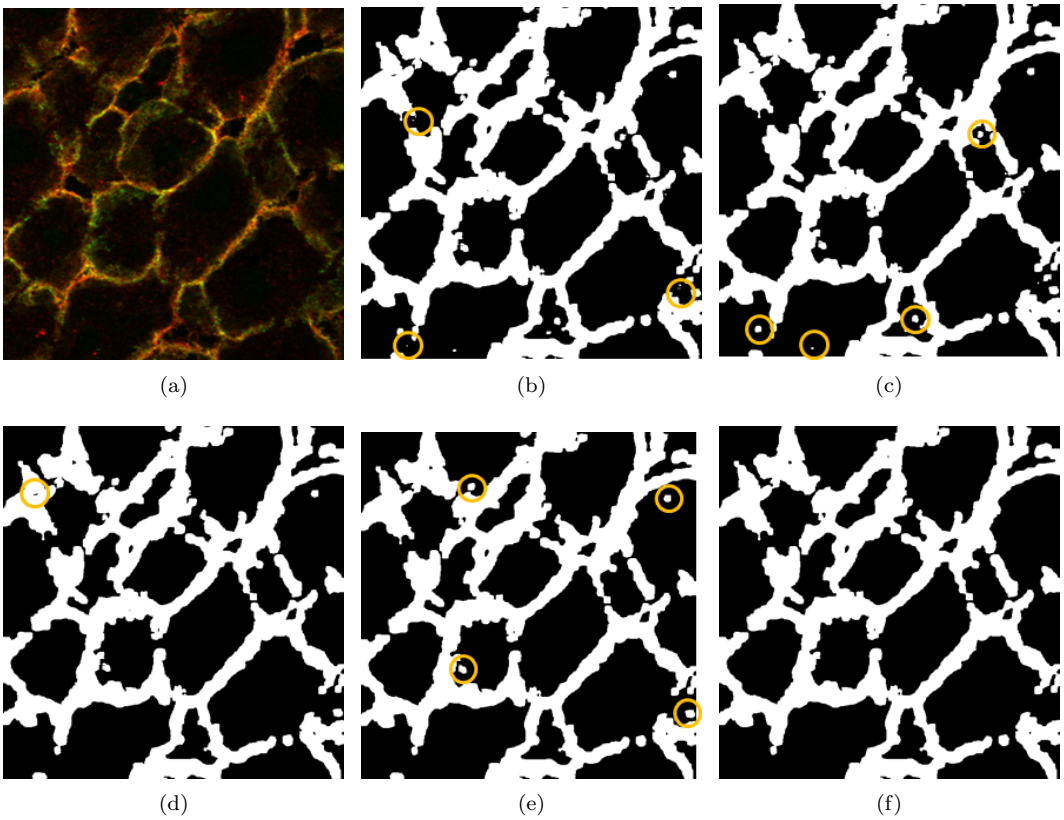


Figure 5.4.1.: Basolateral membranes: structure detection steps I. (a) Initial image; (b) detected foreground regions (white); (c) morphological opening; (d) thresholding; (e) morphological closing; (f) thresholding.

5.4. Application II: structure-based workflow for 1D-Bas process

The second process studied is translocation of Ntcp (sodium taurocholate cotransporting polypeptide) into and from the basolateral membranes in hepatocytes (1D-Bas process). The functional marker (Ntcp) is analyzed in relation to the stable membrane protein Na^+-K^+-ATPase, selected as a structural marker (for biological background see Section 3.1.3). These proteins are immunohistologically stained in liver tissue sections and imaged by confocal fluorescence microscopy.

The structure of the basolateral membrane differs from the canalicular membrane. Therefore, processing parameters have to be adapted, although the same structure-based workflow is applied. In the following, only specific details are given for the 1D-Bas process.

5.4.1. Structure detection

Structure detection is performed according to the scheme presented in Section 5.1.1. It is illustrated here for the basolateral membranes (see Figure 5.4.1).

Foreground regions detected by the Zeta software are first cleaned by morphological opening using a round structuring element of 3 pixels radius. Resulting objects are thresholded and objects having too small area are eliminated (see Table 5.4.1). This threshold, as well as all other referenced later, is selected based on detailed examination of several multi-marker images.

Afterwards, morphological closing is performed with the same structuring element. Areas of

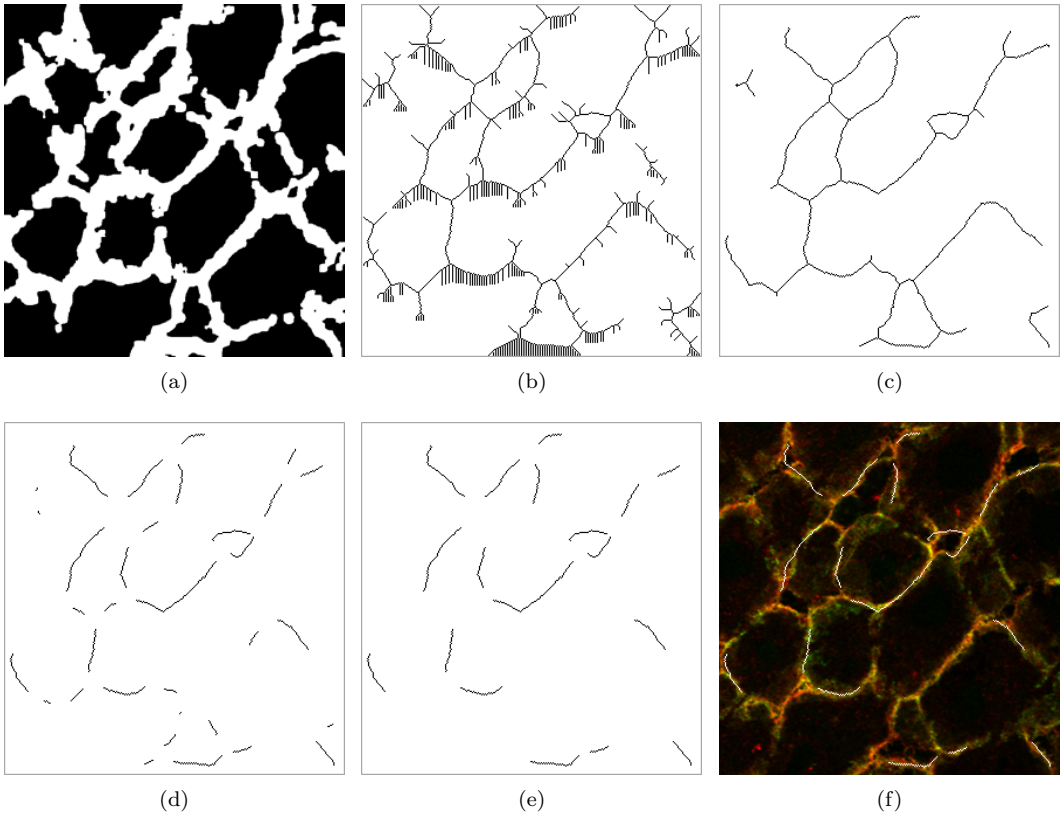


Figure 5.4.2.: Basolateral membranes: structure detection steps II. (a) Cleaned foreground regions; (b) skeletonization; (c) pruning; (d) deletion of branching points and their neighbors; (e) thresholding; (f) resulting skeleton fragments in the initial image.

the resulting foreground regions are again thresholded. This time, bigger objects are eliminated which may correspond to very small basolateral membrane segments. Skeletonization is performed on the resulting foreground regions (see Figure 5.4.2). Subsequent pruning is repeated on the skeleton 4 times to ensure deletion of side branches. Basolateral membranes are long enough so that repeated pruning brings more benefits for structure cleaning than it leads to information loss at the terminals of the skeleton fragments. Subsequently, branching points and their 7 neighbors are deleted. The rest of the skeleton fragments is again thresholded. This thresholding eliminates membrane fragments which are too short for analysis.

Operation	Threshold, pixel
Cleaning after morphological opening	200
Cleaning after morphological closing	400
Delete number of branching points neighbors	7
Cleaning after branching points deletion	15

Table 5.4.1.: Algorithm parameters for the 1D-Bas process.

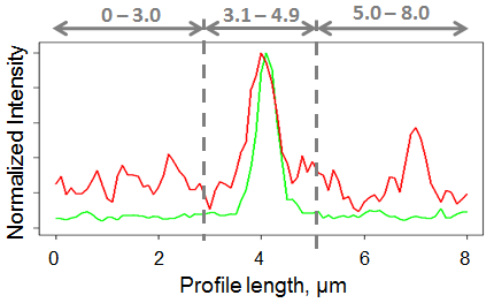


Figure 5.4.3.: Zones for the selection of Na^+-K^+-ATPase profiles.

5.4.2. Profile extraction

Profile extraction is performed according to the same scheme as in case of canalicular membranes (see Sections 5.3.2 and 5.1.2). This illustrates the advantage and general applicability of the structure-based workflow. Although biological structures of the basolateral and the canalicular membranes differ, the same procedures can be applied to extract information on the functional marker distribution.

5.4.3. Profile selection

Three zones are defined for the selection of Na^+-K^+-ATPase profiles (see Figure 5.4.3). The first zone (3.1 - 4.9 μm) should include the intensity peak of the structural marker corresponding to the basolateral membrane. The other two zones (0 - 3.0 and 5.0 - 8.0 μm) cover cytoplasm of the adjacent hepatocytes. No structural marker should be present there, and so these profile parts should be flat.

Based on these zones, following selection criteria are developed. The first criterion rejects those profiles where the peak in the center of the profile is not significantly higher than the tails. In particular, the average intensity in the center (zone 3.1 - 4.9) should be at least twice as large as than the average intensity at the tails:

$$\frac{sum(3.1-4.9)/19}{(sum(0-3.0)+sum(5.0-8.0))/62} \geq 2, \tag{5.4.1}$$

where $sum(3.1-4.9)$, $sum(0-3.0)$ and $sum(5.0-8.0)$ correspond to the integral intensities in the zones 3.1 - 4.9, 0 - 3.0 and 5.0 - 8.0, respectively. The second condition checks whether the tails of the profile are symmetrical enough. Their overall intensities should not differ more than 2.5 times:

$$0.4 \leq \frac{sum(0-3.0)}{sum(5.0-8.0)} \leq 2.5. \tag{5.4.2}$$

The last condition eliminates noisy profiles and those which were extracted at invalid or damaged membrane positions. First, a structural marker profile is processed by median filtering to eliminate high frequency noise. Then, all the local maxima are found and evaluated according to the scheme presented in Figure 5.4.4.

If no maxima are found after median filtering, then it is a straight line. Consequently, it can not be a valid structural marker profile and is filtered out. If exactly one local maximum is detected, then the profile is accepted. If there are further maxima, then they have to be evaluated. In case of a big global maximum and comparatively small second local maximum ($<$ 50%), the profile is accepted. If the second local maximum is $>$ 50% of the global maximum, then the valley (local minimum) between them is evaluated. If the valley is not very deep, the profile is valid. In the other case, the profile is rejected. Some further examples of accepted and rejected profiles are shown in Figure 5.4.5 with the corresponding selection conditions.

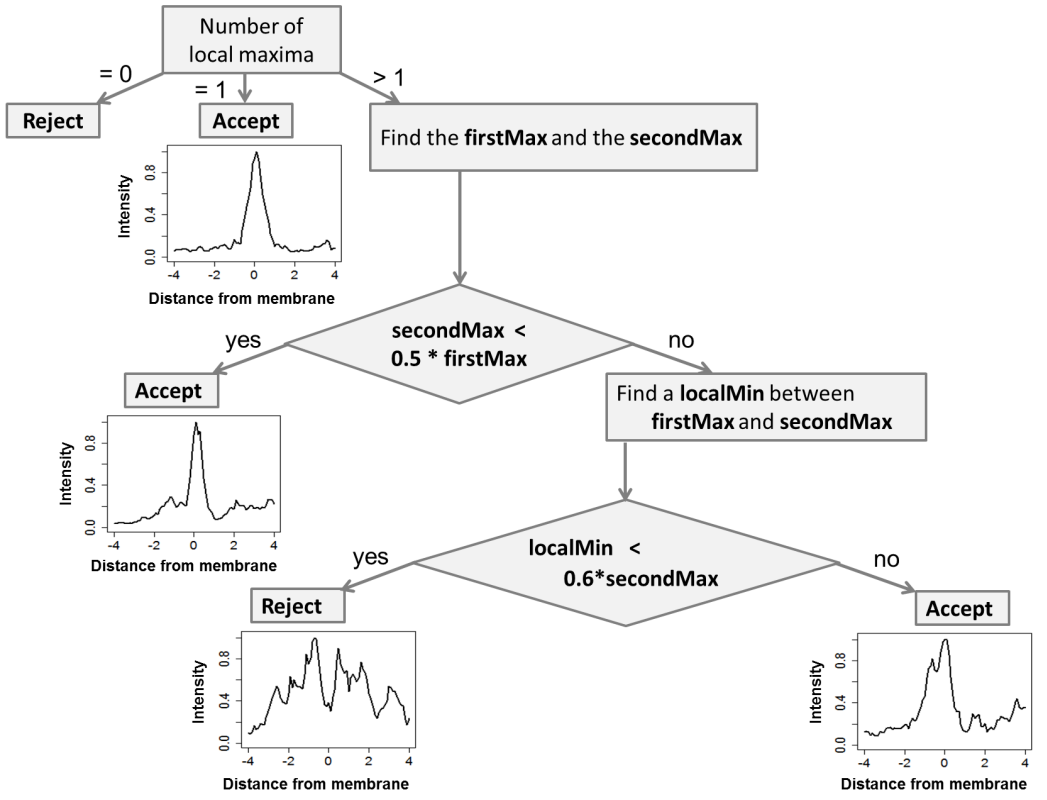


Figure 5.4.4.: Scheme of the selection criterion eliminating noisy and invalid Na^+-K^+-ATPase profiles.

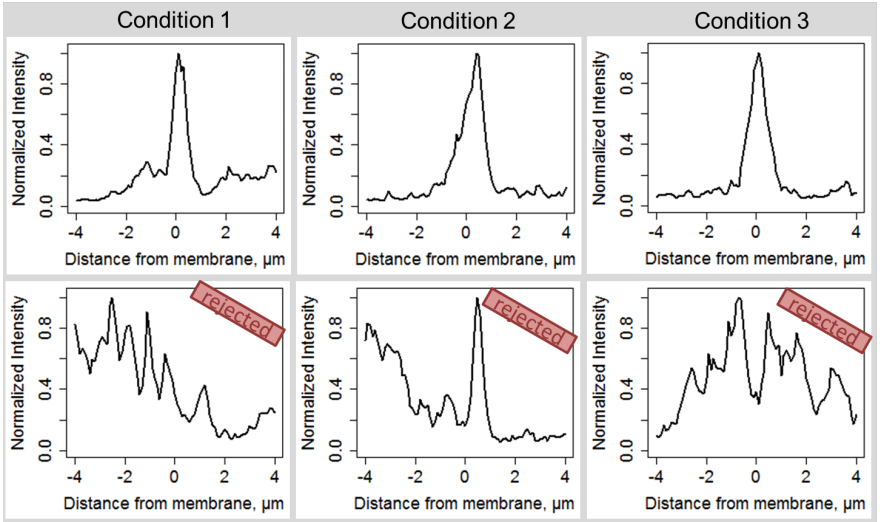


Figure 5.4.5.: Examples of selected and rejected structural marker (Na^+-K^+-ATPase) profiles at basolateral membranes.

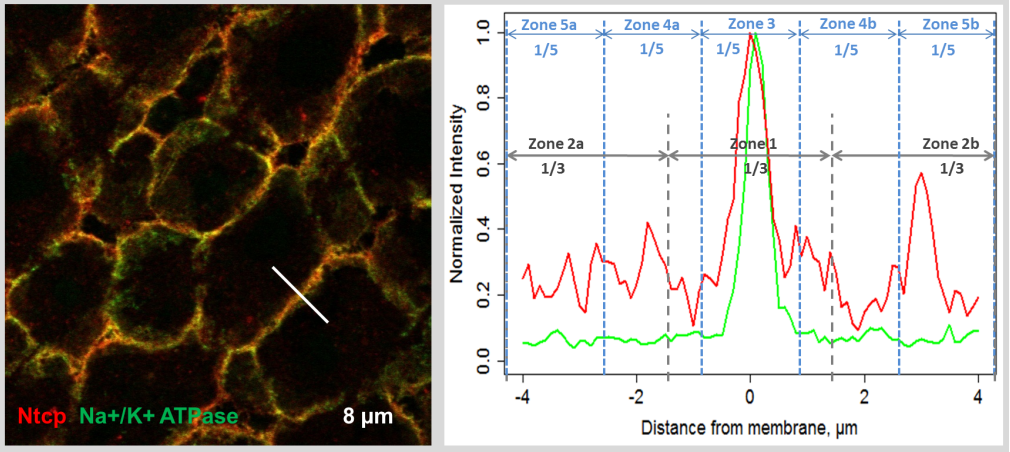


Figure 5.4.6.: Typical basolateral fragment and a zone model the 1D-Bas process.

5.4.4. Profile ranking

Not all the selected profiles are equally representative of the translocation process. The number of selected profiles is also much larger than the number of the manually extracted profiles (n = 10). The quality of the selected profiles can thus be ranked in order to use only the best profiles for translocation quantification.

The first parameter of the selection process is used for ranking. This selection criterion calculates a ratio of the fluorescence intensity in the center of the profile to its intensity at the tails (see Equation 5.4.1). The greater the values of the criterion, the higher the ranks assigned. According the resulting ranks, top n profiles (n = 20, 50, etc.) are taken for further analysis.

5.4.5. Toponomic characterization

Profile zones

The selected profiles are further used for the translocation quantification. A set of biologically relevant zones is defined in the structural marker profiles (see Figure 5.4.6). This zone model is developed specifically for the basolateral membranes (compare to the canalicular zone model in Figure 5.3.5). Zone 1 covers the membrane itself and measures one third of the profile length (2.7 μm). It should include the intensity peak of the structural marker. Zones 2a and 2b also measure 2.7 μm each and cover the cytoplasm of the adjacent hepatocytes. The 1D-Bas process can be characterized by monitoring integral intensities in the zones 1, 2a and 2b.

Finer subdivision of the profiles is performed into smaller zones, each measuring one fifth of the profile length (1.6 μm). Zone 3 covers the central part of the membrane and allows a more precise detection of Ntcp distribution change than zone 1. Zones 4a and 4b cover narrow cytoplasm regions close to the basolateral membrane, while zones 5a and 5b are located closer to the hepatocytes interior. Given the image resolution of 100 nm/pixel, 2.7 and 1.6 μm translate to 27 and 16 pixels, respectively.

Descriptors

Numerical descriptors are developed to quantify the 1D-Bas process. Following relative descriptors of the protein distribution are defined using the zones identified in the Na^+-K^+-ATPase profiles:

$$L = \frac{sum(Zone1)}{sum(Zone2a + Zone2b)}, \tag{5.4.3}$$

$$M = \frac{sum(Zone3)}{sum(Zone4a + Zone4b)}, \tag{5.4.4}$$

$$N = \frac{sum(Zone3)}{sum(Zone5a + Zone5b)}, \tag{5.4.5}$$

$$O = \frac{sum(Zone3)}{sum(Zone4a + Zone4b) + sum(Zone5a + Zone5b)}, \tag{5.4.6}$$

$$P = \frac{sum(Zone4a + Zone4b)}{sum(Zone5a + Zone5b)}, \tag{5.4.7}$$

where *sum(Zone 1)* corresponds to the integral fluorescence intensity in the zone 1, *sum(Zone 2a)* - in the zone 2a, etc. For example, descriptor L is calculated as a ratio of Ntcp fluorescence intensity at the basolateral membrane relative to the Ntcp intensity in the cytoplasm of the adjacent hepatocytes. Values of this descriptor are expected to decrease under cholestatic conditions. Ntcp internalization increases *sum(Zone2a + Zone2b)* and decreases the *sum(Zone 1)* by protein translocation from the membrane into cytoplasm. As the zone lengths are kept constant for all profiles, values of the descriptor can be compared even between images of different experiments.

Other descriptors are based on the shorter zones identified in the profiles, and might lead to more precise translocation quantification. Descriptors M, N and O show the relative Ntcp intensity at the peak of the structural marker (i.e. at the basolateral membrane) to its intensity in the various regions in the cytosol. Descriptor P relates regions of the cytosol to illustrate intra-cellular distribution changes.

5.5. Applications III and IV: Random lines-based workflow for 1D-Can and 1D-Bas processes

The preceding sections cover translocation quantification for the 1D-Can and 1D-Bas processes. They share common processing steps and differ only in the definition of structure-related parameters. Both of them include *structure-based profile extraction*. Alternatively, distribution profiles can be extracted avoiding structure detection images. Such *random profile extraction* is performed along random lines (see Section 4.4.3). Figure 5.5.1 illustrates this for the 1D-Can and 1D-Bas processes.

The profile length (8.0 μm) used is the same for the 1D-Can and 1D-Bas processes (see Section 6.2.5). In turn, original intensity vectors extracted along random lines may be shorter or longer than 8.0 μm. Hence, the procedures applied to the structure-based profiles (selection, ranking, descriptor calculation) have to be modified for these initial vectors. To avoid this, vectors of the length 8.0 μm (*random profiles*) are extracted from the initial vectors that are longer than 8.0 μm. Initial vectors that are shorter than 8.0 μm are ignored.

Random profiles should be extracted without losing any valuable information. All membrane segments that were crossed by the initial random lines have to be depicted in the resulting random profiles. To implement this strategy, random profiles are extracted with overlap.

If an overlap is too short, too many random profiles will depict one and the same membrane fragment. Consequently, if several of them are accepted, the information contained is redundant. Figure 5.5.2 illustrates this case (1). If an overlap is too long, a clearly depicted membrane fragment may be too close to the profile tails (case 3 in Figure 5.5.2). Such profiles will be rejected due to the lack of symmetry. Moreover, membrane fragments may be missing in the extracted random profiles (case 4 in Figure 5.5.2). Therefore, an optimal overlap is required. I set it to 2.4 μm, corresponding to the maximal possible canalicular width (case 2 in Figure 5.5.2).

The overlap chosen (2.4 μm) is smaller than the profile length. Hence, each membrane segment that is crossed by initial random lines will still be depicted in several extracted random

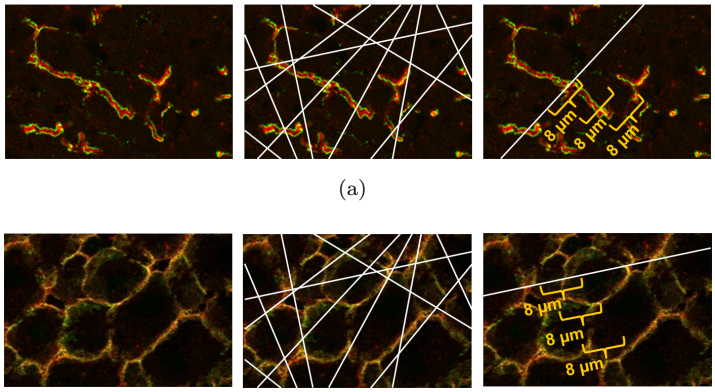


Figure 5.5.1.: Random profile extraction illustrated on canalicular (a) and basolateral (b) membranes. Random lines are drawn in multi-marker images, and individual random profiles are extracted along them.

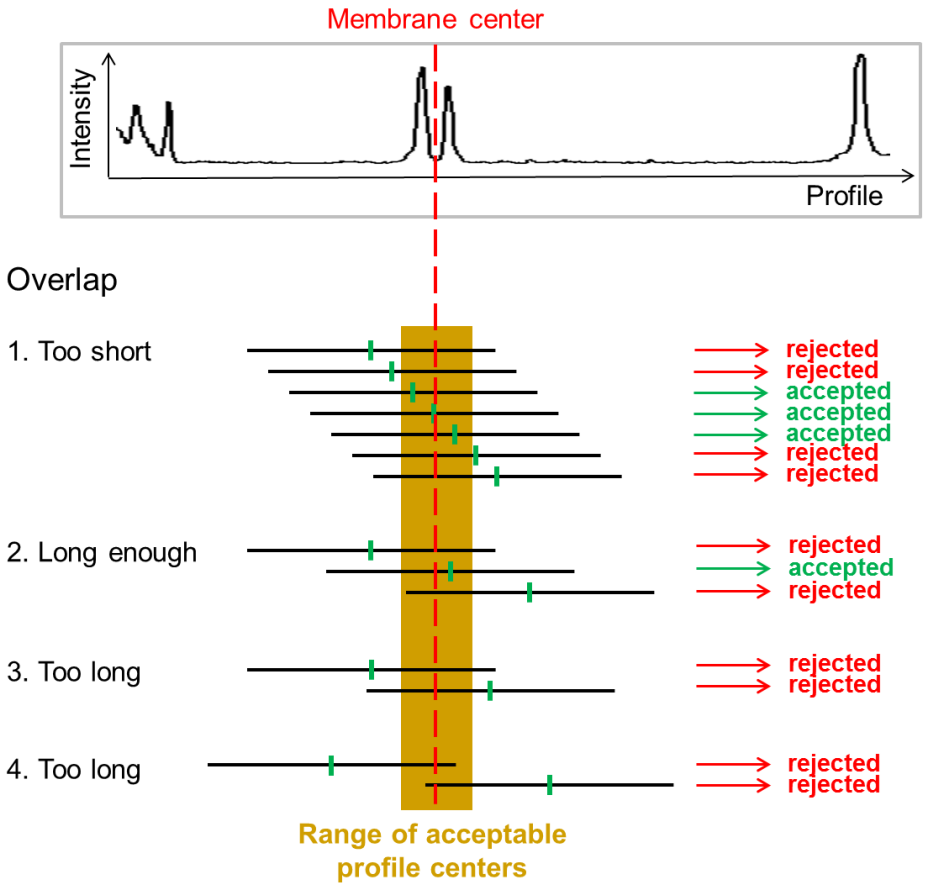


Figure 5.5.2.: Random profiles (black) are extracted with overlap from original intensity vectors. Profile center (green) should not be too far from the center of the membrane (red). If an overlap is too short, too many redundant profiles will be accepted (1). If an overlap is too long, valuable information may be missed in the extracted random profiles (4) or may be lost in the selection process (3).

profiles. However, the center of this membrane segment will be at different positions in all the corresponding random profiles (red line in Figure 5.5.2). Only one of these random profiles may be symmetrical enough and should be selected. To check this symmetry, center of the random profiles (green lines in Figure 5.5.2) are analyzed. They should not be too far from the center of the membrane segment (yellow range in Figure 5.5.2). Therefore, an additional selection criterion is introduced to filter out profiles whose centers are too far from the center of the depicted membrane:

$$\|(CenterOfTheProfile - CenterOfTheMembrane)\| < 1.2, \tag{5.5.1}$$

where $CenterOfTheProfile$ corresponds to the coordinate of the profile center in μm and $CenterOfTheMembrane$ - to coordinate in μm of the membrane center. For canalicular membranes, $CenterOfTheMembrane$ is the center between the tight junctions (Zo-1 intensity peaks); for the basolateral membrane this is the coordinate of the absolute intensity maximum of the membrane marker (Na^+-K^+-ATPase). This criterion excludes redundant acceptance of several random profiles depicting one and the same membrane segment.

Individual random profiles are processed like structure-based profiles. Namely, structure specific profile selection is performed, as described earlier for the 1D-Can (Section 5.3.3) and 1D-Bas (Section 5.4.3) processes. The only difference is the introduction of the additional selection criterion (see Equation 5.5.1). The selected random profiles are analyzed using the numeric descriptors defined for the structure specific zone model (see Section 5.1.3).

The random lines-based workflow can be compared to the structure-based workflow, as well as to the manual workflow, as the same numeric descriptors are calculated for the extracted profiles and used in statistical tests (see Section 5.2).

5.6. Application V: Random lines-based workflow for 2D-Nuc process

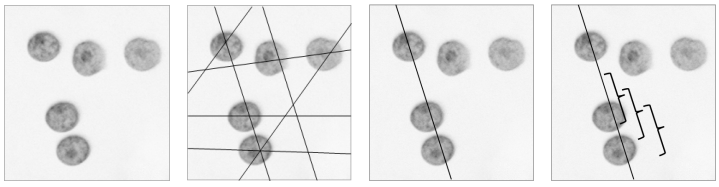
A standard method for quantification of a 2D translocation is structure-based (see Section 3.3). The random lines-based workflow presented in the previous section allows for information extraction without structure detection in images. Thus, it could also be applied to quantify a 2D translocation. I have adapted the random lines-based workflow to quantify glucokinase nucleus-to-cytoplasm translocation in hepatocytes (2D-Nuc process, for biological background see Section 3.1.4).

5.6.1. Profile extraction

Random lines are drawn in the multi-marker images and pixel intensities are extracted along them (see Figure 5.6.1). Each random line may cross several nuclei. Evaluation of such vectors is complicated. Initial long vectors may thus be cut into shorter profiles. These *random profiles* are expected to contain maximally one nucleus. Therefore, an optimal profile length has to be chosen. The size of hepatocyte nuclei varies in the range of 10 - 15 μm. Extracted intensity profiles have to cover a nucleus and a cytoplasmatic region in order to quantify the functional marker translocation. Therefore, the selected profile length should be > 15 μm, but small enough so that no further hepatocytes are covered by one profile.

Rat hepatocytes have a diameter of $20 - 30$ μm [169]. Thus, a range of profile lengths from 20.0 to 30.0 μm was evaluated. The intensity of glucokinase was found to stay nearly constant for distances of more than 14.0 μm from the central point of the nucleus. The length of random profiles is thus set to 28.0 μm.

As explained in the previous section, random profiles are extracted with overlap from the initial long intensity vectors. The optimal overlap was found to be 2.7 μm.



(a) (b) (c) (d)

Figure 5.6.1.: Random profile extraction for the 2D-Nuc process. (a) Initial image; (b), (c) random lines; (d) extracting individual random profiles.

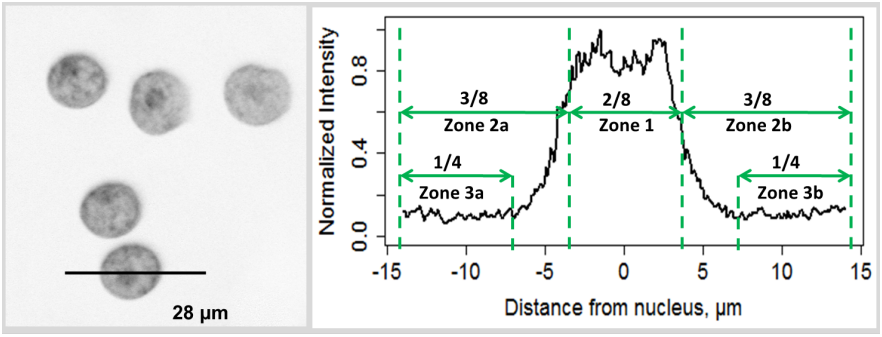


Figure 5.6.2.: Typical image of hepatocyte nuclei and a zone model for profile selection.

5.6.2. Profile selection

Profile selection criteria are developed to filter out random profiles irrelevant to the translocation phenomenon. Here, the profile selection is even more important than in the structure-based workflow. Random lines drawn across the images show even more irrelevant and noisy data. Hence, profile selection has to be able to cope with the less defined input.

The selection is performed only on the structural marker profiles. First, individual random profiles are smoothed by median filtering to eliminate high frequency noise. Then, biologically relevant zones are identified in the random profiles according to the scheme in Figure 5.6.2.

Smoothed structural marker profiles are checked for two conditions. The first one eliminates those profiles which do not have a significant intensity peak in the central part:

$$\frac{Avg(Zone1)}{Avg(Zone3a, Zone3b)} > 4.0, \tag{5.6.1}$$

where $Avg(Zone1)$ is the average intensity value in the central part of the profile, and $Avg(Zone3a, Zone3b)$ is the average intensity at the tails. This condition eliminates noisy profiles and those whose intensity peak is not situated in the central part. The second criterion filters out those profiles whose tails are not symmetrical enough:

$$0.4 \leq \frac{sum(Zone2a)}{sum(Zone2b)} \leq 2.5, \tag{5.6.2}$$

where $sum(Zone2a)$ and $sum(Zone2a)$ correspond to the integral intensities on the left and the right sides of the nucleus. Some examples of accepted and rejected profiles are shown in Figure 5.6.3.

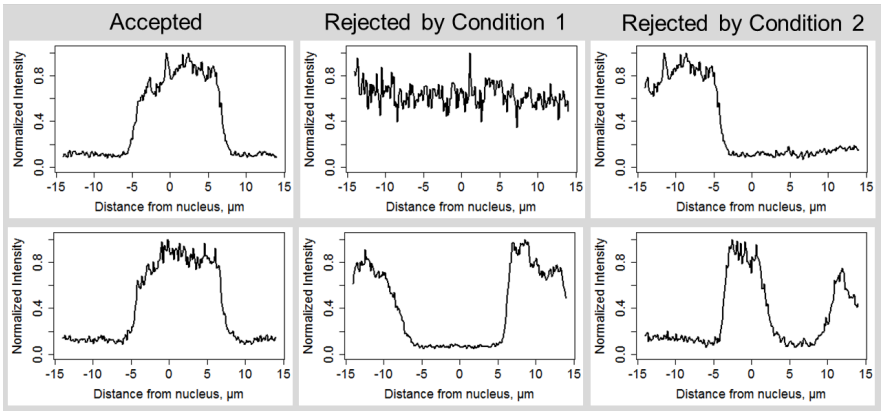
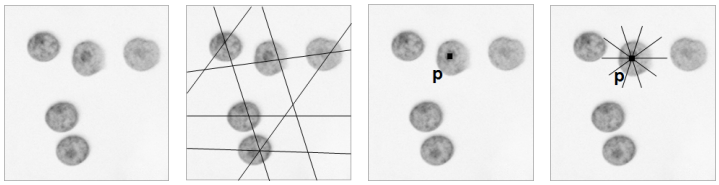


Figure 5.6.3.: Examples of selected and rejected structural marker profiles for the 2D-Nuc process.



(a) (b) (c) (d)

Figure 5.6.4.: Con-centric profile extraction for quantification of the 2D-Nuc process. (a) Initial image; (b) random lines; (c) central point of an accepted profile; (d) further profiles centered at this point.

Con-centric profile extraction

A standard method to evaluate nucleus-to-cytoplasm translocation is structure-based. There, nuclei are detected in images and a certain area around them is segmented as the cytoplasm (see Section 3.3). Nucleus-to-cytoplasm ratio is thus calculated using all pixels belonging to the nuclei and surrounding cytoplasmatic regions.

In contrast, structure detection is avoided in the random lines-based workflow, so the cytoplasm area is also not segmented. Nucleus-to-cytoplasm ratios are calculated based on 1-pixel thick random profiles. If distribution of the functional marker is not homogeneous or anisotropic, random profiles may contain unreliable information. I suggest extracting additional profiles for better *coverage* of the cytoplasmatic and nuclear regions (see Figure 5.6.4).

If an extracted random profile is accepted, then it exhibits an intensity peak in the central part (see Equation 5.6.1). It is then expected to cross one of hepatocyte nuclei. Most probably, the center point of this accepted random profile will be situated in the nuclear region (Figure 5.6.4, (c)). Hence, there may be further acceptable profiles having the same central point. Thus, a number of lines of the optimal profile length is drawn through this point centered at it. Profiles are then extracted along these lines (Figure 5.6.4, (d)). As hepatocyte nuclei exhibit a roundish form, these *con-centric profiles* also lead to a better *coverage* of the cytoplasmatic and the nuclear regions. Extraction of such profiles is further referred to as a *con-centric profile extraction*. The extracted con-centric profiles also undergo the selection procedure.

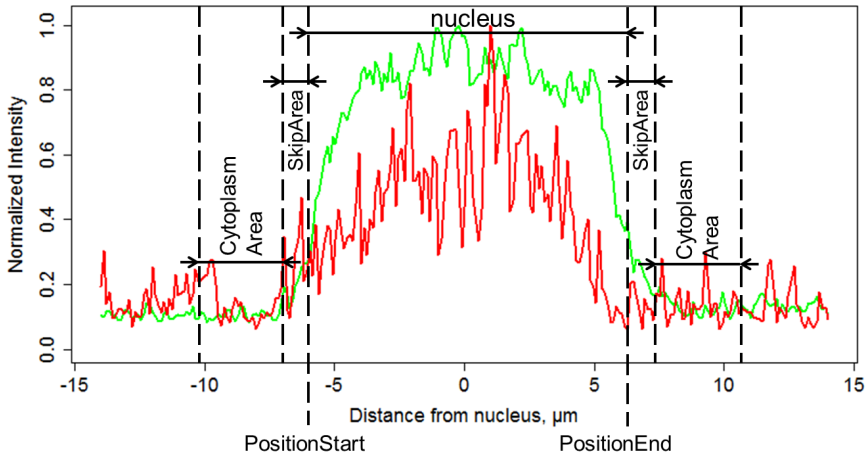


Figure 5.6.5.: Nuclear and cytoplasmatic regions are detected in structural marker profiles (DNA, green) to evaluate functional marker (glucokinase, red) translocation.

5.6.3. Toponomic characterization

Nucleus-to-cytoplasm translocation is quantified by calculation of the functional marker intensity ratio in the nucleus relative to the cytoplasm (see Section 3.3). To achieve this, nuclear and cytoplasmatic regions have to be identified. The random lines-based workflow avoids structure detection in images and performs it only in 1D structural marker profiles (see Figure 5.6.5). Selected random and con-centric profiles are used for this.

To detect a nucleus in a profile, several parameters have to be selected. First, cytoplasmatic regions have to be identified. As cell membranes are not labeled, borders of hepatocytes can not be clearly detected. Therefore, the cytoplasmatic area is identified as a ring around the detected nucleus. I refer to the width of these rings as to the *CytoplasmArea*. Second, the cytoplasmatic area is detected in a certain distance from the nucleus, to ensure their separation. This area to skip is further referred to as the *SkipArea*.

The nuclear regions (see Figure 5.6.5) are found in smoothed structural marker profiles according to the scheme illustrated in Figure 5.6.6. Individual steps are performed as follows:

1. Find the center of the profile (pixel position $Center$) and extract the intensity at this position (I_{Center}).

2. Look for the beginning of the nuclear region. I define the start of the nuclear region at that pixel position where the intensity of the neighboring pixel from the left decreases more than 4 times. Therefore, set initial $PositionStart$ to the $Center$. Move $PositionStart$ to the left and extract the current intensity ($I_{PositionStart}$). Once

$$\frac{I_{Center}}{I_{PositionStart}} \geq 4.0, \tag{5.6.3}$$

   the start of the nuclear region is found at the $PositionStart$.

3. Look for the end of the nuclear region. Similar to the beginning of the nucleus, I define the end of the nuclear region at that pixel position where the intensity of the neighboring pixel from the right decreases more than 4 times. Therefore, set the initial $PositionEnd$ to the $Center$. Move $PositionEnd$ to the right and extract the current intensity ($I_{PositionEnd}$). Once

$$\frac{I_{Center}}{I_{PositionEnd}} \geq 4.0, \tag{5.6.4}$$

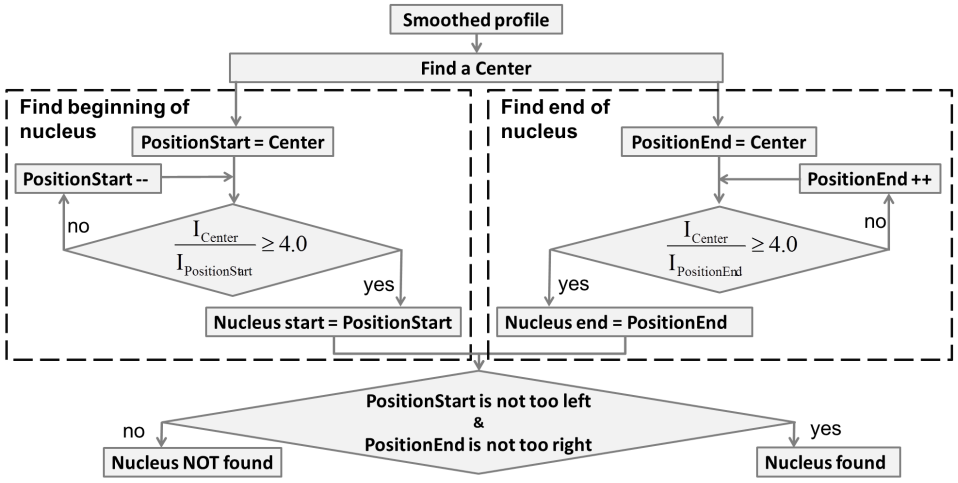


Figure 5.6.6.: Algorithm to find the nuclear region in structural marker profiles.

the end of the nuclear region is found at the *PositionEnd*.

4. Nucleus is found in the region (*PositionStart*, *PositionEnd*), as Figure 5.6.5 illustrates.

5. Check the validity of this nuclear region: it should start at least at at the pixel position (*CytoplasmArea* + *SkipArea*), so that the profile also covers a sufficient cytoplasm region and the area to be skipped on the left side of the nucleus:

$$PositionStart \geq (CytoplasmArea + SkipArea). \tag{5.6.5}$$

   If this is true, continue. Else, reject this profile.

6. Check the validity of this nuclear region: it should end not further than at the pixel position (*profileLength* - *CytoplasmArea* - *SkipArea*), so that the profile also covers a sufficient cytoplasm region and the area to be skipped on the right side of the nucleus:

$$PositionEnd \leq (profileLength - CytoplasmArea - SkipArea). \tag{5.6.6}$$

   If it is true, the detected nuclear region (*PositionStart*, *PositionEnd*) is valid, and this profile can be used for the translocation quantification. Else, reject this profile.

The nucleus-to-cytoplasm intensity ratio is calculated on the functional marker profiles corresponding to the accepted structural marker profiles:

$$ratio = \frac{avg(I_{nucleus})}{avg(I_{cytoplasm})}, \tag{5.6.7}$$

where $avg(I_{nucleus})$ is the average intensity in the region identified as nucleus (*PositionStart*, *PositionEnd*), and $avg(I_{cytoplasm})$ is the average intensity in cytoplasmatic regions on the left (*PositionStart* − *SkipArea* − *CytoplasmArea*, *PositionStart* − *SkipArea*) and on the right side (*PositionEnd* + *SkipArea*, *PositionEnd* + *SkipArea* + *CytoplasmArea*) of the nucleus (see Figure 5.6.5).

This quantification of the 2D-Nuc process can be validated against the established structure-based workflow (see Section 3.3). Therefore, ratios of intensity are calculated by both algorithms. Their means (μ) and standard deviations (σ) are compared.

6. Evaluation and Results

The concept of automated translocation quantification presented in Chapter 4 and defined in Chapter 5 was implemented in a set of algorithms and tested for the 1D-Can, 1D-Bas and 2D-Nuc processes.

This Chapter summarizes statistical results, evaluation and validation of the new automated workflows (see Table 5.0.1). Validation of the **Application I** describes the evaluation of the *structure-based workflow* developed for the 1D-Can process. It is validated against the established manual workflow using the newly defined numeric descriptors. Significance levels of the translocation detected between positive and negative controls are compared. No significant effect is expected to be indicated in negative controls. *Structure-based workflow* for the 1D-Bas process (**Application II**) is assessed similarly.

Further, the *random lines-based workflow* is evaluated for the 1D-Can and 1D-Bas processes (validations of the **Application III** and **IV**, respectively). The *random lines-based workflow* is validated there against the corresponding *structure-based workflows*, because not all datasets have been evaluated manually by my colleagues.

And finally, the *random lines-based workflow* is adapted to quantify 2D-Nuc process (**Application V**). It is validated against the established *structure-based workflow* by the calculated nucleus-to-cytoplasm ratios of the functional marker intensity.

6.1. Validation of the Application I: structure-based workflow for 1D-Can process

6.1.1. Datasets

21 dataset of confocal images of rat liver tissue sections were prepared by my colleagues at the University Clinic Düsseldorf to validate the novel workflows for the 1D-Can process. The images were automatically processed, and the results were compared to those obtained from the manually processed data.

A dataset contains 10 images of different regions of the respective tissue sample. Two proteins, a structural and a functional marker, are immunohistochemically labeled in tissue sections. Canalicular membranes are delineated by tight junctions, where Zonula occludens 1 protein (Zo-1, the structural marker) is labeled. 1D-Can process, translocation of a bile salt export pump (Bsep, the functional marker), is assessed. *Control datasets* include images acquired from tissue sections of the liver incubated under normo-osmolar conditions or perfused with taurocholate (TC). These treatments are not expected to influence the functional marker distribution (see Section 3.4.1). *Test datasets* include images acquired from tissue sections of the liver either incubated under aniso-osmolar conditions (see Section 3.4.1), or perfused with a taurochenodeoxycholate (*TCDC*), or after the bile duct ligature (*BDL*). These treatments may influence distribution of the functional marker, leading to its internalization from canalicular membranes into adjacent hepatocytes, or backwards. For each test dataset, a corresponding control dataset was prepared. For example, control dataset *NormoHyper-1-C* corresponds to the test dataset *NormoHyper-1-T*. As they are compared to each other, notation *NormoHyper-1-C/T* refers to them taken together. Further, *TC-1-C/C* describes two datasets which were obtained either without or with perfusion with TC (taurocholate). Perfusion with TC is not expected to influence the functional marker distribution, hence both these datasets are controls (according to my definition). Table 6.1.1 lists the processed datasets.

Normo-/hyperosmolar		BDL	Perfusion with bile salts	
Control/Test	Control/Control	Control/Test	Control/Control	Control/Test
Osmol-1-C/T	Osmol-1-C/C	BDL-1-C/T	TC-1-C/C	TCDC-1-C/T
Osmol-2-C/T	Osmol-2-C/C	BDL-2-C/T	TC-2-C/C	TCDC-2-C/T
-	Osmol-3-C/C	BDL-3-C/T	TC-3-C/C	TCDC-3-C/T
-	-	-	TC-4-C/C	TCDC-4-C/T
Z-stacks				
-	-	BDL-z1-C/T	TC-z1-C/C	TCDC-z1-C/T
-	-	BDL-z2-C/T	-	-
-	-	BDL-z3-C/T	-	-

Table 6.1.1.: Datasets for the 1D-Can process.

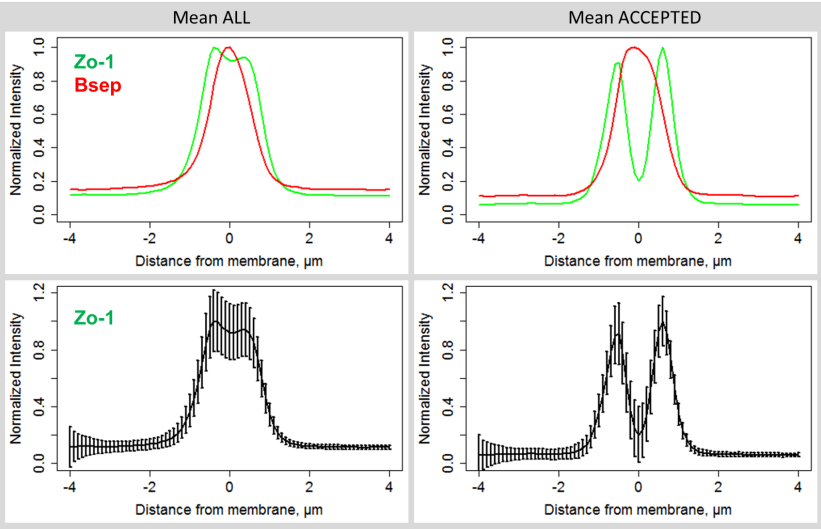


Figure 6.1.1.: Averages of all extracted profiles do not correspond to the biological structure of tight junctions. Averages of selected profiles exhibit two clear Zo-1 intensity peaks and reduced standard deviation.

The expected result of the automated translocation analysis is to detect a changed amount of Bsep in the canalicular membranes and in the cytoplasm in test datasets relative to control datasets. Furthermore, no false positive effects should be detected by the structure-based workflow. Comparison of control datasets should thus not show any significant translocation.

6.1.2. Profile extraction

The structure-based workflow was applied to the datasets. Image processing followed by the automated profile extraction resulted in approx. 4000 profiles per image.

Averages of all extracted profiles were calculated for each dataset for illustration. Structural marker (Zo-1) profiles are expected to exhibit two clear intensity peaks corresponding to tight junctions. Averages over all extracted Zo-1 profiles did not comply with the expected shape (see Figure 6.1.1, left). Furthermore, standard deviations were high due to a high level of variation of individual profiles. A possible reason for this might be profile extraction at damaged canalicular membranes. To exclude such profiles from the evaluation, profile selection was applied.

		Difference of peak height <			
		5%	15%	25%	35%
Valley height <	10%	32	107	151	194
	20%	45	166	230	294
	30%	56	179	264	333
	40%	58	187	273	351

Table 6.1.2.: Numbers of accepted profiles (out of 2871) at different thresholds for the selection criteria (for the 1^{st} image of the dataset *BDL-1-C*).

6.1.3. Profile selection

Profile selection filtered out profiles extracted at damaged and asymmetrical canalicular membranes. This resulted in the reduced standard deviation of the average profiles. Furthermore, averages of selected structural marker profiles exhibited two clear narrow intensity peaks, corresponding to tight junctions (see Figure 6.1.1).

Profile selection was tested for robustness and stability. Two of the four criteria were varied, and numbers of accepted profiles were reported. The first criterion constrains the height difference allowed for two Zo-1 peaks. The second threshold limits the relative central valley height between these peaks. Table 6.1.2 shows respective numbers of accepted profiles on the example of the first image of the dataset *BDL-1-C*. The rejection percentage was found to continuously vary in the range tested. Hence, any threshold in this range can be used.

Profile selection reduced the number of profiles from approx. 4000 to approx. 150 per image. The number of accepted profiles always significantly exceeded 10, which is the number of profiles extracted manually. Consequently, the automated workflow does not only speed up the process, but also increases the number of data points available for analysis.

Profile selection was found to increase significance of the detected translocation (see Table 6.1.4 and Table 6.1.5 for the datasets *Osmol-1-C/T* and *BDL-1-C/T*, respectively). Statistical tests are performed 100 times on 100 vs. 100 randomly sampled profiles. P-values given in the following are median p-values from such 100 experiments (see Section 5.1.3).

6.1.4. Profile ranking

Selected profiles were ranked using two selection criteria (difference of peak intensities and peak - valley contrast). From approx. 150 accepted profiles per image, n top-ranked profiles were evaluated. n was varied between 10 and 50 due to the following reasons. First, the number of manually extracted profiles is 10. Therefore, this is the smallest number of the top-ranked profiles that should be tested. Second, the largest n could be set to 150, which was the average number of the automatically extracted and accepted profiles. Still, some images led to the selection of only 50 profiles. Hence, n was varied in the range $10 - 50$.

Statistical results of WRSTs on control-control datasets varied only slightly for different n of top-ranked selected profiles. WRSTs did not detect any significant translocation using any tested n. Table 6.1.3 illustrates an example for the dataset *TC-2-C/C*.

In all test-control datasets, WRSTs on top-ranked profiles detected translocation of a higher significance than WRSTs on all selected profiles. Table 6.1.5 and Table 6.1.4 illustrate this for the datasets *BDL-1-C/T* and *Osmol-1-C/T*, respectively.

WRTSs on control-test datasets yielded slightly different results for the tested n of the top-ranked profiles. $n = 20$ led to either equal or higher significance levels of the detected translocation than the larger n (30, 40 or 50) (see Table 6.1.4 and Table 6.1.5). Ranking sorts profiles by their *structural quality* which is expressed by their *similarity* to the *ideal* structure of tight junctions. Significance of the detected translocation was found to decrease when profiles of worse *structural quality* were included into the calculation. Hence, the objective of ranking is achieved.

Descriptor	P-value						
	All	Accepted	Top 10	Top 20	Top 30	Top 40	Top 50
X	5.5 e-1	4.6 e-1	2.0 e-1	5.7 e-1	5.3 e-1	5.6 e-1	4.7 e-1
Y	5.0 e-1	5.0 e-1	5.7 e-1	4.2 e-1	5.0 e-1	4.1 e-1	3.4 e-1
Z	4.2 e-1	3.9 e-1	4.1 e-1	5.8 e-1	6.0 e-1	4.9 e-1	5.0 e-1
A	5.4 e-1	5.0 e-1	3.3 e-1	6.5 e-1	5.5 e-1	5.9 e-1	5.8 e-1
B	5.0 e-1	4.8 e-1	4.6 e-1	5.7 e-1	4.7 e-1	3.3 e-1	2.9 e-1
C	5.6 e-1	5.2 e-1	2.8 e-1	3.4 e-1	4.8 e-1	2.4 e-1	2.8 e-1
D	5.0 e-1	4.9 e-1	4.0 e-1	6.4 e-1	5.6 e-1	4.4 e-1	4.7 e-1
E	5.3 e-1	4.9 e-1	3.3 e-1	5.2 e-1	5.2 e-1	3.0 e-1	3.7 e-1
F	7.3 e-2	8.8 e-2	3.5 e-3	3.6 e-2	2.3 e-2	6.5 e-4	1.9 e-4
Variance	8.7 e-2	3.4 e-1	6.4 e-2	1.1 e-1	4.3 e-1	4.9 e-1	3.0 e-1

Table 6.1.3.: Ranking of accepted profiles does not influence WRST results on control-control datasets. Dataset *TC-2-C/C*.

Descriptor	P-value						
	All	Accepted	Top 10	Top 20	Top 30	Top 40	Top 50
X	1.1 e-2	5.5 e-3	1.6 e-4	6.8 e-5	3.3 e-4	2.4 e-3	1.0 e-3
Y	7.8 e-8	4.7 e-14	1.15 e-18	3.9 e-20	2.4 e-15	1.7 e-14	6.7 e-15
Z	7.9 e-10	2.0 e-18	1.17 e-22	7.3 e-23	2.3 e-20	3.1 e-20	1.0 e-20
A	1.2 e-4	1.4 e-6	8.3 e-10	6.9 e-11	1.4 e-7	1.1 e-6	4.0 e-7
B	4.9 e-9	6.7 e-17	2.7 e-21	2.2 e-22	1.5 e-18	6.4 e-18	2.2 e-18
C	5.4 e-13	6.1 e-20	1.7 e-25	2.5 e-25	3.6 e-22	1.3 e-21	1.2 e-21
D	7.2 e-15	4.4 e-23	5.9 e-26	9.0 e-27	6.8 e-25	1.1 e-25	9.6 e-26
E	4.6 e-15	1.1 e-22	1.4 e-27	1.5 e-27	5.6 e-26	4.3 e-26	5.1 e-26
F	1.2 e-11	1.0 e-13	7.3 e-17	1.7 e-17	6.0 e-15	1.5 e-14	1.2 e-14
Variance	3.0 e-10	4.3 e-13	5.3 e-20	1.3 e-17	2.7 e-14	1.1 e-14	1.2 e-14

Table 6.1.4.: Profile selection and ranking increase significance of the detected translocation effect. Dataset *Osmol-1-C/T*.

Descriptor	P-value						
	All	Accepted	Top 10	Top 20	Top 30	Top 40	Top 50
X	4.9 e-1	3.9 e-1	1.7 e-1	6.2 e-2	2.7 e-2	1.9 e-2	1.6 e-1
Y	3.6 e-2	1.6 e-4	1.6 e-2	4.7 e-3	1.5 e-2	9.0 e-3	9.5 e-3
Z	9.9 e-4	1.2 e-5	7.4 e-5	3.5 e-4	6.9 e-4	3.5 e-4	2.5 e-4
A	2.5 e-1	6.3 e-2	6.8 e-1	5.4 e-1	5.6 e-1	4.2 e-1	4.2 e-1
B	3.5 e-3	1.9 e-5	4.2 e-4	9.3 e-4	2.0 e-3	8.1 e-4	5.5 e-4
C	1.0 e-4	2.2 e-8	4.6 e-10	5.1 e-10	6.1 e-10	4.8 e-9	1.7 e-8
D	1.8 e-5	4.0 e-9	1.7 e-12	1.2 e-12	1.6 e-12	5.5 e-10	3.1 e-10
E	7.4 e-6	7.9 e-10	1.7 e-13	9.0 e-13	5.1 e-13	1.5 e-11	1.8 e-11
F	5.9 e-2	5.4 e-8	2.0 e-6	4.7 e-4	2.7 e-4	4.6 e-4	7.1 e-4
Variance	5.3 e-4	1.8 e-4	2.2 e-7	3.1 e-9	1.3 e-8	5.9 e-7	1.4 e-6

Table 6.1.5.: Profile selection and ranking increase significance of the detected translocation effect. Dataset *BDL-1-C/T*.

Descriptor	*BDL-3-C/T*		*TCDC-2-C/T*	
	Top 10	Top 20	Top 10	Top 20
x	7.1 e-5	5.2 e-6	3.4 e-2	7.6 e-3
Y	6.3 e-14	7.3 e-16	1.4 e-3	2.2 e-5
Z	4.3 e-9	1.1 e-12	2.3 e-2	4.4 e-4
A	7.3 e-10	2.4 e-11	8.9 e-3	1.4 e-3
B	1.4 e-11	6.1 e-15	4.4 e-3	5.0 e-5
C	3.4 e-14	1.3 e-13	2.7 e-3	1.1 e-4
D	5.4 e-5	1.0 e-8	3.3 e-1	3.8 e-2
E	2.7 e-8	1.1 e-11	2.5 e-2	1.7 e-3
F	1.5 e-4	1.6 e-4	1.3 e-3	7.7 e-11
Variance	3.4 e-4	5.0 e-7	6.1 e-2	3.9 e-3

Table 6.1.6.: Significance of the detected translocation effect depends on the number of the top-ranked profiles used. Dataset *BDL-3-C/T* and *TCDC-2-C/T*.

WRSTs on $n = 20$ top-ranked selected profiles led to the detection of a slightly more significant translocation effect than WRSTs on $n = 10$ in 3 datasets (see Table 6.1.4 and Table 6.1.6). In the rest control-test datasets, WRSTs on 10 or 20 top-ranked profiles indicated comparable significance levels of the functional marker translocation (e.g. Table 6.1.5).

Generally, I have not observed the usage of neither $n = 10$ nor 20 top-ranked selected profiles leading to generally higher significance levels. Still, for three datasets $n = 20$ was found to be better. Therefore, the 1D-Can process is further evaluated on 20 top-ranked profiles. Calculations on 20 profiles indicate higher significance levels and are fast.

6.1.5. Optimal profile width selection

Various profile widths in the range of 1 - 7 pixels (0.1 - 0.7 μm with pixel size of 100 nm) were evaluated to find the optimal one. Generally, narrow profiles (width = 1 pixel) are accepted only if they exactly correspond to all selection criteria. Thus, such individual profiles have to meet the selection requirements by themselves. Neighborhoods of these profiles are not taken into account. A narrow profile can be accepted even if skeleton pixels neighboring to its center yield profiles that do not meet the selection requirements. Such accepted profiles may be situated at damaged biological structures.

Wider profiles (width $>$ 1 pixel), in turn, can be accepted even if their individual components are not perfect, but the average corresponds to the selection criteria. Thus, a neighborhood of every profile is accounted for. This resembles the logic of manual profile extraction, where only sufficiently long undamaged structures are analyzed. On the other hand, too wide profiles that include clean and symmetrical structural marker profiles can be rejected if these profiles are averaged with those taken at damaged canalicular membranes. A compromise has to be achieved by selecting the optimal profile width.

Figure 6.1.2 illustrates average accepted profiles of various widths for the third image of the dataset *Osmol-1-C/T*. All these averages exhibit two clear Zo-1 intensity peaks corresponding to tight junctions. These Zo-1 peaks are symmetrical, clear, narrow and separated by a valley. Hence, selected profiles were extracted on undamaged, symmetrical and clearly depicted canalicular membranes.

A shape of the molecule distribution function does not seem to be strongly influenced by the different profile widths tested. WRSTs were performed on the profiles of different widths to evaluate this statistically. Profile width = 3 led to the highest significance levels of the detected translocation (see Table 6.1.7 for the dataset *Osmol-1-C/T*). Profile width = 3 seems to represent a good compromise between too narrow profiles (width = 1) and too wide profiles (width $\geq$ 5). Too narrow profiles do not account for the neighborhood of the pixel, at which the profile was extracted. Too wide profiles seem to average too large membrane segments.

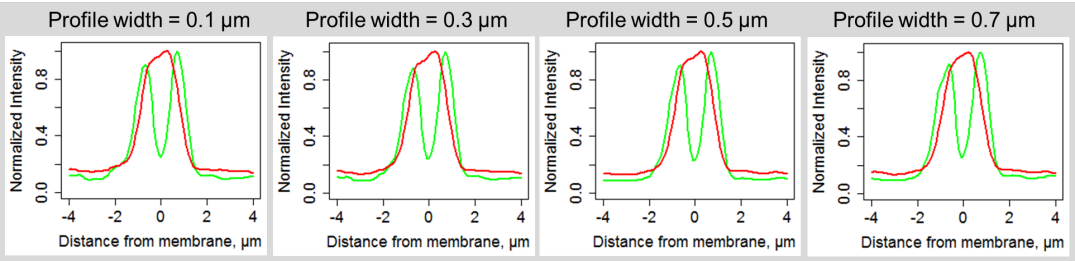


Figure 6.1.2.: Plots of average automatically extracted and selected profiles of various width (0.1 – 0.7 μm) for the 1D-Can process. Red - Bsep, green - Zo-1.

Descriptor	P-value			
	0.1 μm	0.3 μm	0.5 μm	0.7 μm
X	1.7 e-2	6.8 e-5	2.4 e-4	5.7 e-4
Y	4.2 e-16	3.9 e-20	1.6 e-15	4.4 e-14
Z	4.2 e-21	7.3 e-23	7.0 e-21	2.0 e-21
A	2.8 e-7	6.9 e-11	4.9 e-8	4.2 e-7
B	2.0 e-19	2.2 e-22	3.4 e-19	1.5 e-18
C	3.4 e-23	2.5 e-25	1.1 e-22	9.1 e-22
D	7.9 e-27	9.0 e-27	9.4 e-25	9.8 e-28
E	1.1 e-26	1.5 e-27	3.2 e-26	1.8 e-27
F	1.7 e-14	1.7 e-17	3.2 e-18	1.0 e-15
Variance	7.6 e-13	1.3 e-17	3.7 e-18	1.2 e-18

Table 6.1.7.: Profile width = 0.3 μm leads to the highest significance levels of the detected translocation. P-values from 20 top-ranked selected automatically extracted profiles. Dataset *Osmol-1-C/T*.

Such profiles are more biased towards long clearly depicted membrane regions. Furthermore, computational cost grows with the increasing profile width. Therefore, a profile width of 3 pixel was chosen for further translocation quantifications. In other words, profiles are every time extracted along 3 parallel lines and averaged according to the scheme in Figure 5.1.3.

6.1.6. Comparison of the structure-based workflow to the manual workflow

4 datasets (*BDL-1-C/T*, *BDL-2-C/T*, *BDL-3-C/T* and *Osmol-1-C/T*) were also evaluated by the manual workflow (see Section 3.2). 10 profiles per image were extracted manually by an expert. Descriptors and statistical variances were computed for these profiles. These values were further compared to the values calculated from the top 20 profiles extracted and selected by the structure-based workflow.

Means and standard deviations of the descriptor values indicate a strong correspondence between the results from the structure-based workflow and the manual workflow (see Table 6.1.8). The structure-based workflow is considered to be valid, as parameters of the automatically extracted profiles are distributed similarly to those extracted manually. Statistical evaluations on the automatically and manually extracted profiles are also expected to be comparable. WRSTs were performed to validate the structure-based workflow statistically.

Statistical tests

The structure-based workflow was further validated against the manual workflow by results of statistical tests. First, control datasets (*BDL-2-C* and *BDL-3-C*) were compared for the functional marker distribution. No false positive translocation effects should be detected there

Descriptor	Structure-based		Manual	
	μ	σ	μ	σ
X	0.552	0.169	0.595	0.150
Y	1.11	0.528	1.13	0.481
Z	0.766	0.338	0.715	0.266
A	0.361	0.124	0.380	0.110
B	0.446	0.194	0.434	0.164
C	1.93	0.624	1.97	0.639
D	1.35	0.449	1.15	0.380
E	0.806	0.247	0.729	0.224
F	0.558	0.185	0.514	0.149
Variance	0.056	0.017	0.056	0.015

Table 6.1.8.: Descriptor values μ and σ calculated from 20 top-ranked automatically extracted and selected profiles (structure-based workflow) and 10 manually extracted profiles (manual workflow). Dataset *BDL-1-C*.

Descriptor	P-value	
	Manual	Structure-based
X	7.3 e-2	3.1 e-4
Y	8.6 e-4	2.0 e-2
Z	6.1 e-3	2.1 e-1
A	3.6 e-3	4.5 e-3
B	2.0 e-3	9.7 e-1
C	3.2 e-3	6.1 e-1
D	9.6 e-3	2.4 e-1
E	4.3 e-3	4.0 e-1
F	8.4 e-2	4.8 e-3
Variance	7.2 e-2	5.5 e-1

Table 6.1.9.: Significance of translocation detected in control-control dataset by the structure-based workflow (20 top-ranked profiles) and by the manual workflow (10 profiles). Datasets *BDL-2-C* and *BDL-3-C*.

(p-value > 0.05). Evaluation by the manual workflow did not yield a consistent decision on the presence or absence of a significant translocation (see Table 6.1.9). There, only 3 descriptors suggested that no Bsep internalization into hepatocytes took place (X, F, Variance). WRSTs on top 20 selected automatically extracted and selected profiles also yielded inconsistent results. Six of ten descriptors (Z, B, C, D, E, Variance) suggested that Bsep distribution has not changed. As comparison of control datasets should not detect significant Bsep translocation, the structure-based workflow performed better than the manual workflow: 6 vs. 3 descriptors yielded expected results.

Further, control and test datasets were compared. Test datasets include images of livers either perfused in hyper-osmolar buffer (e.g. *Osmol-1-T*) or livers with induced cholestasis (e.g. bile duct ligature in *BDL-1-T*), or livers perfused with TCDC (e.g. *TCDC-1-T*). These conditions may lead to Bsep translocation from the canalicular membranes into the cytoplasm, what is expected to be visible compared to control datasets. Four datasets were evaluated by the manual workflow (*BDL-1-C/T*, *BDL-2-C/T*, *BDL-3-C/T* and *Osmol-1-C/T*).

In all these datasets, both the manual and the structure-based workflow indicated a significant translocation of the functional marker (see Table 6.1.10 and Table 6.1.11). For the dataset *Osmol-1-C/T*, all descriptors indicated a significant effect on profiles extracted by both workflows. For the other 3 datasets, a significant translocation was detected by the majority of the

Descriptor	*BDL-1-C/T*		BDL-3-C/T		BDL-2-C/T	
	Manual	Structure-based	Manual	Structure-based	Manual	Structure-based
X	1.9 e-1	6.2 e-2	2.0 e-1	5.2 e-6	3.9 e-1	7.6 e-4
Y	4.8 e-4	4.7 e-3	1.9 e-17	7.3 e-16	4.1 e-9	4.6 e-7
Z	8.9 e-5	3.5 e-4	6.3 e-14	1.1 e-12	4.1 e-4	3.9 e-5
A	2.9 e-1	5.4 e-1	8.4 e-9	2.4 e-11	5.7 e-3	1.4 e-5
B	8.5 e-5	9.3 e-4	2.5 e-16	6.1 e-15	1.7 e-6	3.4 e-6
C	1.3 e-6	5.1 e-10	3.9 e-21	1.3 e-13	1.8 e-15	2.6 e-7
D	9.3 e-8	1.2 e-12	8.1 e-15	1.0 e-8	9.9 e-7	2.2 e-3
E	4.0 e-8	9.0 e-13	1.3 e-18	1.1 e-11	1.5 e-10	1.8 e-5
F	6.5 e-1	4.7 e-4	5.5 e-1	1.6 e-4	1.9 e-11	7.9 e-10
Variance	6.8 e-10	3.1 e-9	8.9 e-15	5.0 e-7	3.8 e-9	2.9 e-3

Table 6.1.10.: Significance of translocation detected in control-test datasets by the structure-based workflow (20 top-ranked profiles) and by the manual workflow (10 profiles). Datasets *BDL-1-C/T*, *BDL-3-C/T* and *BDL-2-C/T*.

Descriptor	P-value	
	Manual	Structure-based
X	1.3 e-3	6.8 e-5
Y	1.6 e-21	3.9 e-20
Z	3.3 e-24	7.3 e-23
A	2.0 e-10	6.9 e-11
B	5.7 e-24	2.2 e-22
C	3.7 e-26	2.5 e-25
D	4.3 e-27	9.0 e-27
E	1.1 e-27	1.5 e-27
F	6.2 e-14	1.7 e-17
Variance	3.4 e-17	1.3 e-17

Table 6.1.11.: Significance of translocation detected in control-test dataset by the structure-based workflow (20 top-ranked profiles) and by the manual workflow (10 profiles). Dataset *Osmol-1-C/T*.

descriptors.

Significance levels of the translocation were comparable for both workflows in one dataset (*Osmol-1-C/T*). For the other 3 datasets, either one or another workflow detected a more significant effect based on different descriptors. However, manual workflow indicated mostly higher significance levels than the structure-based workflow. Still, I consider the results of the structure-based workflow comparable to the manual evaluation.

Almost all descriptors (except X, A and F) were able to indicate a significant Bsep translocation in the manually extracted profiles from the dataset *BDL-1-C/T*. Evaluation of the automatically extracted, selected and ranked profiles improved this result. Only two descriptors (X and A) were unable to detect the functional marker translocation (see Table 6.1.10). The best descriptors yielding the smallest p-values were C, D and E both for the manual and for the structure-based workflow.

6.1.7. Descriptors

The rest datasets were not evaluated manually but were still processed by the structure-based workflow in order to study the newly defined descriptors.

Descriptor	P-value		
	TCDC-3-C/T	*TCDC-4-C/T*	*TCDC-1-C/T*
X	1.9 e-1	5.0 e-1	1.8 e-2
Y	4.4 e-3	1.5 e-7	6.0 e-1
Z	1.6 e-4	7.4 e-12	2.7 e-1
A	4.7 e-1	1.5 e-2	3.3 e-1
B	5.4 e-4	3.5 e-10	3.8 e-1
C	4.9 e-6	3.0 e-11	4.3 e-2
D	1.2 e-7	1.8 e-14	5.8 e-3
E	1.9 e-7	1.1 e-13	6.4 e-3
F	2.2 e-6	1.9 e-8	2.2 e-2
Variance	4.3 e-7	2.3 e-11	5.8 e-1

Table 6.1.12.: Significance of translocation detected in control-test datasets by the structure-based workflow (20 top-ranked profiles). Datasets *TCDC-3-C/T*, *TCDC-4-C/T* and *TCDC-1-C/T*.

Table B.0.1 illustrates these results for the control-control datasets. Descriptor F was found to falsely detect a significant translocation in most datasets (5 of 8) processed by the structure-based workflow. This descriptor seems to be inappropriate for Bsep translocation quantification.

Descriptor X was the second worst with 4 false positive effects. Other newly defined descriptors performed better. They detected a false positive effect in 2 (descriptors Y, Z, A, B, D) or 3 (descriptors C, E) datasets. The statistical variance led a false positive translocation detection in 3 datasets. Herewith, it performed worse than 5 newly defined descriptors.

TC-4-C/C and *Osmol-1-C/C* do not seem to be good examples of control-control datasets. There, a significant translocation was detected by 9 of 10 (or 6 of 10, respectively) descriptors. The significance levels were also high (up to e-11). Excluding these datasets from the consideration, newly defined descriptors indicated absence of a significant Bsep translocation in 5 of 6 control-control datasets.

Table B.0.2 summarizes results of the structure-based workflow for the control-test datasets. Descriptors X and A led to the highest false negative rate (3 of 9, see Table 6.1.12 and Table B.0.2). The same descriptors were among the worst for control-control datasets.

Descriptors C, D and E led to the highest significance levels (see Table 6.1.11, Table 6.1.12 and Table B.0.2) and did not lead to any false negative translocation detection. Descriptors C, D and E are thus considered to be the most representative for Bsep translocation quantification. They also include the same zones (2, 3 and 4) in their formula (see Equation 5.3.11, 5.3.12 and 5.3.13). Hence, zones 2 (the tight junctions), 3 and 4 (cytoplasmatic regions) seem to contain the most relevant information for the quantification of the 1D-Can process.

In contrast, zone 1 (interior of a canaliculus) seems to contain inconclusive information. Zone 1 is included in the formula of the descriptors which performed worse, such as X, A and F (see Equation 5.3.6, 5.3.9 and 5.3.14). Zone 1 seems to be a region that is not representative for the quantification of Bsep translocation. Therefore, descriptors that include this zone were not considered further. Keeping these findings in mind, only descriptors C, D and E are taken into account in the upcoming evaluations.

Descriptor F is the only descriptor that is not based on zones of the fixed length. It performed worst on control-control datasets. It also led to 2 false negative translocation detections on 4 control-test datasets evaluated manually. Therefore, zones with varying lengths are found unreliable for translocation quantification. As expected, zones of the fixed length provide a better statistical basis for translocation quantification.

As discussed in the previous sections, statistical variance was used in earlier work (Schmitt et al. [142]) to characterize intensity distribution profiles (see Section 4.4.4). Significance levels of the translocation detected using the statistical variance and the newly defined descriptors were

Descriptor	P-value	
	TCDC-1-C/T	*TC-1-C/C*
C	4.3 e-2	5.6 e-1
D	5.8 e-3	3.8 e-1
E	6.4 e-3	4.1 e-1
Variance	5.8 e-1	5.4 e-3

Table 6.1.13.: WRSTs on the statistical variance lead to false positive and false negative translocation detection.

compared. Table 6.1.10, Table 6.1.11 and Table 6.1.12 illustrate some examples of the control-test datasets, where significance of the translocation detected by WRSTs on the statistical variance was several orders lower than in WRSTs on descriptor D. As Table B.0.2 illustrates, in all control-test datasets, significance levels of the translocation detected using descriptors were either higher or the same than using the statistical variance. These statistical results support the usage of the newly defined descriptors for the translocation quantification. Extracted profiles do not necessarily exhibit a Gaussian form. The descriptors, in contrast to the statistical variance, do not make any assumptions on the shape function of the distribution profiles. Hence, their usage as parameters to represent such profiles is more appropriate.

Furthermore, WRSTs on the statistical variance did not detect any significant translocation in one control-test dataset *TCDC-1-C/T* where the newly defined descriptors indicated a significant effect. On one control-control dataset (*TC-1-C/C*), the statistical variance indicated a significant effect while selected descriptors suggested its absence. In total, the statistical variance performed worse than the most relevant defined descriptors (C, D and E) and was thus not considered further for quantification of the 1D-Can process.

Significance of translocation indicated by statistical tests and descriptor values

Translocation is quantitatively analyzed by WRSTs on descriptor values. P-values obtained are used to describe the significance of the effect indicated by statistical tests. To support this approach, the p-values were observed together with the respective descriptor values. Descriptor D was selected for this evaluation, as one of the most representative for the 1D-Can process. Values of the descriptor D were analyzed for the datasets processed. Mean and standard deviation were calculated (μ and σ), as well the overlap of the intervals [μ - σ, μ + σ] of descriptor D values from two datasets (control and control, or control and test).

Table 6.1.14 illustrates the results sorted by the p-values. Interestingly, *Osmol-1-C/T* is the only dataset, where the intervals [μ - σ, μ + σ] of descriptor D values from the control and the test dataset do not overlap. The corresponding difference of the mean values is the highest and the p-value is the lowest (e-27), indicating the most significant translocation effect.

Generally, greater differences between $\mu(1)$ and $\mu(2)$ and shorter overlaps lead to lower p-values. Either increasing the overlap, or decreasing the absolute difference between $\mu(1)$ and $\mu(2)$ results in worse separation of the datasets (1) and (2), and in lower p-values. For example, datasets *TCDC-4-C/T* and *BDL-3-C/T* were found to have a similar overlap between the control and test datasets (0.462 vs. 0.486). However, absolute difference between $\mu(1)$ and $\mu(2)$ is higher in the dataset *TCDC-4-C/T*, than in the dataset *BDL-3-C/T* (0.864 vs. 0.346), leading to the higher significance of the indicated translocation (e-14 vs. e-8, respectively).

WRSTs perform ranking of individual values to calculate the test statistics. Therefore, consideration of only μ and σ values is insufficient to reproduce the statistical results. Nevertheless, the correlation between p-values and $|\mu(1) - \mu(2)|$ was found to be high (correlation coefficient = 0.932).

For further statistical validation of the approach, differences $|\mu(1) - \mu(2)|$ were compared between the control-control and control-test datasets. WRST resulted in a p-value 9.4 e-4 (unpaired test, confidence level = 0.95). Herewith, the difference of descriptor D values between

Dataset	Descriptor D values						P-value
	$\mu(1)$	$\sigma(1)$	$\mu(2)$	$\sigma(2)$	$\lvert\mu(1) - \mu(2)\rvert$	Overlap	
Osmol-1-C/T	2.72	0.89	1.27	0.51	1.451	-0.050	9.0 e-27
TCDC-4-C/T	1.66	0.48	2.52	0.84	0.864	0.462	1.8 e-14
BDL-1-C/T	1.35	0.45	0.94	0.26	0.410	0.308	1.2 e-12
BDL-3-C/T	1.29	0.48	0.95	0.35	0.346	0.486	1.0 e-8
TCDC-3-C/T	1.50	0.69	2.06	0.83	0.560	0.957	1.2 e-7
Osmol-2-C/T	1.47	0.53	2.13	1.15	0.655	1.028	1.5 e-4
BDL-2-C/T	1.41	0.62	1.18	0.57	0.223	0.969	2.2 e-3
TCDC-1-C/T	2.16	0.79	1.88	0.88	0.275	1.399	5.8 e-3
TC-3-C/C	1.83	0.51	1.67	0.52	0.167	0.856	2.5 e-2
TCDC-2-C/T	1.71	0.53	1.96	0.76	0.254	1.040	3.8 e-2
TC-4-C/C	1.57	0.51	1.72	0.56	0.147	0.918	1.1 e-1
BDL-2-C/BDL-3-C	1.35	0.45	1.29	0.48	0.061	0.870	2.4 e-1
TC-1-C/C	1.26	0.32	1.30	0.34	0.039	0.625	4.0 e-1
Osmol-1-C/C	1.32	0.54	1.26	0.52	0.056	0.999	5.0 e-1
Osmol-3-C/C	1.36	0.49	1.32	0.49	0.040	0.941	5.5 e-1
TC-2-C/C	2.45	0.95	2.40	0.86	0.049	1.760	6.4 e-1
Osmol-2-C/C	1.46	0.53	1.35	0.44	0.114	0.858	7.0 e-1

Table 6.1.14.: Significance of statistical tests and descriptor values (sorted by p-values). (1) - Control datasets, (2) - the second control dataset (in a control-control pair) or the test dataset (in a control-test pair).

Dataset	Canalicular width			
	Median	Mean	1^{st} quartile	3^{rd} quartile
Osmol-1-C	1.43	1.56	1.29	1.71
Osmol-1-T	1.86	1.83	1.57	2.14

Table 6.1.15.: Distribution of canalicular width (in μm) in the dataset *Osmol-1-C/T*.

the control and test datasets was indicated statistically, and does not occur by chance. WRST on $|\sigma(1) - \sigma(2)|$ also led to a significant result (p-value = 2.7 e-3).

These findings support the validity of the quantitative translocation analysis using numeric descriptors. The difference between the control and test datasets was found to be significantly greater than the difference between the control and control datasets. Furthermore, p-values were found to correlate to the difference in descriptor values. Therefore, referring to p-values when discussing the significance of the translocation effect (i.e. difference between distributions of the calculated descriptor values) is considered valid.

6.1.8. Restriction of canalicular width

If a structural marker (Zo-1) profile exhibits two clear intensity peaks corresponding to tight junctions, the distance between these peaks indicates the canalicular width at this particular point. Canalicular width is known to vary from 0.8 to 2.5 μm. I have evaluated impact of restricting the canalicular width on the translocation quantification.

Canalicular widths were studied in the dataset *Osmol-1-C/T*. Their distribution was found to be different under normo- and hyper-osmolar conditions (see Table 6.1.15).

The width restriction was performed so that still the majority of the profiles was included into evaluation. The 1^{st} and 3^{rd} quartiles of the canalicular width were calculated for the profiles extracted from the control dataset and for the profiles extracted from the test dataset. The intervals (1^{st} quartile, 3^{rd} quartile) of the canalicular width were intersected for these

Dataset	Canalicular width	Descriptor		
		C	D	E
Osmol-1-C/T	All widths	2.5 e-25	9.0 e-27	1.5 e-27
	1.2 – 2.2 μm	1.3 e-20	4.4 e-24	6.5 e-25
	1.4 – 2.0 μm	2.4 e-24	1.0 e-23	6.5 e-26
TCDC-3-C/T	All widths	4.9 e-6	1.2 e-7	1.9 e-7
	1.2 – 2.2 μm	3.8 e-6	2.8 e-7	1.1 e-7
	1.4 – 2.0 μm	1.5 e-7	2.6 e-8	3.7 e-9
TCDC-4-C/T	All widths	3.0 e-11	1.8 e-14	1.1 e-13
	1.2 – 2.2 μm	1.3 e-11	2.2 e-14	1.8 e-14
	1.4 – 2.0 μm	1.1 e-12	1.4 e-13	5.7 e-14
Osmol-2-C/C	All widths	3.0 e-1	6.6 e-1	5.8 e-1
	1.2 – 2.2 μm	3.7 e-1	1.9 e-1	2.5 e-1
	1.4 – 2.0 μm	4.6 e-1	1.6 e-1	1.9 e-1

Table 6.1.16.: Influence of canalicular width restriction on significance of the detected translocation (structure-based workflow).

conditions. The resulting width interval 1.29 – 2.14 μm was rounded to 1.2 – 2.2 μm. To evaluate the impact of further width restriction, an even smaller width range was tested (1.4 – 2.0 μm).

Descriptor values were calculated for 20 top-ranked profiles accepted either with or without width restriction. For 3 evaluated control-control datasets (*TC-1-C/C*, *TC-2-C/C* and *Osmol-2-C/C*), no significant translocation was indicated in profiles with restricted canalicular width (see an example in Table 6.1.16). Table B.0.3 shows the complete data.

Significance of the translocation detected in 4 control-test datasets (*Osmol-1-C/T*, *BDL-3-C/T*, *TCDC-3-C/T* and *TCDC-4-C/T*) varied slightly more. For the selected descriptors C, D and E, significance levels increased or decreased by several orders of magnitude. Interestingly, restriction of the canalicular width to 1.2 – 2.2 μm led to the detection of a mostly less significant translocation, than restriction to the range 1.4 – 2.0 μm. However, the highest increase of the significance levels was only by 2 orders of magnitude and was not achieved by all descriptors. Therefore, canalicular width was not restricted in further evaluations.

6.1.9. Impact of microscopy settings on translocation quantification

Biological processes being three-dimensional by nature are frequently studied on 2D data (see Section 4.2.1). I suggested a 3D information extraction from confocal z-stacks by combination of the original 2D optical layers. *New average confocal* images were obtained as averages of three neighboring confocal layers in z-stacks (see Table 6.1.1).

Synthetic wide-field images were generated to compare their information content to that from the confocal data. It was not possible to record the same biological samples also by wide-field microscopy. I superimpose all images from each of the z-stacks in order to simulate these data (see Figure 6.1.3). The obtained images can be considered an approximation to those which could be acquired by wide-field microscopy. The thickness of each confocal layer is 0.40 μm, while the distance between them is 0.35 μm. Consequently, a significant part of the sample thickness (approx. 3 of 4 μm) is covered by the recorded z-stack containing 8 optical layers.

And finally, maximum intensity projections (*MIP*) [15, 137] were obtained for each z-stack. All data were processed by the structure-based workflow.

The generated and the original confocal images were compared by two criteria. First, the number of selected profiles is considered. It is an important factor, as more data points provide a better statistical basis for the analysis. Figure 6.1.4 presents the results for the dataset *TC-z1-C/C*. For the new average confocal images (red line), the number of the accepted profiles were

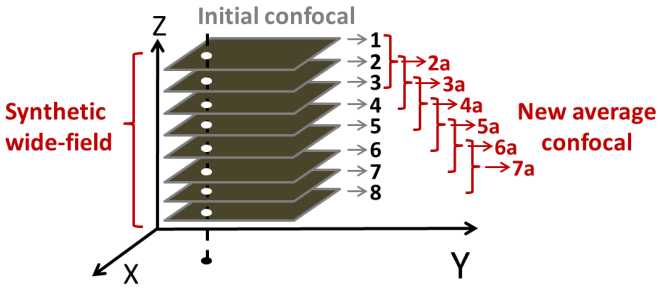


Figure 6.1.3.: Illustration of 3D information extraction from a confocal z-stack: synthetic wide-field images and new average confocal layers (2a, 3a, etc.).

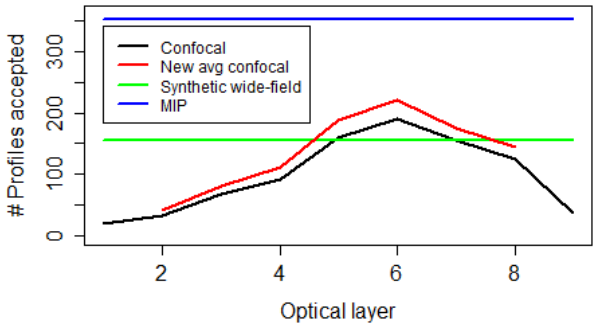


Figure 6.1.4.: Number of accepted profiles depending on a microscopy and/or image combination method. Dataset *TC-z1-C/C*.

often larger than those for the initial confocal layers (black line). This supports the hypothesis that more canaliculi are rescued due to 3D information extraction. MIP collected the brightest points through all confocal layers. Hence, most of the canalicular membranes were used for the analysis and the largest number of data points was extracted (blue line). In contrast, synthetic wide-field images contained the information from blurred regions (out-of-focus light). Consequently, fewer profiles were acceptable there (green line).

Second, the images were compared regarding the quality of the information contained. If an image is not sharp and contrasty the extracted intensity profiles will not be representative for the real distribution of the functional marker. The discrimination between the control and test datasets may be unreliable. Therefore, the quality of the extracted data is assessed by the significance of the detected translocation. When comparing control and test datasets, lower p-values indicate a greater difference, and thus better relevance of the image data. Comparison of control datasets, in turn, should indicate no significant effect (p-value > 0.05).

The best layer was found in each z-stack as the most sharp and yielding the greatest number of profiles selected. Such best layers were also evaluated separately to compare their information content to that from the complete z-stack.

Control-control (*TC-z1-C/C*) and control-test (*BDL-z1-C/T*, *BDL-z2-C/T*, *BDL-z3-C/T* and *TCDC-z1-C/T*) datasets were evaluated based on the profiles automatically extracted from the images (see Table 6.1.17). False positive translocation in the dataset *TC-z1-C/C* was detected only using synthetic wide-field images. All other approaches based on confocal images have not detected any significant change in Bsep distribution. For the datasets *BDL-z1-C/T*, *BDL-z2-C/T* and *BDL-z3-C/T*, synthetic wide-field images did not yield any acceptable intensity profile. All other methods based on confocal images and their combinations detected significant translocation effect (p-value ≤ 0.05), as expected. WRSTs on synthetic wide-field images of the dataset *TCDC-z1-C/T* detected a several orders less significant translocation

Images	P-value				
	TC-z1-C/C	BDL-z1-C/T	BDL-z2-C/T	BDL-z3-C/T	TCDC-z1-C/T
Synthetic wide-field	6.8 e-5	-	-	-	2.6 e-6
MIP	5.3 e-1	3.4 e-9	4.0 e-19	7.2 e-16	8.2 e-15
All confocal	5.0 e-1	1.8 e-6	4.7 e-17	1.2 e-15	7.9 e-10
The best confocal	4.0 e-1	1.0 e-11	2.3 e-18	3.0 e-12	1.5 e-8
New avg confocal (all)	4.9 e-1	3.9 e-11	3.3 e-17	1.3 e-18	1.7 e-10
New avg confocal (best)	2.6 e-1	1.2 e-19	2.6 e-18	1.4 e-15	6.3 e-9

Table 6.1.17.: Significance of translocation detected by the structure-based workflow based on original images of confocal z-stacks and images generated from them. WRSTs on descriptor D values.

effect than WRSTs on other combinations of original confocal layers.

The most significant translocation effect was detected in two datasets (*BDL-z1-C/T*, *BDL-z3-C/T*) using the new average confocal images. In two other datasets, MIP represented the most relevant information source for the translocation quantification.

Interestingly, only for two datasets (*BDL-z1-C/T* and *BDL-z2-C/T*), WRSTs on the best confocal layer led to the higher significance levels than tests on all confocal layers. For other test-control datasets, significance levels were slightly lower. This is a surprising finding, as the best confocal layer always yielded the largest number of selected profiles. In contrast, very few profiles were selected in the uppermost and the lowest layers, taken on the top and the bottom of the biological sample, respectively (see Figure 6.1.4).

Generally, 3D information extraction by generation of new average confocal layers was found to have no impact on the result (e.g. datasets *TC-z1-C/C*, *BDL-z2-C/T*, *TCDC-z1-C/T*) or led to higher significance levels of the detected translocation (e.g. datasets *BDL-z1-C/T*, *BDL-z3-C/T*). If a canaliculus does not run parallel to the focal plane, the extracted intensity profiles will be discarded due to the lack of symmetry. This can be changed if the information from 3 neighboring layers is combined. A simple averaging allowed rescuing of more data points and led to higher significance levels of the detected translocation (e.g. e-11 vs. e-19 for the dataset *BDL-z1-C/T*, and e-12 vs. e-15 for the dataset *BDL-z3-C/T*).

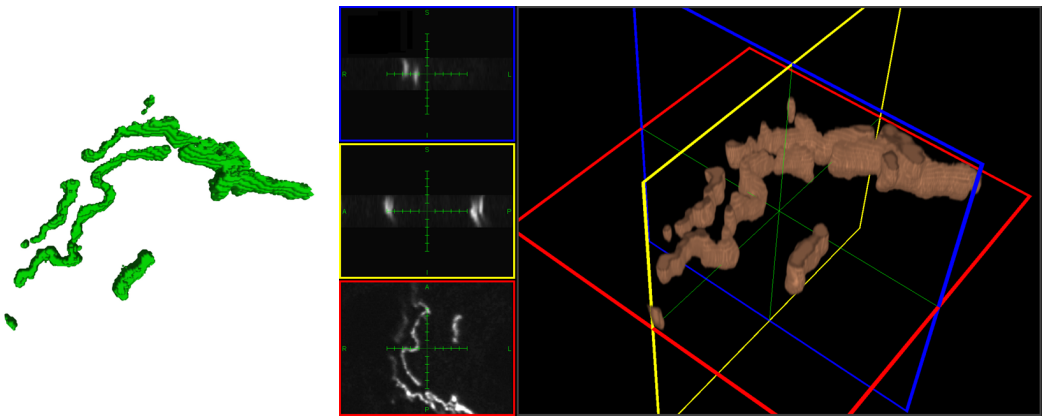
Canalicular structures were segmented (using BioImageXD [82]) and studied in 3D in order to better understand these findings. Figure 6.1.5, (a) illustrates tight junctions of an individual canaliculus. The cross section shows that the tight junctions are slightly shifted relative to each other along the Z-axis (Figure 6.1.5, (b), top left). Therefore, only the middle (the best) confocal layer can capture both their intensities. It can thus depict most of the canalicular membranes symmetrically leading to the largest number of profiles selected (see Figure 6.1.4). This confirms my findings for the z-stacks processed (see Figure 6.1.4).

New average confocal layers capture more intensity from both of such shifted tight junctions than the original confocal layers. Consequently, they represent an even richer and more complete information source. Statistical tests confirmed this by higher significance levels of the detected translocation (see Table 6.1.17).

6.2. Validation of the Application II: structure-based workflow for 1D-Bas process

6.2.1. Datasets

13 datasets of confocal images of rat liver tissue sections were prepared by my colleagues at the University Clinic Düsseldorf to validate the structure-based workflow for the 1D-Bas process.



(a) (b)

Figure 6.1.5.: (a) Representative canaliculus segmented and reconstructed by its tight junctions from a z-stack. (b) Its tight junctions in 3D with lateral projections.

The datasets were evaluated, and the results were compared to those obtained from the manually processed data. As in case of the 1D-Can process, each dataset contains 10 images of different regions of the respective tissue samples. Translocation of the functional marker Ntcp (sodium taurocholate cotransporting polypeptide) is evaluated in relation to the basolateral structural marker Na^+-K^+-ATPase (1D-Bas process).

Similar to the datasets for the 1D-Can process, *control datasets* include images acquired from tissue sections of the liver incubated under normo-osmolar conditions or perfused with taurocholate (TC). These treatments are not expected to influence the functional marker distribution (see Section 3.4.1). *Test datasets* include images acquired from tissue sections of the liver perfused with taurochenodeoxycholate (TCDC, see Section 3.4.1). Functional marker is expected to be internalized from the basolateral membranes into the adjacent hepatocytes. For each test dataset, a corresponding control dataset was prepared. For example, control dataset *TCDC-1-C* corresponds to the test dataset *TCDC-1-T*. As they are compared to each other, notation *TCDC-1-C/T* refers to them taken together. Further, *TC-1-C/C* describes two datasets which were obtained either without or with perfusion with TC (taurocholate). Perfusion with TC is not expected to influence the functional marker distribution, hence both these datasets are controls (according to my definition). Table 6.2.1 lists the processed datasets.

Automated translocation quantification is expected to detect a decrease of Ntcp in the basolateral membranes and its increase in the cytoplasm in test datasets relative to control datasets. Furthermore, no false positive effects should be detected by the structure-based workflow when comparing control datasets.

Perfusion with bile salts	
Control / Control	Control / Test
TC-1-C/C	TCDC-1-C/T
TC-2-C/C	TCDC-2-C/T
TC-3-C/C	TCDC-3-C/T
TC-4-C/C	TCDC-4-C/T
TC-5-C/C	TCDC-5-C/T
TC-6-C/C	TCDC-6-C/T
TC-7-C/C	-

Table 6.2.1.: Datasets for the 1D-Bas process.

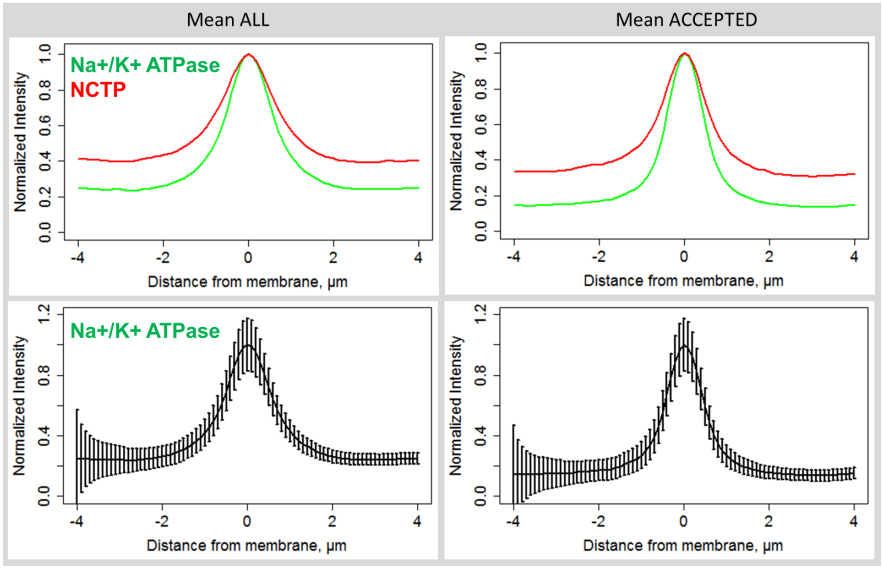


Figure 6.2.1.: Profile selection makes the intensity peak in structural marker profiles narrower and reduces standard deviation (1D-Bas process).

6.2.2. Profile extraction

The structure-based workflow was applied to the datasets. Approx. 20000 profiles per image were extracted. Averages of all extracted profiles were calculated for each dataset for illustration. Average structural marker (Na^+-K^+-ATPase) profiles are expected to exhibit one narrow intensity peak corresponding to the basolateral membrane. Although averages over all extracted Na^+-K^+-ATPase profiles exhibited such a peak (see Figure 6.2.1, left), it was not narrow. Standard deviations were high due to a high level of variation of individual profiles. This indicated that profiles taken at damaged basolateral membranes might have been included into the evaluation. To exclude these profiles, the profile selection was performed.

6.2.3. Profile selection

Profile selection made the average Na^+-K^+-ATPase profiles significantly narrower (see Figure 6.2.1, right). The distribution of selected Na^+-K^+-ATPase profiles exhibited a significant reduction of the standard deviation. The level of profile variation has decreased, indicating the profile selection at symmetrical undamaged basolateral membranes only.

Profile selection reduced the number of profiles from approx. 20000 to approx. 2000 per image. Figure 6.2.1 illustrates improvements of the average plots of all profiles (n = 12572) after the profile selection (n = 1582) for the first image of the dataset *TCDC-3-C/T*. The number of accepted profiles always significantly exceeded 10, which is the number of manually extracted data points. Consequently, the automated workflow does not only speed up the process, but also increases the number of data points available for analysis.

Profile selection was also found to reduce the rate of false negative translocation detection. Evaluation of one control-test dataset (*TCDC-2-C/T*) did not indicate functional marker translocation (p-value > 0.05) by the majority of the descriptors calculated from all extracted profiles. However, statistically significant translocation was found using selected profiles only (see Table 6.2.2). Statistical tests are performed 100 times on 100 vs. 100 randomly sampled profiles. P-values given in the following are median p-values from such 100 experiments (see Section 5.1.3).

Table 6.2.2 illustrates an extreme case when functional marker translocation can be missed when evaluating all extracted profiles. This was, however, not the case for all processed test-

Descriptor	P-value	
	All	Accepted
L	3.4 e-2	9.1 e-3
M	1.7 e-1	5.4 e-3
N	7.1 e-2	1.2 e-3
O	1.0 e-1	1.4 e-3
P	7.0 e-1	4.0 e-3
Variance	1.2 e-4	2.0 e-4

Table 6.2.2.: Profile selection makes translocation effect visible. Dataset *TCDC-2-C/T*.

Descriptor	P-value		
	All	Accepted	Top 20
L	1.1 e-1	8.4 e-2	1.3 e-5
M	1.7 e-2	1.6 e-3	3.6 e-10
N	8.8 e-3	5.3 e-4	4.5 e-12
O	8.4 e-3	3.9 e-4	5.8 e-13
P	3.7 e-2	1.3 e-2	7.0 e-8
Variance	1.6 e-4	5.3 e-6	4.5 e-8

Table 6.2.3.: Profile selection and ranking increase significance of the detected translocation effect. Dataset *TCDC-1-C/T*.

control datasets. Still, significance of the translocation was higher for most, when evaluated on selected profiles only (see an example in Table 6.2.3).

6.2.4. Profile ranking

Selected profiles were ranked by the relative intensity of the central peak. This ranking criterion describes the profile *similarity* to the *ideal* basolateral membrane.

The ranking influenced significance of the detected translocation. In most test-control pairs, evaluation of the top-ranked profiles only led to the detection of more significant functional marker translocation than based on all selected profiles (see Table 6.2.3).

From approx. 2000 accepted profiles per image, n top-ranked profiles were evaluated. n was varied between 10 and 500 due to the following reasons. First, the number of manually extracted profiles is usually 10. Therefore, this is the smallest number of the top-ranked profiles that should be tested. Second, the largest n could be set to 2000, but some images led to the selection of 500 profiles. Hence, n was varied in the range 10 – 500.

Statistical results of WRSTs on control-control datasets did not vary much for the tested n of top-ranked selected profiles. Only WRSTs on 10 top-ranked profiles detected a significant translocation in the control-control dataset *TC-3-C/C* (see Table 6.2.4).

For all test-control datasets, translocation of a higher significance was detected in WRSTs on top-ranked profiles than in WRSTs on all selected profiles. Table 6.2.3 illustrates this for the dataset *TCDC-1-C/T*.

Statistical significance of the translocation in control-test datasets varied depending on n. Generally, the stronger the restriction (or the smaller n), the more significant effect was detected. WRSTs on top 10 and 20 profiles always led to the detection of a more significant effect than WRSTs on 50 - 500 profiles (see Tables 6.2.5 and 6.2.6). Hence, statistical tests on profiles of a better *structural quality* lead to more significant translocation detection.

For one control-test dataset, WRSTs on top 10 profiles detected translocation of the highest significance (see Table 6.2.5). For the rest control-test datasets, significance levels were either equal, or slightly higher for WRSTs on top 20 profiles (see Table 6.2.6). Therefore, $n = 20$ is considered the best number of top-ranked selected profiles to quantify Ntcp translocation.

Descriptor	P-value						
	Accepted	Top 10	Top 20	Top 50	Top 100	Top 200	Top 500
L	5.0 e-1	3.1 e-1	4.5 e-1	2.5 e-1	3.8 e-1	4.7 e-1	4.6 e-1
M	5.0 e-1	1.9 e-2	1.8 e-1	2.0 e-1	1.9 e-1	4.2 e-1	5.0 e-1
N	5.2 e-1	6.0 e-3	3.0 e-1	3.0 e-2	6.5 e-2	3.4 e-1	4.6 e-1
O	5.0 e-1	5.0 e-3	5.7 e-1	4.0 e-2	7.2 e-2	3.3 e-1	4.7 e-1
P	5.0 e-1	3.2 e-2	7.1 e-1	4.0 e-2	1.1 e-1	3.7 e-1	4.8 e-1
Variance	4.9 e-1	1.7 e-1	4.9 e-1	9.1 e-2	1.3 e-1	3.2 e-1	5.0 e-1

Table 6.2.4.: Influence of profile ranking on significance of the detected translocation. Dataset *TC-3-C/C*.

Descriptor	P-value					
	Top 10	Top 20	Top 50	Top 100	Top 200	Top 500
L	8.8 e-5	1.3 e-5	1.2 e-5	6.8 e-4	3.7 e-3	2.0 e-2
M	1.2 e-11	3.6 e-10	4.0 e-8	2.6 e-6	2.3 e-5	9.9 e-5
N	1.1 e-14	4.5 e-12	2.2 e-10	1.1 e-7	1.6 e-6	8.1 e-6
O	5.5 e-15	5.8 e-13	6.7 e-11	4.4 e-8	2.2 e-7	6.4 e-6
P	1.5 e-10	7.0 e-8	5.1 e-7	2.6 e-5	1.6 e-4	6.0 e-4
Variance	7.9 e-7	4.5 e-8	2.9 e-8	8.5 e-9	2.3 e-7	6.6 e-7

Table 6.2.5.: Influence of profile ranking on significance of the detected translocation. Dataset *TCDC-1-C/T*.

Descriptor	P-value					
	Top 10	Top 20	Top 50	Top 100	Top 200	Top 500
L	2.9 e-5	2.4 e-4	3.1 e-2	5.6 e-3	1.9 e-4	2.1 e-4
M	2.3 e-6	2.0 e-7	4.2 e-9	6.0 e-9	6.8 e-9	2.4 e-7
N	1.8 e-11	9.9 e-12	3.8 e-10	7.3 e-10	3.9 e-10	9.8 e-9
O	5.3 e-10	5.3 e-11	3.5 e-11	6.1 e-11	4.1 e-11	2.2 e-9
P	4.1 e-10	6.8 e-10	3.1 e-7	4.6 e-7	1.8 e-7	1.6 e-6
Variance	9.2 e-3	4.4 e-4	2.6 e-4	1.8 e-5	2.9 e-6	3.8 e-6

Table 6.2.6.: Influence of profile ranking on significance of the detected translocation. Dataset *TCDC-2-C/T*.

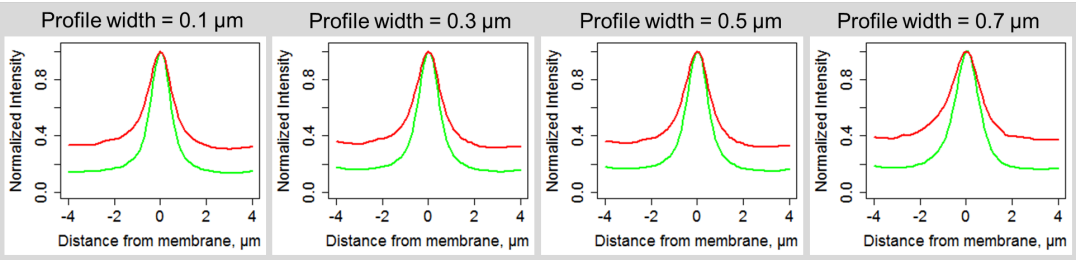


Figure 6.2.2.: Plots of average automatically extracted and selected profiles of various width (0.1 – 0.7 μm) for the 1D-Bas process. Red - NTCP, green - Na^+-K^+-ATPase.

Descriptor	P-value			
	0.1 μm	0.3 μm	0.5 μm	0.7 μm
L	1.3 e-5	2.8 e-5	3.2 e-6	2.1 e-5
M	3.6 e-10	1.3 e-11	1.0 e-11	1.7 e-11
N	4.5 e-12	9.3 e-11	3.4 e-11	8.2 e-12
O	5.8 e-13	1.8 e-12	1.3 e-12	1.4 e-12
P	7.0 e-8	7.7 e-7	5.1 e-7	1.0 e-8
Variance	2.6 e-7	1.5 e-6	3.0 e-6	1.1 e-10

Table 6.2.7.: Influence of the profile width on the significance levels of the detected translocation. P-values from 20 top-ranked selected automatically extracted profiles for the dataset *TCDC-1-C/T*.

6.2.5. Optimal profile width and length

Various profile widths in the range of 1 – 7 pixels (0.1 – 0.7 μm with pixel size of 100 nm) were evaluated to find the optimal one. Generally, narrow profiles (width = 1 pixel) are accepted only if they exactly correspond to all selection criteria. Neighborhoods of these profiles are not taken into account. A narrow profile can thus be accepted even if skeleton pixels neighboring to its center yield profiles that do not meet the selection requirements. Such accepted profiles may be situated at damaged biological structures.

Wider profiles (width > 1 pixel), in turn, can be accepted even if their individual components are not perfect, but their average corresponds to the selection criteria. A neighborhood of every profile is accounted for. This resembles the logic of manual profile extraction, where only sufficiently long undamaged structures are analyzed. On the other hand, wide profiles are biased towards undamaged symmetrical membranes. A compromise has thus to be achieved by selecting the optimal profile width.

Average structural marker profiles of all tested widths exhibited a narrow peak corresponding to the structural marker of the basolateral membrane (see Figure 6.2.2).

All tested widths also led to comparable significance levels of the detected translocation (see an example in Table 6.2.7). However, computational cost grows with the increasing profile width. Approx. 20000 individual profiles are extracted from a typical confocal image of stained basolateral membranes. Increasing the profile width would lead to a growth of the computation time. Hence, the width of 1 pixel (0.1 μm) was chosen as a compromise between the benefit and computational cost.

Optimal profile length was found to be 8 μm, as for the 1D-Can process (see Section 5.3.2). 8 μm were found sufficient to cover the basolateral membrane and significant parts of the hepatocytes cytoplasm to illustrate variations in the functional marker distribution. At the pixel size = 100 nm, this length corresponds to 81 pixels (nearest odd value).

Descriptor	Structure-based		Manual	
	μ	σ	μ	σ
L	0.665	0.205	0.653	0.156
M	0.788	0.178	0.791	0.146
N	1.686	0.952	1.535	0.826
O	0.508	0.163	0.506	0.128
P	2.038	0.713	1.890	0.829
Variance	2.4 e-4	7.3 e-5	1.89 e-4	7.03 e-5

Table 6.2.8.: Descriptor values μ and σ calculated from 20 top-ranked automatically extracted and selected profiles (structure-based workflow) and 10 manually extracted profiles (manual workflow). Dataset *TCDC-2-C/T*.

Descriptor	*TC-3-C/C*		*TC-6-C/C*	
	Manual	Structure-based	Manual	Structure-based
L	5.9 e-1	4.5 e-1	9.9 e-1	2.6 e-1
M	8.7 e-1	1.8 e-1	7.8 e-1	1.5 e-1
N	6.3 e-1	3.0 e-1	3.8 e-1	3.6 e-1
O	7.0 e-1	5.7 e-1	8.0 e-1	5.4 e-2
P	6.3 e-1	7.1 e-1	1.1 e-1	6.3 e-2
Variance	4.3 e-4	4.9 e-1	2.9 e-1	3.5 e-1

Table 6.2.9.: Significance of translocation detected in control-control datasets by the structure-based workflow (20 top-ranked profiles) and by the manual workflow (10 profiles). Datasets *TC-3-C/C* and *TC-6-C/C*.

6.2.6. Comparison of the structure-based workflow to the manual workflow

11 of 13 datasets (see Table 6.2.1 except for *TC-7-C/C* and *TCDC-6-C/T*) were also evaluated by the manual workflow (see Section 3.2). 10 profiles per image were extracted manually by an expert. Descriptors and statistical variances were computed for these profiles. These values were further compared to the values calculated from the top 20 profiles extracted and selected by the structure-based workflow.

Means and standard deviations of the descriptor values were studied to validate the automated profile extraction. A strong correspondence between the automated and manual results was found (see Table 6.2.8). The structure-based workflow is considered to be valid, as parameters of the automatically extracted profiles are distributed similarly to those extracted manually. Statistical evaluations on the automatically and manually extracted profiles are thus expected to be comparable. To validate the structure-based workflow statistically, WRSTs are performed on these data.

Statistical tests

The structure-based workflow was validated against the manual workflow by the results of statistical tests. First, control datasets were compared for the functional marker (Ntcp) distribution. No false positive translocation effects should be detected there (p-value > 0.05).

For 2 of 6 control-control datasets, the results of the manual and the structure-based workflow demonstrated a good correlation (see Table 6.2.9). No significant translocation of the functional marker was indicated by any of these methods. One exception is the evaluation of the statistical variance in the dataset *TC-3-C/C*. However, none of the newly defined descriptors indicated a false positive effect.

In four other control-control datasets, the results of the structure-based workflow differed from the manually obtained results. In the dataset *TC-2-C/C*, the manual workflow indicated

Descriptor	*TC-2-C/C*	
	Manual	Structure-based
L	9.6 e-2	5.8 e-2
M	9.7 e-1	2.3 e-2
N	1.5 e-1	1.7 e-3
O	4.4 e-1	1.7 e-3
P	4.9 e-2	1.3 e-2
Variance	7.3 e-2	2.0 e-2

Table 6.2.10.: Significance of translocation detected in control-control dataset by the structure-based workflow (20 top-ranked profiles) and by the manual workflow (10 profiles). Dataset *TC-2-C/C*.

Descriptor	*TC-1-C/C*		*TC-4-C/C*		*TC-5-C/C*	
	Manual	Structure-based	Manual	Structure-based	Manual	Structure-based
L	2.3 e-2	5.5 e-1	3.0 e-2	5.2 e-1	1.1 e-2	6.9 e-1
M	3.5 e-2	5.8 e-1	3.0 e-3	3.7 e-1	7.9 e-1	4.3 e-1
N	1.8 e-4	5.3 e-1	5.5 e-3	6.3 e-1	7.2 e-3	3.3 e-1
O	5.5 e-4	5.2 e-1	1.6 e-3	6.7 e-1	1.0 e-1	3.3 e-1
P	5.4 e-4	5.9 e-1	9.7 e-2	4.8 e-1	1.5 e-3	2.0 e-1
Variance	1.0 e-3	1.6 e-1	1.9 e-1	1.5 e-1	3.9 e-2	4.7 e-1

Table 6.2.11.: Significance of translocation detected in control-control datasets by the structure-based workflow (20 top-ranked profiles) and by the manual workflow (10 profiles). Datasets *TC-1-C/C*, *TC-4-C/C* and *TC-5-C/C*.

no Ntcp translocation while the structure-based workflow detected a significant effect (see Table 6.2.10). In three other datasets, manual workflow detected a significant functional marker translocation, while the structure-based workflow indicated its absence (see Table 6.2.11). Provided that the control-control datasets have been correctly prepared, the structure-based workflow performed better than the manual workflow with 1 vs. 3 false positive translocation detections.

Further, control and test datasets were compared. Functional marker translocation is expected to be visible from the control to test datasets (p-value ≤ 0.05). Both, the manual and the structure-based workflow, indicated a significant translocation in 3 of 5 control-test datasets (see Table 6.2.12). For 2 of these datasets (*TCDC-1-C/T* and *TCDC-2-C/T*), significance levels of the translocation detected by the structure-based workflow were even higher than indicated by the manual workflow.

In two other control-test datasets (*TCDC-4-C/T* and *TCDC-5-C/T*), the manual workflow did not yield conclusive results. Three of six descriptors indicated a significant translocation (see Table 6.2.13). The structure-based workflow, in turn, indicated a significant translocation in one of these datasets (*TCDC-4-C/T*) by the majority of the descriptors. The other dataset *TCDC-5-C/T* does not seem to be a good control-test example. Excluding this dataset from the consideration, the structure-based workflow indicated a significant translocation in 4 of 4 control-test datasets.

6.2.7. Descriptors

The rate of false positive translocation detection in control-control datasets was found to be comparable for all newly defined descriptors (see Table B.0.4).

In 6 control-test datasets, descriptors L and P performed worst (3 false negatives). Descriptors M, N, O and the statistical variance led to the false negative translocation detection in 1

Descriptor	*TCDC-1-C/T*		*TCDC-2-C/T*		*TCDC-3-C/T*	
	Manual	Structure-based	Manual	Structure-based	Manual	Structure-based
L	1.3 e-2	1.3 e-5	1.1 e-2	2.4 e-4	3.8 e-2	6.7 e-1
M	2.0 e-6	3.6 e-10	1.2 e-6	2.0 e-7	8.0 e-12	4.6 e-2
N	6.0 e-6	4.5 e-12	6.5 e-8	9.9 e-12	1.5 e-10	3.2 e-2
O	4.2 e-7	5.8 e-13	5.0 e-8	5.3 e-11	6.0 e-12	1.8 e-2
P	3.9 e-3	7.0 e-8	9.0 e-6	6.8 e-10	1.2 e-5	1.3 e-1
Variance	7.5 e-10	2.9 e-8	5.7 e-7	4.4 e-6	1.6 e-4	9.6 e-3

Table 6.2.12.: Significance of translocation detected in control-test datasets by the structure-based workflow (20 top-ranked profiles) and by the manual workflow (10 profiles). Datasets *TCDC-1-C/T*, *TCDC-2-C/T* and *TCDC-3-C/T*.

Descriptor	*TCDC-4-C/T*		*TCDC-5-C/T*	
	Manual	Structure-based	Manual	Structure-based
L	1.2 e-3	5.2 e-1	9.0 e-1	7.2 e-1
M	3.8 e-1	2.7 e-2	5.0 e-2	6.3 e-1
N	7.8 e-2	3.1 e-2	9.0 e-2	3.4 e-1
O	5.7 e-2	1.2 e-2	5.0 e-2	4.2 e-1
P	7.5 e-4	1.2 e-1	1.4 e-1	3.0 e-1
Variance	1.1 e-3	1.6 e-5	3.0 e-3	5.2 e-1

Table 6.2.13.: Significance of translocation detected in control-test datasets by the structure-based workflow (20 top-ranked profiles) and by the manual workflow (10 profiles). Datasets *TCDC-4-C/T* and *TCDC-5-C/T*.

dataset (see Table B.0.5).

Among descriptors, the highest significance levels of the detected translocation were achieved by M, N and O both for the manual and for the structure-based workflows. These descriptors also include the same profile zones (3, 4 and 5) in their formula (see Equation 5.4.4, 5.4.5 and 5.4.6). Hence, zones 3, 4 and 5 seem to be the most representative for the Ntcp translocation detection. In contrast, zones 1 and 2 seem to contain less conclusive information. Zones 1 and 2 might also be too large for the reliable translocation evaluation. These zones are included in the formula of the descriptor L (see Equation 5.4.3) which performed worse (see Table 6.2.12) and even missed translocation effects.

It was discussed in the previous sections that the statistical variance was used to characterize molecule distribution profiles (see Section 4.4.4). I compared the significance levels of the translocation detected using the statistical variance to the significance of the effect detected using the newly defined descriptors.

First, false positive translocation was detected in one control-control dataset when tested on the statistical variance but not on the descriptors (*TC-3-C/C*). Generally, the false positive rate was comparable for the descriptors and the statistical variance. Second, significance levels of the detected translocation were mostly higher, when evaluated using descriptors (see Table B.0.5). These results support quantification of the 1D-Bas Process using the newly defined descriptors. Distribution profiles do not necessarily exhibit a Gaussian form. Descriptors, in contrast to the statistical variance, do not make any assumptions on the shape function of the profiles. Hence, their usage as parameters is more appropriate. Therefore, only descriptors M, N and O are taken into account in further evaluations.

6.3. Validation of the Application III: random lines-based workflow for 1D-Can process

I defined a novel method for translocation quantification avoiding time consuming and complicated structure detection in images (see Section 5.5). This random lines-based workflow has been developed and tested for the 1D optimal transport processes (1D-Can and 1D-Bas). Further, it has been validated against the structure-based workflow.

As described earlier, the random lines-based workflow starts with the information extraction along random lines drawn in images (see Section 4.4.3). To generate intensity profiles comparable to structure-based profiles, segments of a pre-defined length are extracted from the original random lines. The length of choice depends on a biological problem studied and is the same as in the structure-based profile extraction. The optimal profile length both for the 1D-Can and 1D-Bas processes is found to be 8.0 μm. Individual random profiles are extracted with overlap (2.4 μm) in order to use all valuable non-redundant information.

1D-Can and 1D-Bas processes share common features, but differ in structure specific details. Therefore, the random lines-based workflow is validated for each of them separately.

6.3.1. Optimal number of random lines

As no structure detection is performed in the random lines-based workflow, the extracted information fully depends on the initial random lines drawn in the multi-marker images. Therefore, the first parameter to choose is the number of these lines.

Random lines should cross a sufficient number of membranes present in an image to deliver a result comparable to the structure-based profile extraction. The number of the random lines should thus be sufficiently large. On the other hand, different random lines can cross one and the same membrane fragment several times leading to the extraction of redundant information. Hence, too many lines may be excessive. Furthermore, the random profile extraction should take not much more time than the structure-based profile extraction.

Various numbers of random lines ($\#RandLines$) are evaluated for the images from the earlier mentioned datasets. The number of the accepted random profiles and as well as the time required for the profile extraction and selection are compared to the structure-based workflow. Table 6.3.1 illustrates this on the example of the 5^{th} image from the dataset *Osmol-1-C* (1512 $\times$ 1512 pixel). Experiments with the same $\#RandLines$ are repeated 3 times, and mean values of the calculated parameters are reported.

First, very few initial random lines (20 and 50) do not permit selecting a sufficient number of random profiles. The selection of at least the same number of profiles as in the established manual workflow ($n = 10$) needs $\#RandLines \geq 100$. The respective computation times are still significantly smaller than in the structure-based workflow. For example, total computation time for $\#RandLines = 100$ is approx. 7 times shorter than for the structure-based workflow (T_{total}: 41.4 vs. 284.7 seconds, respectively). Random profile extraction is faster than the structure-based, as slow structure detection in images with manual training of the algorithm is avoided.

Second, average descriptor values change only slightly when $\#RandLines$ is varied in the range 100 – 1500. The random profile extraction appears to be a robust approach. I refer here and in the following only to the descriptors C and D, as they were found to be the most relevant for the quantification of the 1D-Can process.

Third, the smallest $\#RandLines$ that permits selecting a comparable number of profiles to the structure-based workflow is 1500. The computation times are in this case also comparable (e.g. T_{total}: 280.3 vs. 284.7 seconds). According to these findings, $\#RandLines = 1000$ appears to be a good compromise. The random lines-based workflow is still faster than the structure-based workflow, while the number of profiles selected and distributions of descriptor values are comparable.

Figure 6.3.1 illustrates a small fragment of the 5^{th} image from the dataset *Osmol-1-C*. Centers

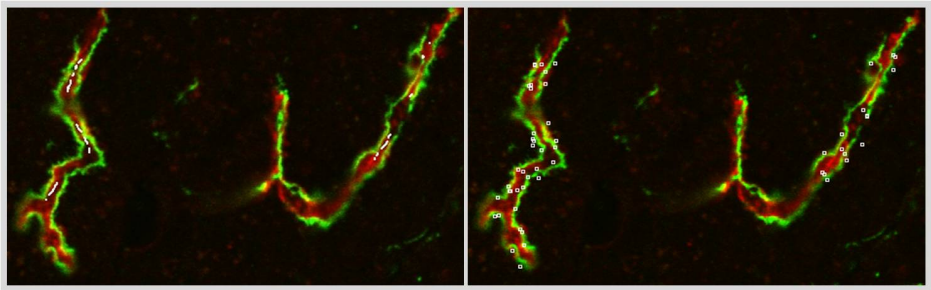


Figure 6.3.1.: Centers of accepted profiles extracted by the structure-based workflow (left) and by the random lines-based workflow (right) for the 1D-Can process.

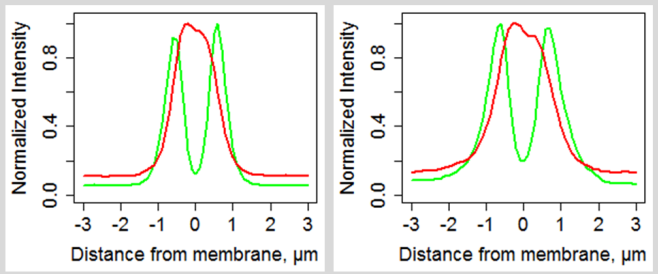


Figure 6.3.2.: Averages over accepted profiles extracted by the structure-based workflow (left) and by the random lines-based workflow (right) for the 1D-Can process.

of the accepted profiles generated by these two methods are marked. These positions show a good correlation. Furthermore, averages over the accepted profiles are also comparable, as Figure 6.3.2 illustrates. Herewith, the random profile extraction is considered to be a valid alternative to the structure-based profile extraction.

6.3.2. Comparison of the random lines-based workflow to the structure-based workflow

I showed that the random lines-based workflow extracts profiles at positions correlating well to the centers of profiles extracted by the structure-based workflow. Average profiles were also found to be similar. Further validation of the random lines-based workflow was performed using numeric descriptors. The newly defined descriptors and the statistical variance were computed and compared for the profiles extracted using both automated workflows.

Means and standard deviations of the descriptor values indicate a strong correspondence between the results from the structure-based workflow and the random lines-based workflow (see Table 6.3.2). The random lines-based workflow is considered to be valid, as parameters of the extracted profiles are distributed similarly to those extracted by the already validated structure-based workflow.

WRSTs have been performed on the descriptors calculated from the random profiles to validate the random lines-based workflow statistically. Functional marker translocation is assessed using 20 top-ranked accepted profiles. The random lines-based workflow is expected to detect a significant translocation when comparing control and test datasets, and to indicate its absence in control-control pairs. 17 datasets of images for the 1D-Can process (except z-stacks) were evaluated.

The random lines-based workflow indicated absence of a significant translocation in 5 of 8 evaluated control-control datasets (see Table B.0.1). In the dataset *TC-4-C/C*, both the structure-based and the random lines-based workflow detected a significant translocation effect.

#RandLines	# Acc Prof	$T_{profExtract}$, c	$T_{profSelect}$, sek	T_{total}, sek	μ(C)	μ(D)
Structure-based workflow						
-	204	260.3	271.8	284.7	3.01	2.84
Random lines-based workflow						
20	1.33	21.2	21.3	27.2	2.55	1.72
50	7.33	25.6	25.9	32.3	2.59	2.29
100	14.0	33.3	33.5	41.4	2.13	2.08
150	25.3	40.2	40.4	50.0	2.05	1.87
200	29.3	46.1	46.4	56.2	2.29	2.00
300	51.3	59.8	60.1	74.0	2.16	1.97
500	74.7	86.6	86.9	103.0	2.22	2.11
1000	137.7	163.1	163.8	199.2	2.26	2.13
1500	235.3	229.6	230.3	280.3	2.17	2.01

Table 6.3.1.: Comparison of the structure-based and the random lines-based workflows. Dataset *Osmol-1-C*.

Descriptor	Structure-based		Random lines-based	
	μ	σ	μ	σ
X	0.564	0.132	0.578	0.201
Y	1.13	0.382	1.03	0.512
Z	0.731	0.241	0.710	0.341
A	0.372	0.092	0.363	0.144
B	0.441	0.142	0.414	0.199
C	1.99	0.495	1.72	0.520
D	1.30	0.341	1.21	0.442
E	0.781	0.186	0.701	0.226
F	0.564	0.162	0.511	0.149
Variance	0.058	0.013	0.058	0.016

Table 6.3.2.: Descriptor values μ and σ calculated from 20 top-ranked selected profiles automatically extracted by the structure-based and the random lines-based workflow. Dataset *TC-1-T*.

Descriptor	*TC-4-C/C*		*TC-2-C/C*		*BDL-2-C/BDL-3-C*	
	Structure-based	Random lines-based	Structure-based	Random lines-based	Structure-based	Random lines-based
C	6.1 e-4	4.0 e-2	3.4 e-1	1.2 e-1	6.1 e-1	3.0 e-2
D	1.1 e-1	2.7 e-1	6.4 e-1	3.2 e-3	2.4 e-1	6.1 e-1
E	1.0 e-2	8.5 e-2	5.2 e-1	8.2 e-3	4.0 e-1	3.1 e-1

Table 6.3.3.: Significance of translocation detected in control-control datasets by the structure-based and the random lines-based workflows (on 20 top-ranked profiles). Datasets *TC-4-C/C*, *TC-2-C/C* and *BDL-2-C/BDL-3-C*.

Descriptor	*BDL-2-C/T*		*TCDC-1-C/T*	
	Structure-based	Random lines-based	Structure-based	Random lines-based
C	2.6 e-7	1.0 e-6	4.3 e-2	4.4 e-2
D	2.2 e-3	3.1 e-6	5.8 e-3	2.5 e-3
E	1.8 e-5	2.7 e-7	6.4 e-3	4.5 e-3

Table 6.3.4.: Significance of translocation detected in control-test datasets by the structure-based and the random lines-based workflows (on 20 top-ranked profiles). Datasets *BDL-2-C/T* and *TCDC-1-C/T*.

This dataset, however, does not seem to be a good example.

For two other control-control datasets (*TC-2-C/C* and *BDL-2-C/BDL-3-C*), the random lines-based workflow performed slightly worse. At least one of the selected descriptors (*C*, *D*, *E*) indicated there a significant translocation (see Table 6.3.3), while the structure-based workflow indicated its absence.

Evaluation of control-test pairs was successful for all 9 datasets. For 6 datasets, all 3 descriptors *C*, *D* and *E* detected a significant translocation (see Table B.0.2). Table 6.3.4 illustrates this on the example of the datasets *BDL-2-C/T* and *TCDC-1-C/T*. Here, the random lines-based workflow indicated comparable or even higher significance levels than the structure-based workflow. However, for most compared control-test datasets, resulting p-values were 3 - 4 orders of magnitude higher and thus less significant.

For 3 datasets (*TCDC-2-C/T*, *TCDC-4-C/T* and *BDL-1-C/T*), a significant translocation effect was indicated by 2 of 3 selected descriptors (see Table 6.3.5). Still, for all control-test datasets, at least one newly defined descriptor indicated a significant translocation effect. Hence, the random lines-based workflow for the 1D-Can process has not missed any significant translocation, while being faster than the structure-based workflow and absolutely objective.

Descriptor	*TCDC-2-C/T*		*TCDC-4-C/T*		*BDL-1-C/T*	
	Structure-based	Random lines-based	Structure-based	Random lines-based	Structure-based	Random lines-based
C	1.1 e-4	1.8 e-2	3.0 e-11	1.2 e-1	5.1 e-10	6.0 e-5
D	3.8 e-2	5.7 e-2	1.8 e-14	3.2 e-3	1.2 e-12	4.0 e-6
E	1.7 e-3	1.9 e-2	1.1 e-13	8.2 e-3	9.0 e-3	6.6 e-1

Table 6.3.5.: Significance of translocation detected in control-test datasets by the structure-based and the random lines-based workflows (on 20 top-ranked profiles). Datasets *TCDC-2-C/T*, *TCDC-4-C/T* and *BDL-1-C/T*.

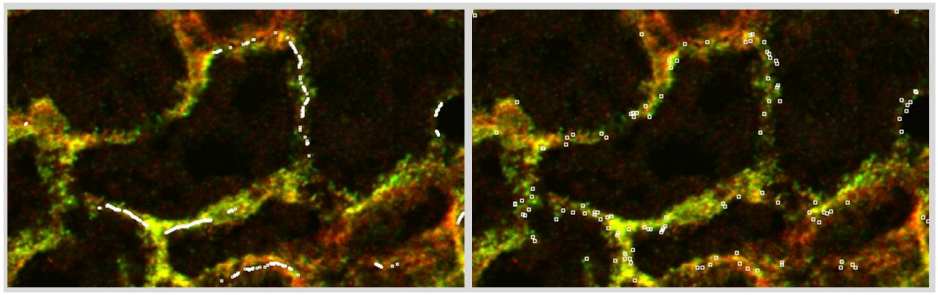


Figure 6.4.1.: Centers of accepted profiles extracted by the structure-based workflow (left) and by the random lines-based workflow (right) for the 1D-Bas process.

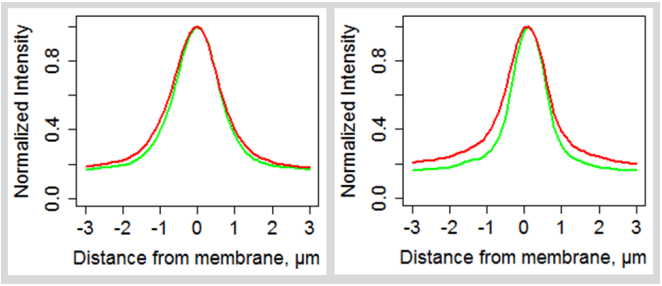


Figure 6.4.2.: Averages over accepted profiles extracted by the structure-based workflow (left) and by the random lines-based workflow (right) for the 1D-Bas process.

6.4. Validation of the Application IV: random lines-based workflow for 1D-Bas process

The random lines-based workflow has been applied for the quantification of Ntcp internalization at the basolateral membranes (1D-Bas process) and validated against the structure-based workflow.

Profile extraction parameters ($\#RandLines$ = 1000 and overlap for profile extraction = 2.4 μm) chosen for the 1D-Can process are also found to be well tuned for the 1D-Bas process. Figure 6.4.1 illustrates a small fragment of the 1^{st} image from the dataset *TC-3-C*. Centers of the accepted profiles generated by the structure-based and the random lines-based workflows are marked. These positions show a good correlation. Averages over the accepted profiles are also found to be comparable, as Figure 6.4.2 illustrates. Herewith, the random profile extraction for basolateral membranes is considered to be a valid alternative to the structure-based profile extraction.

6.4.1. Comparison of the random lines-based workflow to the structure-based workflow

I showed that distribution profiles are extracted by the random lines-based workflow at positions that correlate well to the central profile positions found in the structure-based workflow. Further statistical validation of the random lines-based workflow was performed based on descriptor values. Therefore, the datasets of images for the 1D-Bas process (see Table 6.2.1) were processed according to the random lines-based workflow. Mean (μ) and standard deviation (σ) of the descriptors were compared to these values calculated from the structure-based profiles.

A good correspondence of μ and σ values was found for two automated workflows. Table 6.4.1 illustrates this for the dataset *TC-3-CC*. C1 and C2 in this table correspond to two control datasets that constitute the dataset *TC-3-CC*. Interestingly, μ descriptor values were consis-

Descriptor	Structure-based				Random lines-based			
	μ(C1)	σ(C1)	μ(C2)	σ(C2)	μ(C1)	σ(C1)	μ(C2)	σ(C2)
L	0.633	0.185	0.630	0.168	0.728	0.179	0.736	0.165
M	0.779	0.176	0.779	0.158	0.829	0.144	0.834	0.137
N	1.46	0.855	1.40	0.706	1.95	0.920	1.94	0.843
O	0.488	0.150	0.484	0.133	0.565	0.131	0.569	0.121
P	1.81	0.822	1.75	0.683	2.31	0.914	2.30	0.842
Variance	0.052	0.024	0.052	0.017	0.061	0.016	0.059	0.013

Table 6.4.1.: Descriptor values μ and σ calculated from 20 top-ranked selected profiles automatically extracted by the structure-based and the random lines-based workflow. Dataset *TC-3-CC*.

Descriptor	*TC-3-C/C*		*TC-7-C/C*	
	Structure-based	Random lines-based	Structure-based	Random lines-based
M	1.8 e-1	7.9 e-2	1.5 e-1	5.3 e-1
N	3.0 e-1	2.7 e-1	8.6 e-2	5.7 e-1
O	5.7 e-1	1.8 e-1	5.4 e-1	6.4 e-1

Table 6.4.2.: Significance of translocation detected in control-control datasets by the structure-based and the random lines-based workflows (on 20 top-ranked profiles). Datasets *TC-3-C/C* and *TC-7-C/C*.

tently higher for the random lines-based workflow. However, for both automated workflows, μ values did not differ much between C1 and C2. This suggests that the random lines-based workflow extracts reliable information on molecule distribution. Standard deviations of descriptor values were comparable for both automated workflows.

Further validation of the random lines-based workflow was performed by statistical tests. WRSTs were performed on 20 top-ranked selected random profiles. The results were compared to the results obtained on top 20 selected structure-based profiles. The random lines-based workflow is expected to detect a significant translocation when comparing control and test datasets, and to indicate its absence in control-control pairs. Only descriptors M, N and O are considered, as they were found to be the most relevant for the evaluation of the 1D-Bas process.

For 6 control-control datasets, the results of the random lines-based workflow correlated well to the results of the structure-based workflow (see Table B.0.4). In 5 datasets (except *TC-2-C/C*), both automated workflows indicated no significant translocation of the functional marker (see examples in Table 6.4.2).

In the dataset *TC-2-C/C*, both automated workflows detected a significant translocation effect (see Table B.0.4). Only in the dataset *TC-5-C/C*, the random lines-based workflow indicated a significant translocation that was not detected by the structure-based workflow (see Table 6.4.3). Interestingly, manual workflow also indicated a significant effect in this dataset.

In 4 control-test datasets (*TCDC-1-C/T*, *TCDC-2-C/T*, *TCDC-3-C/T* and *TCDC-6-C/T*), both the random-lines based and the structure-based workflow indicated a significant translocation of the functional marker (see examples in Table 6.4.4). Significance of the translocation detected by the random-lines based workflow was generally several orders lower than for the structure-based workflow. Table B.0.5 shows the full data.

In the dataset *TCDC-5-C/T*, none of the workflows (random lines-based, structure-based, manual) indicated a significant Ntcp translocation (see Table B.0.5). Probably, preparation of the test dataset was not successful in this case.

And finally, the random lines-based workflow did not indicate a significant translocation in the dataset *TCDC-4-C/T* (see Table 6.4.3). In contrast, it was detected by the structure-based

Descriptor	*TC-5-C/C*		*TCDC-4-C/T*	
	Structure-based	Random lines-based	Structure-based	Random lines-based
M	4.3 e-1	1.5 e-2	2.7 e-2	1.0 e-1
N	3.3 e-1	1.5 e-5	3.1 e-2	1.8 e-1
O	3.3 e-1	1.2 e-4	1.2 e-2	1.3 e-1

Table 6.4.3.: The structure-based and the random lines-based workflows yield different results for the datasets *TC-5-C/C* and *TCDC-4-C/T*.

Descriptor	*TCDC-1-C/T*		*TCDC-2-C/T*		*TCDC-3-C/T*	
	Structure-based	Random lines-based	Structure-based	Random lines-based	Structure-based	Random lines-based
M	3.6 e-10	5.9 e-4	2.0 e-7	7.7 e-7	4.6 e-2	3.8 e-3
N	4.5 e-12	1.4 e-7	9.9 e-12	7.4 e-5	3.2 e-2	1.6 e-2
O	5.8 e-13	1.8 e-6	5.3 e-11	2.0 e-6	1.8 e-2	2.1 e-3

Table 6.4.4.: Significance of translocation detected in control-test datasets by the structure-based and the random lines-based workflows (on 20 top-ranked profiles). Datasets *TCDC-1-C/T*, *TCDC-2-C/T* and *TCDC-3-C/T*.

workflow. In total, the random lines-based workflow yielded results well correlating to the results of the structure-based workflow in 11 of 13 evaluated datasets.

The random lines-based workflow was found approx. 8 times faster than the structure-based workflow (see examples in Table 6.4.5). Many more profiles are originally extracted in the random lines-based workflow (e.g. approx. 55000 vs. 12000). However, extraction of numerous random profiles is still much faster than structure-based profile extraction.

6.5. Validation of the Application V: random lines-based workflow for 2D-Nuc process

The random lines-based workflow is based on profile extraction avoiding structure detection in images. Hence, this method is expected to be more generally applicable than structure specific approaches. I have tested it for the quantification of the 2D glucokinase translocation (2D-Nuc process, see Section 3.1.4) and validated it against the established structure-based workflow (see Section 3.3).

6.5.1. Datasets

Samples of rat hepatocytes were prepared by my colleagues at the University Clinic Düsseldorf. Hepatocytes were incubated with various glucose concentrations (dataset *Gluc-1*). In other experiments, hepatocytes were incubated with glucose and some added reagents (e.g. In-

Image	Structure-based			Random lines-based		
	# Profiles extracted	# Profiles accepted	T_{total}, min	# Profiles extracted	# Profiles accepted	T_{total}, min
1	12297	1810	16.02	55882	917	1.84
2	10778	3449	13.59	55722	1560	1.81

Table 6.4.5.: Structure-based workflow compared to random lines-based workflow for two images from the dataset *TC-3-C*.

Incubated with glucose	Incubated with glucose and other reagents		
	DMSO	Insulin	Reagents
Gluc-1	GlucDMSO-1	GlucIns-1	GlucR-1
-	-	GlucIns-2	GlucR-2
-	-	-	GlucR-3
-	-	-	GlucR-4
-	-	-	GlucR-5
-	-	-	GlucR-6

Table 6.5.1.: Datasets for the 2D-Nuc process.

sulin, DMSO (Dimethyl sulfoxide), etc.). The respective datasets are referred to as *GlucIns-1*, *GlucIns-2* and *GlucDMSO-1* (see Table 6.5.1). Datasets *GlucR-1*, *GlucR-2*, etc. refer to hepatocytes incubated with glucose and some *reagents*. *Reagents* encode other combinations of chemicals tested by my colleagues. This code is used as the results have not been published yet. In the dataset *GlucR-6*, hepatocytes were incubated with a constant glucose concentration but varying concentrations of the reagent 6.

10 datasets of confocal images (2048 × 2048 pixel) were acquired with the final pixel size of 110 nm. A dataset contains 12 images of different regions of the respective cultural plate. Distribution of the functional marker (glucokinase) is automatically quantified by the nucleus-to-cytoplasm ratio. This ratio is dependent on the glucose concentration in the medium (see Section 3.1.4) and on the action of reagents added. The ratio is expected to change under different experimental conditions. Datasets were processed according to the random lines-based workflow. The results were compared to those obtained by the established structure-based approach.

6.5.2. Optimal number of random lines

The optimal profile length (28.0 μm) and the required overlap for extracting individual random profiles (2.7 μm) were chosen earlier (see Section 5.6.1). In the following, other parameters for the optimal profile extraction will be selected.

The extracted information fully depends on the initial random lines, because no structure detection in images is performed in the random lines-based workflow. Therefore, the first parameter to choose is the number of these lines. Similar to the **Application III** and **Application IV**, random lines should cross enough biological structures studied, but should not extract redundant information. Random profile extraction should also not last much longer than the structure-based profile extraction.

Various numbers of random lines were evaluated. The resulting nucleus-to-cytoplasm intensity ratio and total calculation time are compared to the established structure-based workflow. Table 6.5.2 illustrates this on a sample image from the dataset *GlucIns-1*. Experiments with the same number of random lines were repeated 3 times, and mean values of the calculated parameters are reported. For this evaluation, the number of con-centric profiles, the width of the ring that is skipped around the nucleus and the width of the cytoplasm are kept constant ($\#ConcentricProfiles = 10$, $SkipArea = 1.0$ μm and $CytoplasmArea = 4.0$ μm, respectively).

First, few initial random lines (10 - 50) do not cross a visually acceptable number of nuclei in the images. Hence, resulting random profiles will not cover a representative set of nuclei. Second, average nucleus-to-cytoplasm intensity ratios change only slightly when $\#RandLines$ is varied from 50 to 150, suggesting that the random profile extraction is a robust approach. Third, when $\#RandLines \geq 100$ the computation time becomes comparable to that in the established structure-based workflow (e.g. T_{total}: 249.4 vs. 242.0 seconds). Such a number of the initial random lines needs to be chosen, that the random lines-based workflow yields a ratio comparable to the structure-based workflow, while not being much slower. Therefore, the $\#RandLines = 100$ appears to be a good compromise.

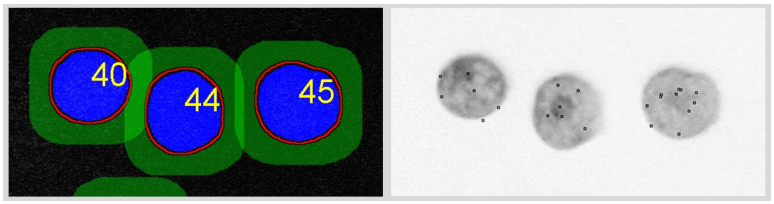


Figure 6.5.1.: Left: Structure detection in images. Blue - nucleus, green - cytoplasm, red - area to skip. Right: Centers of accepted random profiles found by the random profile extraction for the 2D-Nuc process.

Results of the random profile extraction were validated visually against the results of structure detection in the established workflow. Figure 6.5.1 illustrates a small fragment of a sample image from the dataset *GlucIns-1* processed by the structure-based workflow and the random lines-based workflow. Centers of the selected random profiles correspond well to the nuclear and cytoplasmatic regions segmented in the structure-based workflow.

6.5.3. Optimal number of con-centric profiles

Once the number of random lines is chosen, further profile extraction parameters can be optimized. The first parameter is the number of con-centric profiles ($\#ConcentricProfiles$). Various $\#ConcentricProfiles$ were tested. Table 6.5.3 reports mean parameters from 3 experiments on the earlier mentioned image from the dataset *GlucIns-1*.

First, nucleus-to-cytoplasm intensity ratio appeared to be not very sensitive to $\#ConcentricProfiles$. It varied slightly in the range 1.75 – 1.89. However, $\#ConcentricProfiles = 3$ or 10 led to the ratio closest to the ratio obtained by the established structure-based workflow (1.79 and 1.81 vs. 1.80). Second, the computation time increases with the $\#ConcentricProfiles$. $\#ConcentricProfiles = 10$ requires almost the same time as the established structure-based workflow. Third, con-centric profiles are introduced for better coverage of the nuclear and cytoplasmatic regions. Hence, too few con-centric profiles are insufficient to achieve this aim.

Figure 6.5.2 illustrates a small fragment of the sample image from the dataset *GlucIns-1* and positions of the extracted con-centric profiles depending on their number. The larger the $\#ConcentricProfiles$, the better coverage of the nuclear and cytoplasmatic regions is ensured. Therefore, the largest $\#ConcentricProfiles$ is chosen which still leads to the computation time comparable to that of the established structure-based workflow. Thus, $\#ConcentricProfiles$ is thus set to 10.

$\#RandLines$	T_{total}, sec	μ(Ratio)
Structure-based workflow		
-	242.0	1.80
Random lines-based workflow		
10	63.2	1.87
20	89.2	1.71
50	153.8	1.83
75	188.1	1.80
100	249.4	1.81
150	335.8	1.85

Table 6.5.2.: Structure-based workflow compared to random lines-based workflow for the 2D-Nuc process.

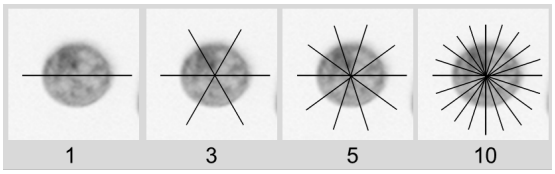


Figure 6.5.2.: Con-centric profile extraction for the 2D-Nuc process. $\#ConcentricProfiles$ is varied from 1 to 10.

6.5.4. Optimal cytoplasm area

Nuclear and cytoplasmatic regions have to be identified to quantify intensity ratio of the functional marker (see Section 5.6.3). The nucleus is detected in structural marker profiles (see Figure 5.6.6). Nuclear borders are found at those positions where the structural marker intensity decreases more than 4 times. This method yields only approximate nuclear regions. For that reason, I require some space to be skipped around the nuclear borders before the surrounding cytoplasm is *detected* (*SkipArea*). Such area is also used in the established structure-based workflow (see red regions in Figure 6.5.1). There, *SkipArea* is set to 0.5 μm. The structure-based workflow identifies nuclei more precisely than the random lines-based workflow. Thus, to ensure an even larger separation of the nuclear and cytoplasmatic regions, I set *SkipArea* = 1.0 μm in the random lines-based workflow.

The width of the cytoplasmatic region (*CytoplasmArea*) is the next parameter to be set. In the established structure-based workflow, such *CytoplasmArea* was also used and set to 4.0 μm. I have evaluated different *CytoplasmArea* in the range 2.0 - 7.0 μm and the influence of this parameter on the nucleus-to-cytoplasm ratio.

Table 6.5.4 reports mean parameters from experiments repeated 3 times on the sample image from the dataset *GlucIns-1*. The intensity ratio varied slightly when *CytoplasmArea* was changed. The ratio is thus found to be robust to the changes in the *CytoplasmArea*. Therefore, the *CytoplasmArea* is set to 4.0 μm, as no benefits for other values have been detected. This *CytoplasmArea* is also found to be small enough, so that *detected* cytoplasm areas minimally overlap with other nuclei.

Evaluation of the other datasets supported $\#RandLines = 100$, $\#ConcentricProfiles = 10$ and *CytoplasmArea* = 4.0 μm as well tuned parameters for quantification of the 2D-Nuc process. Table 6.5.5 illustrates this for the datasets *GlucR-1* and *GlucR-2*.

6.5.5. Comparison of the random lines-based workflow to the structure-based workflow

I showed that the random lines-based workflow extracts distribution profiles at positions well corresponding to the nuclear and cytoplasmatic regions detected by the structure-based workflow. Further, statistical validation of the random lines-based workflow was performed. There-

$\#ConcentricProfiles$	T_{total}, sec	μ(Ratio)
Structure-based workflow		
-	242.0	1.80
Random lines-based workflow		
1	131.0	1.89
3	128.1	1.79
5	158.5	1.75
10	249.4	1.81

Table 6.5.3.: Extended profile extraction for quantification of the 2D-Nuc process.

CytoplasmArea, μm	μ(Ratio)	σ(Ratio)
2.0	2.19	0.90
3.0	2.24	0.92
4.0	2.21	0.95
5.0	2.27	0.80
6.0	2.19	0.89
7.0	2.06	0.88

Table 6.5.4.: Influence of *CytoplasmArea* on the nucleus-to-cytoplasm ratio for quantification of the 2D-Nuc process.

g/L glucose	*GlucR-1*				*GlucR-2*			
	Structure-based		Random lines-based		Structure-based		Random lines-based	
	μ	σ	μ	σ	μ	σ	μ	σ
0	2.26	0.80	2.44	0.99	4.01	1.15	4.28	1.70
0	2.76	0.76	2.91	1.17	3.47	1.08	3.69	1.52
1	1.73	0.41	1.69	0.56	2.05	0.56	2.04	0.72
1	1.90	0.55	1.81	0.76	2.82	0.83	2.86	1.31
2	1.26	0.16	1.27	0.27	2.61	0.94	2.88	1.38
2	1.22	0.16	1.23	0.25	2.43	0.87	2.75	1.28
3	1.02	0.13	1.05	0.22	2.93	1.13	3.09	1.68
3	0.89	0.09	0.88	0.17	3.16	0.95	3.13	1.31
4	0.99	0.10	0.97	0.21	3.08	0.91	3.07	1.28
4	1.11	0.13	1.14	0.23	3.21	1.15	3.04	1.57
4.5	0.94	0.18	0.95	0.23	3.08	0.99	3.04	1.42
4.5	0.93	0.09	0.93	0.17	1.96	0.74	2.01	0.99

Table 6.5.5.: Nucleus-to-cytoplasm ratios obtained by the random lines-based workflow and the established structure-based workflow. Datasets *GlucR-1* and *GlucR-2*.

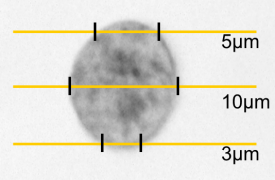


Figure 6.5.3.: Profiles cover various width of the nuclear region.

fore, nucleus-to-cytoplasm intensity ratios calculated from the accepted random profiles were compared to the ratios calculated by the established structure-based workflow. The ratios calculated by these workflows were comparable and correlated well for all datasets (see Table B.0.6). The standard deviation of ratios obtained by the random lines-based workflow was consistently higher. Table 6.5.5 illustrates this on the example of the datasets *GlucR-1* and *GlucR-2*.

6.5.6. Restriction of nuclear width for profile selection

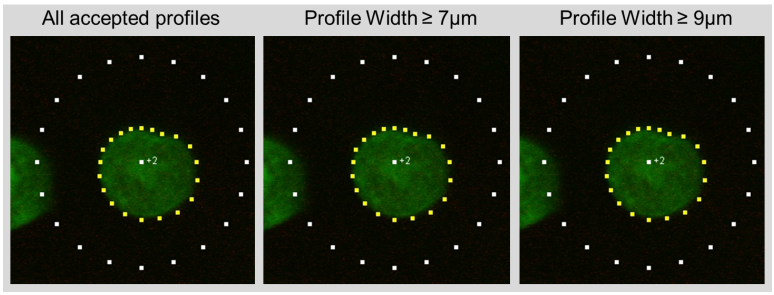
The selection criteria for structural marker profiles were presented earlier (see Section 5.6.2). Only profiles with an intensity peak in the central part and relatively symmetrical flat tails are accepted. The results obtained illustrate that the defined selection procedure allows calculating reasonable nucleus-to-cytoplasm intensity ratios (e.g. Table 6.5.5). Nevertheless, I investigated the influence of another selection criterion.

The nuclear region is detected in every profile (see Section 5.6.3). The identified nuclear region is now additionally compared to a width threshold. Profiles with too narrow nuclear regions are rejected. Figure 6.5.3 illustrates the motivation for such a selection. Three shown profiles cross a nucleus at different positions. Varying nuclear and cytoplasmatic regions are *covered* by these profiles. As nucleus-to-cytoplasm intensity ratio has to be calculated, intensity profiles are expected to *cover* a large part of the nucleus, and not only regions close to their borders. Therefore, the profile extracted in the middle yielding a nuclear width approx. 10 μm is more reliable for such an evaluation, than two other profiles (yielding widths approx. 5 and 3 μm).

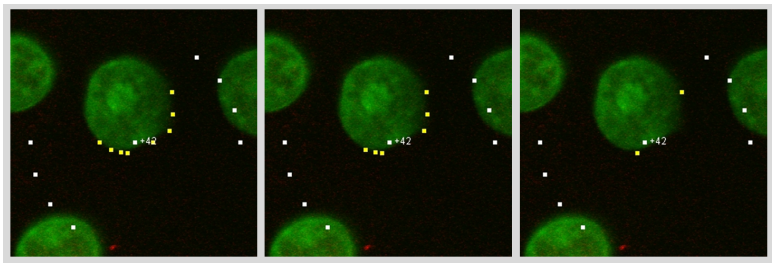
A threshold is required to filter out profiles crossing nuclei too close to their borders. As size of hepatocyte nuclei varies in rats in the range of 10 – 15 μm, I evaluate 7 μm as the minimal nuclear width that has to be covered by an acceptable profile. A width of 7 μm allows coverage of at least half width of a large nucleus. 9 μm are also evaluated to find potential benefits from an even more restricted profile selection.

All datasets were processed using the restricted profile selection. Figure 6.5.4 illustrates some example positions of the extracted profiles. Centers and ends of the profiles are marked by white dots. Yellow dots indicate nuclear regions detected in these profiles. Figure 6.5.4, (a) shows profiles generated by the extended profile extraction centered in the nucleus. All their detected nuclear regions are wider than 9 μm and will thus be accepted. Figure 6.5.4, (b) shows profiles extracted close to the border of a nucleus. One of them is narrower than 7 μm and is rejected first. Two further profiles are narrower than 9 μm and are rejected later.

Restriction of the profile width has slightly increased the calculated nucleus-to-cytoplasm ratios (see an example in Table 6.5.6). Standard deviation of the ratios was not found to decrease. Difference of ratios calculated for different experimental conditions was comparable for all profiles and for those with the restricted width (see an example in Figure 6.5.5). The random lines-based workflow is thus found to be robust for the quantification of the 2D-Nuc process. Profile width was not restricted in further experiments as its benefits were not detected.



(a)



(b)

Figure 6.5.4.: Restriction of the nuclear region width to ≥ 7 or 9 μm. White points - profile start and end, yellow points - detected nuclear region.

g/L glucose	Structure-based		Random lines-based					
	μ	σ	All widths		Width ≥ 7 μm		Width ≥ 9 μm	
			μ	σ	μ	σ	μ	σ
0	2.72	0.91	2.71	1.25	2.87	1.25	3.06	1.25
1	1.97	0.65	1.86	0.84	1.90	0.84	1.99	0.85
2	1.37	0.32	1.42	0.44	1.44	0.44	1.48	0.45
3	1.30	0.18	1.30	0.27	1.32	0.27	1.32	0.26
4	1.19	0.15	1.20	0.24	1.21	0.24	1.23	0.24
4.5	1.10	0.13	1.10	0.24	1.10	0.24	1.10	0.24

Table 6.5.6.: Glucokinase nucleus-to-cytoplasm ratios obtained with or without nuclear width restriction. Dataset *GlucDMSO-1*.

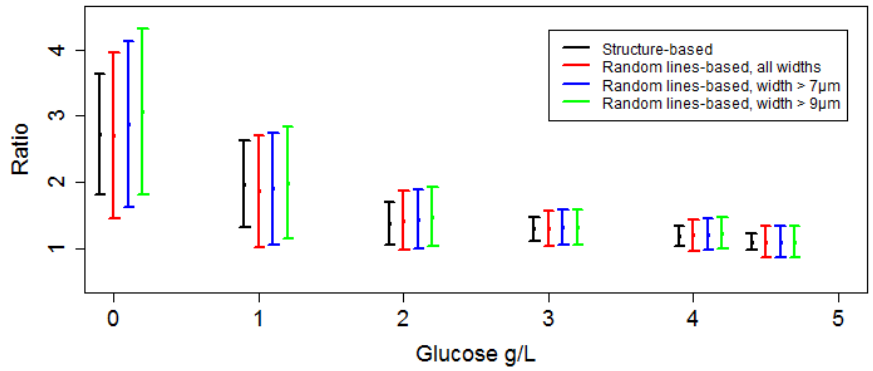


Figure 6.5.5.: Glucokinase nucleus-to-cytoplasm ratios obtained with or without nuclear width restriction. Dataset *GlucDMSO-1*.

7. Discussion

The lack of generally applicable translocation quantification approaches motivated to develop an automated workflow for this purpose. In the course of the work, translocation processes were analyzed. In particular, I defined a model for the widespread directed translocations proceeding orthogonally to a membrane structure (1D optimal transport model). This enabled the development of the automated workflow that is expected to be applicable to quantitatively analyze all such translocations.

Two automated translocation quantification workflows were implemented: one based on structure detection in images (structure-based workflow) and one avoiding this operation (random lines-based workflow). Both methods were validated against the established analyses (either manual or automated). My automated workflows are not optimized for any particular type of a biological structure and should thus be applicable to many translocation processes.

Furthermore, the random lines-based workflow has a potential for evaluations beyond the translocation quantification. Owing to the implemented content-independent information extraction strategy, various biological (and other) images may be evaluated.

7.1. Structure-based workflow for 1D translocation

The first automated workflow was developed for two model biological processes (1D-Can and 1D-Bas) corresponding to the optimal 1D transport model. Structure detection is performed in images using a machine learning algorithm. Although manual training of the foreground detection is required, the applied algorithm is not optimized to any particular biological structure.

7.1.1. Image processing and profile extraction

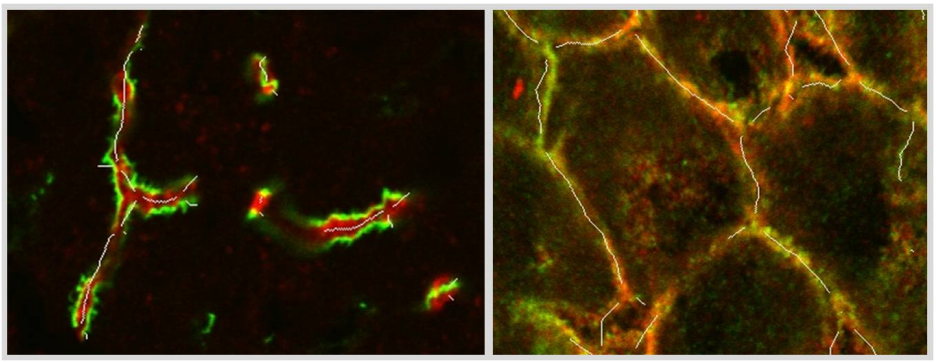
The results of the structure detection were always reliable according to the visual analysis. Image processing cleaned the detected foreground regions by morphological operations and thresholding. These procedures eliminated small membrane fragments and noise. Further, skeletonization was performed to yield medial axis of the membranes. Skeletons were mostly created at valid positions avoiding damaged biological structures (see Figure 7.1.1, (a)). The skeletons were not always perfectly centered in membranes because structure detection is not trained to produce regions perfectly fitting to the membrane borders. Consequently, the extracted distribution profiles were not always perfectly centered in the membranes. However, this is not problematic, because zones are defined relative to peaks in the profiles.

7.1.2. Profile selection and ranking

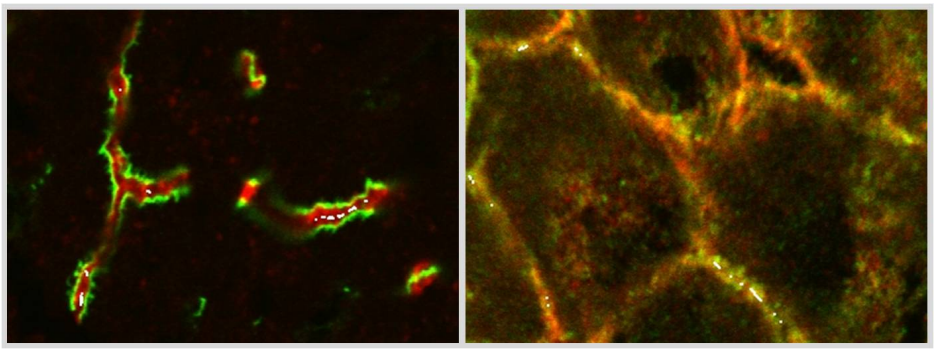
The number of the automatically extracted profiles almost always significantly exceeded the number of the manually extracted data points ($n = 10$). Therefore, among numerous extracted profiles those could be selected which more clearly depict the biological structure.

Profile selection led to meaningful visual and statistical results. Skeleton pixel positions corresponding to damaged or asymmetrical regions were filtered out (see Figure 7.1.1, (b)). Translocation evaluation based on selected profiles only was found to be more reliable and yield higher significance levels compared to the evaluation of all extracted profiles.

Only objective criteria based on general biological and physical knowledge were used for the profile selection. Still, these were found to be sufficient to clean the extracted data from noise and to produce reliable statistical results.



(a)



(b)

Figure 7.1.1.: Central positions of extracted (a) and accepted (b) profiles (white). Canalicular (left) and basolateral (right) membranes.

Construction of the profile selection criteria represented a challenging task. Expert knowledge and decision whether to evaluate a certain membrane fragment or not had to be expressed in formulas. This was possible only owing to the close collaboration with biologists and doctors from the University Clinic Düsseldorf. They routinely evaluate biological images and thus could provide us with numerous examples of acceptable and unacceptable biological structures. These data are essential for the *translation* of human decisions into mathematical expressions. Hence, the presented selection criteria were developed to follow the human expert logic.

Selected profiles corresponded better to the considered biological structure than all extracted profiles. This indicates that only profiles extracted at undamaged symmetrical canalicular membranes were selected. The number of selected profiles was still sufficiently large to perform further data sorting. Therefore, the selected profiles were ranked by structural quality.

Significance of the detected translocation effect was higher on the top-ranked profiles than on all selected profiles. 20 top-ranked profiles were found to be optimal for the quantification. Significance levels were higher when evaluated on top 20 than on the larger numbers of profiles. Hence, ranking achieved its purpose by increasing the statistical significance with the ranked structural quality of the profiles. The ranking procedure is thus found to be robust.

Interestingly, top 10 profiles led to either equal or lower significance levels than top 20 profiles. Even a false positive translocation was detected on top 10 profiles. Selecting only 10 profiles for the evaluation thus seems to be a too strict limitation.

7.1.3. Structure-specific descriptors vs. statistical variance

Biologically meaningful zones were identified in the profiles and enabled creation of process-specific descriptors. Depending on a particular translocation process, different descriptors were found to be more relevant for the quantification. For example, among descriptors developed for the 1D-Can process, those were found to be less relevant that include intensity in the central canalicular part. The central part of the canalicular membrane seems to be a region where distribution of the functional marker varies significantly even among control datasets. Therefore, descriptors including this zone were not representative for Bsep translocation quantification. In turn, when assessing Ntcp translocation (1D-Bas process), those descriptors performed better which included the integral intensity in the central part of the membranes. This illustrates the advantage of my descriptor-based approach. Numerous descriptors can be developed for any biological structure in question to optimally quantify the translocation. The number of potential descriptors is not limited, and maybe even better combinations of zonal intensities could be created for considered translocation processes.

Almost all newly defined descriptors are based on fixed-length zones. The only descriptor (F for the 1D-Can process) is based on zones of varying length that are every time computed for a particular profile. In total, this descriptor performed worst for the quantification of the 1D-Can process. It led to the highest false positive rate in control-control datasets and was among 3 descriptors leading to the highest false negative rate on control-test datasets. This suggests that intensity comparison between different images is unreliable using zones of varying length. Therefore, the usage of the descriptors based on the fixed-length zones is supported.

The developed descriptors were shown to be more relevant for the translocation quantification than the statistical variance that was used in previous works. Hence, subdivision of profiles into zones and combination of integral intensity values into descriptors is advantageous. This also supports the general applicability of the zone model and descriptors: these can be created without any assumptions on the mathematical model of the molecule distribution function.

7.1.4. Different microscopy types

I compared wide-field and confocal fluorescence microscopy images regarding the information content for the translocation quantification. Clear advantages of the confocality were found in my experiments. Interestingly, working with all confocal layers from z-stacks was not an

optimal solution. Redundant information and low quality of the data from the uppermost and the bottom layers are the possible reasons.

The best confocal layer provided sufficient information for the translocation detection. Significance levels were further increased by evaluating averages of 3 neighboring original confocal layers. The thickness of an optical layer of approx. $3 \times 0.4\ \mu\text{m} = 1.2\ \mu\text{m}$ was found to be suitable for the translocation quantification. This setting can be used when acquiring data by a conventional wide-field microscope.

Advances of microscopy allowed investigating sophisticated biological questions which would be impossible without high resolution and/or 3D analysis. Nevertheless, there are still applications for which conventional 2D wide-field microscopy is sufficient. In general, the simplest, the fastest and the cheapest working method has to be found for every purpose in order to avoid excessive costs, both in time and money.

7.1.5. Structure-based workflow vs. manual workflow

The developed structure-based workflow was validated against the established manual workflow. First, descriptor values calculated from the automatically and manually extracted profiles were compared. A good correlation was found. As descriptors are parameters of the distribution profiles, the respective profiles are also considered to be comparable.

Second, statistical results obtained from the automatically extracted and selected profiles were validated against the results from the manually obtained data. False negative translocation was indicated by the manual workflow in 1 of 9 control-test datasets (taken together 1D-Can and 1D-Bas processes). In one dataset, both the manual and the structure-based workflow indicated no significant effect. This dataset, however, does not seem to be a good example. Herewith, the structure-based workflow has not missed any significant translocation that was indicated by the manual workflow. Significance levels were higher either in one or another workflow.

The results of these workflows differed more for control-control datasets. In 2 of 7 datasets, both the structure-based and the manual workflow indicated no significant translocation. In 4 datasets, the structure-based workflow indicated no effect, while a significant translocation was detected in the manually extracted data. Only in one control-control dataset, the structure-based workflow indicated a significant translocation that was not detected manually. Summarizing these findings, I consider that the structure-based workflow performed better than the manual workflow in my experiments.

Third, the structure-based workflow is much faster than the manual workflow. Fourth, it yields objective and reproducible results.

7.1.6. Validation by biological methods

The statistical results of the sample preparation, labeling, microscopy and image processing were compared to those yielded from biological evaluation. In particular, my colleagues from the University Clinic Düsseldorf performed flow cytometry to evaluate the functional marker translocation.

As presented in Mühlfeld et al. [111], HepG2 cell clones expressing Ntcp with an intracellular EGFP- and an extracellular FLAG-tag were prepared. The cells were treated with bile salts (TC and TCDC) and analyzed by flow cytometry. Ntcp in the basolateral membrane was quantified by the extracellularly localized FLAG-tag in unpermabilized cells. Total Ntcp was measured by EGFP-fluorescence. Fluorescence intensity distributions are shown in cumulative histograms. The total amount of Ntcp was assessed by the fluorescence of the intracellular EGFP-tag. It has changed neither under TC, nor under TCDC treatment (Figure 7.1.2, (b)). In turn, the extracellular FLAG-associated fluorescence was reduced by TCDC (Figure 7.1.2, (a)). This suggests that TCDC reduces Ntcp at the basolateral membrane by retrieval into intracellular compartments [111]. Herewith, the functional marker translocation under TCDC but not under TC treatment is independently validated by a biological method.

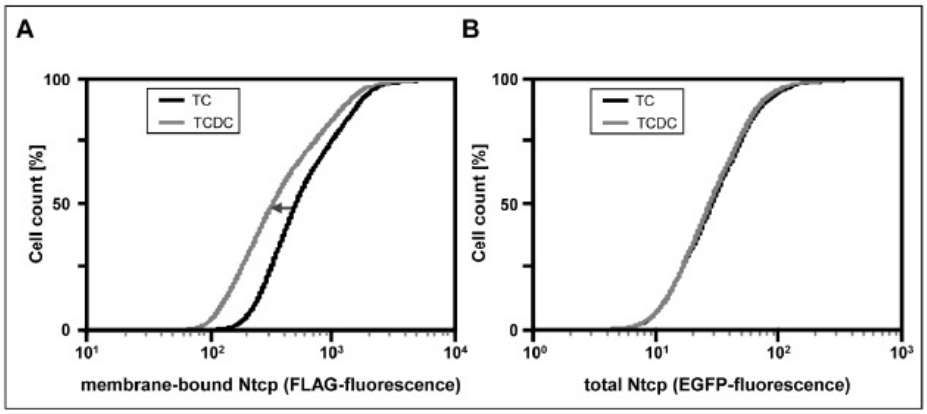


Figure 7.1.2.: Flow cytometry results (from Mühlfeld et al. [111]).

7.2. Random lines-based workflow for 1D translocation

The aim of this thesis is to develop a generally applicable translocation quantification workflow not being optimized to any particular biological structure. To achieve this, the most sensitive step in the developed structure-based workflow, the structure detection, was omitted.

7.2.1. Advantages

The structure-based workflow includes two structure detection steps. First, structure detection in images is performed. Further, profiles are extracted in relation to the membranes detected. Extracted profiles undergo a selection procedure which requires a signal detection in every profile (of the structural marker). The random lines-based workflow optimizes this approach. It avoids structure detection in images by a different information extraction strategy. Random lines are drawn through the images, and distribution profiles are extracted. It makes the whole workflow more robust and generally applicable. Signal detection is now performed only on the extracted 1D profiles. This enables processing of images of almost any biological structure that contain membranes or separated compartments. For instance, neurite outgrowth might be analyzed using the statistical data extracted along numerous random lines.

Avoiding machine learning, which is used in the structure-based workflow for the foreground detection, excludes subjective training of the process. And finally, working with numerous random lines permits parallel calculations on several CPUs saving time.

7.2.2. Random lines-based workflow vs. structure-based workflow

The random lines-based workflow was compared to the structure-based workflow. Depending on the settings, either one could be faster for the 1D-Can process. When the training time of the foreground detection is included, the random lines-based workflow was mostly faster. For the 1D-Bas process, the random lines-based workflow was approx. 8 times faster.

The difference in computation time between the two automated workflows is greater for the 1D-Bas process than for the 1D-Can process due to the membrane structure. The total area of the basolateral membranes depicted in a standard image is much larger than the area of the canalicular membranes. Therefore, the structure-based workflow requires more time for structure detection, image processing, skeletonization and extraction of profiles in relation to the membranes detected. The random lines-based workflow, in turn, does not become much slower because of the presence of much more biological structures. This illustrates the advantage of the random lines-based workflow.

Statistical results of the structure-based and the random lines-based workflows were compared. In 9 of 15 control-control datasets, none of these workflows indicated a significant translocation. In one dataset, both indicated a significant effect. In the rest 5 control-control datasets, only one of the automated workflows detected a significant effect.

For control-test datasets, the results were better. A significant translocation was indicated by both the structure-based and the random lines-based workflows in 13 of 15 datasets. For one dataset, all workflows (including the manual) indicated no effect. Results of the two automated workflows were contradictory only in one control-test dataset (*TCDC-4-C/T* for 1D-Bas process).

Performance of the random lines-based workflow seems to be better on control-test datasets. Further evaluations have to be performed to better validate my automated approaches. Availability of reliable biological data is a core requirement for this.

Published application of the structure-based workflow

Mühlfeld et al. [111] applied the structure-based workflow for quantitative analysis of the 1D-Can and 1D-Bas processes. For their experimental setup, it was required that the membranes under study remain unchanged in test datasets, relative to control datasets. To achieve this, the generic structure-based workflow was used to monitor distribution changes of the structural markers by calculation of structure-specific numeric descriptors. Quantitative analysis of the descriptor values allowed filtering out those datasets, where the membrane structure has changed after the treatment. This application illustrates again the powerful approach implemented in the automated workflows. Numeric descriptors can be calculated for any stained protein of interest to perform data analysis or processing according to the required scheme.

As the focus of my dissertation is the development of the automated workflows and not a validation of biological effects, I have not performed such filtering of the data.

7.3. Random lines-based workflow for 2D translocation

The advantage of the random lines-based workflow is its wide application area. I conducted successful tests on a different type of biological structure and process. In particular, 2D nucleus-to-cytoplasm translocation of glucokinase (2D-Nuc process) was quantified.

Profile extraction and selection were performed according to the same scheme as in the previously discussed 1D cases. Only structure specific parameters had to be changed, such as profile length, for instance. Signal detection was carried out on 1D profiles and was easily adapted based on the knowledge on the biological structure. The workflow itself, however, did not have to be changed.

7.3.1. Random lines-based workflow vs. structure-based workflow

The random lines-based workflow was compared to the structure-based workflow. In the considered 2D case, a nucleus-to-cytoplasm ratio of the functional marker intensity is evaluated. The ratios computed by these two methods correlated very well, so that the random lines-based workflow was validated. Furthermore, it has the same advantages, as described for the 1D case: it is objective, reproducible, fast, and can be calculated on parallel CPUs.

8. Conclusions

Technological advances led to the development of various techniques producing gigabytes of data in a high-throughput manner. For instance, in biological research, methods like sequencing, micro-array analysis, imaging, in particular automated microscopy, and many others generate data that have to be analyzed to access the valuable information.

While generally applicable methods have been developed for some biological questions, like colocalization assessment, there are other less investigated areas. Translocation quantification is one of them. Although translocation processes are found in all living organisms, only few assessment methods have been reported so far. Many of them are manual analyses or special software optimized for a particular biological process and structure. Generally applicable automated translocation quantification methods are still to be created.

In the course of this work, I have developed two automated workflows for the translocation quantification. The first one includes structure detection in biological images (structure-based workflow), while the second one omits it (random lines-based workflow). Both these workflows were tested on the model 1D and 2D translocation processes. They were found to be valid, as statistical results comparable to the established quantification methods were produced.

8.1. Application area

To develop the automated workflows, I defined a general model for a range of translocation processes. I also limited the input data to images. Images are a widely used biological information source and image acquisition techniques are greatly developed.

Several requirements are set to the translocations of biomolecules to be quantifiable by my automated workflows. First, a functional marker has to be detectable in the input images. It could be, for example, immunohistochemically labeled.

Second, translocation of the functional marker should include a biological barrier, e.g. a membrane structure. The translocation can either proceed across this membrane, or represent an integration into it. Availability of such a reference structure is essential for the quantification based on the images of different samples. As membrane structures are widespread in living organisms, this condition should not limit the application area too much. Furthermore, membranes delineate organelles, where concentrations of biomolecules can differ from those in the cytoplasm. Consequently, concentration gradients run across the membranes indicating possible translocation trajectories.

Third, the reference structure should be also detectable in the input images. Therefore, one of its components (a structural marker) must be taggable. I claim that for almost any membrane such a structural marker can be found. For instance, it may be a structural protein with constitutive expression and unchanged localization under various experimental conditions.

Fourth, the developed workflows assume that translocation proceeds along the concentration gradients according to the presented 1D optimal transport model. Numerous processes meet this requirement. It also motivates the information extraction along these gradients. Directions of the gradients are identified orthogonally to the membrane structures.

Despite these limitations, most translocation processes should still be quantifiable by the developed workflows.

8.2. Comparison of the established manual workflow and the developed automated workflows

The existing manual workflow was taken as a reference in this thesis. The developed automated workflows were compared to it, as well as to each other.

The manual workflow was found to be the slowest, apart from being subjective and error-prone simply due to a human factor. Still, significance levels indicated by the manually extracted data were mostly the highest. The automated workflows, in turn, are fast, robust and objective. The random lines-based workflow was mostly faster than the structure-based workflow. Their statistical results correlated similarly to those from the manual workflow. However, significance levels were mostly several orders lower. Assuming that the manual workflow yields correct results, in most datasets both novel workflows also correctly indicated presence or absence of a significant effect (10 and 12 of 16 datasets for the structure-based and the random lines-based workflows, respectively).

The random lines-based workflow has some significant advantages over the structure-based workflow. First, a complicated image processing operation such as structure detection is avoided. It significantly reduces the programming effort and omits inclusion of frequently too specific models. The method is more generally applicable, as it is not optimized to any type of biological structure. A successful application of the random lines-based workflow to the 2D glucokinase translocation illustrated that. Second, the random lines-based workflow can be computed on parallel CPUs making the calculations even more effective. Significance levels of the detected translocation were mostly lower for the random lines-based workflow. Still, the false negative and false positive rates were lower than in the structure-based workflow.

Table 8.2.1 summarizes these findings. Qualities are ranked by (+), (++) and (+++), the latter being the best. The random lines-based workflow outperforms the other two methods when considering all characteristics together.

8.3. Achievements

The developed automated workflows represent an alternative to the previously used manual workflow. Application of the automated methods saves the valuable experts' time which can be invested into other tasks.

1D model translocations of biomolecules meeting the requirements were successfully quantified by the developed workflows. Classical problems arising during segmentation of biological structures were avoided in the random lines-based workflow, making it faster and even more generally applicable. Positive results for the 2D model translocation process represent an evidence for this. All three considered cases illustrated plausibility of the theoretically suggested generic translocation quantification.

To more extensively test the validity of the developed workflows, numerous further experiments should be conducted and validated against the manually extracted data. However, I expect that almost any translocation process corresponding to the optimal 1D transport model and meeting all the requirements (see Section 8.1) will be quantifiable by the developed workflows.

Feature	Manual workflow	Automated workflow	
		Structure-based	Random lines-based
Time	+	++	+++
Significance level	+++	++	+
Programming effort	+++	+	++
General applicability	+	++	+++

Table 8.2.1.: Comparison of the translocation quantification methods.

9. Outlook

Translocation processes are found in every living organism and are crucially important for the physiological balance. Therefore, translocation quantification may shed light on vital biological processes.

The first step of my future work may be a development a generic application package for the translocation quantification. The automated workflows implemented in the course of this work can serve as prototypes. Provided that a translocation process in question meets all the requirements (see Section 8.1), following steps could be performed in a GUI:

- Set location of the image data, including datasets to be compared.
- Set data extraction parameters depending on the image data and biological structures studied:
  - Image resolution,
  - Profile length,
  - #RandomLines,
  - #CircularProfiles,
  - Overlap for extraction of random profiles.
- Set profile selection parameters depending on the biological structures studied:
  - Need previous filtering of the data?
  - How does a *good* signal look like (e.g. number of intensity peaks)?
  - Allowed width of the signal,
  - Allowed noise level,
  - Any other parameters derived from biological or physical knowledge.
- Set profile ranking parameters depending on the biological structures studied. These may be some of the profile selection criteria.
- Define a zone model (biologically meaningful zones):
  - How many?
  - Of which width?
  - Localized where?
- Define structure specific descriptors. These may be, for instance, ratios of zonal integral intensity values.
- Select a statistical test to be used and a significance level (e.g. Wilcoxon rank sum test, p-value ≤ 0.05).
- Select sampling options:
  - How many descriptor values should be randomly selected from a dataset?
  - How many times should this sampling be performed?
  - Report all p-values, their median?

- Select data files to be written out:
  - Profile information,
  - Graphics, plots, etc.

Second, my workflows could be adapted to quantify processes in human liver tissue. Rat liver tissue, used for the development, is quite different from human samples. Human liver is not as homogeneous, as the one from rat due to unhealthy life style or genetic disorders. Autofluorescence of the human tissue is also much stronger that in rats. Hence, structure specific parameters may have to be optimized. For instance, canaliculi in human livers were found to be approx. 0.4 μm thinner than in rat. Therefore, zone model for canalicular membranes may have to be optimized respectively.

Third, physiologically different regions of liver tissue may be analyzed separately. In other words, translocation may be evaluated individually in each region. Therefore, regions of interest (ROI) shall be automatically identified in tissue sections by specific markers. Such ROIs may be regions close to the portal vein, liver artery or bile ducts. Protein distribution will be analyzed in each of these regions in order to compare the intra- and interregional variation.

Fourth, the developed workflows could and should be tested for other biomolecule translocations. For example, translocation of other membrane transporter proteins in other tissues may be quantified. Fifth, the developed random lines-based workflow may be applied for other biological questions beyond the translocation. Valuable statistical data may be gathered by information extraction along numerous random lines drawn across images. For example, neurite outgrowth may be assessed with this method.

And finally, the main objective is to introduce my algorithms into clinical practice. They may assist experts who routinely evaluate biological images by saving their time and/or providing an orientation point for analysis. To achieve this, my program shall be integrated in a hardware-software platform TopoScan [165], developed at the Fraunhofer Institute for Applied Information Technology (FIT). TopoScan can automatically acquire microscopic images which then shall be automatically evaluated by my algorithms. The developed workflows possess the qualities required for high-throughput experiments. Therefore, they may be used for translocation evaluation on a large scale.

A. Detailed Calculations

A.1. Bilinear interpolation

Digital images provide pixel-wise intensity information. However, intensity of intermediate pixel positions may also be required by some algorithms. It can be calculated from the given intensity raster using the bilinear interpolation [130]. Bilinear interpolation is one of the basic resampling techniques for computer vision and image processing. It can be applied for texture mapping, scaling up or transformation of an image [154]. A given pixel raster contains information on intensity at the positions (x_1,y_1),(x_1,y_2),(x_2,y_1) and (x_2,y_2) equal to I_{11},I_{12},I_{21} and I_{22}, respectively. Intensity I_{xy} at the position $(x,\ y)$ is unknown, but can be computed according to the following equations [131]:

$$I_{R_1} \approx \frac{(x_2 - x) \times I_{11} + (x - x_1) \times I_{21}}{x_2 - x_1}, \tag{A.1.1}$$

$$I_{R_2} \approx \frac{(x_2 - x) \times I_{12} + (x - x_1) \times I_{22}}{x_2 - x_1}, \tag{A.1.2}$$

$$I_{xy} \approx \frac{(y_2 - y) \times I_{R_1} + (y - y_1) \times I_{R_2}}{y_2 - y_1}, \tag{A.1.3}$$

where I_{R_1} and I_{R_2} are the intensities computed by a linear interpolation at the positions $R_1(x,y_1)$ and $R_2(x,y_2)$, respectively. They are linearly interpolated again to yield the required intensity at the position $(x,\ y)$ I_{xy}. Hence, the bilinear interpolation consists of two linear interpolations performed sequentially.

A.2. Statistical tests

Numerous statistical tests have been developed for different question settings, variable features, etc. For example, some tests require the knowledge on variable distribution. Some even require this distribution to be normal. However, this information is not always available. Some tests also require the compared samples to be paired. In other words, every measurement from one sample has to correspond to a measurement from another sample. In biological experiments, it is not always the case, unless time series are studied. In our experiments, variables from different biological samples are compared which do not belong to a time series. The prior knowledge on variables distribution is also lacking. Consequently, only unpaired generally applicable tests can be used. One of the tests meeting all these requirements is the *Wilcoxon Rank Sum Test* [174]. It is a widely used nonparametric test [99] which compares observations from the independent sets. WRST was also extended to arbitrary sample sizes. This test estimates equality or inequality of two distributions [131]. The null hypothesis states that the distributions of both groups are equal. To test this, the mean rank of one distribution in a combined sample of both distributions is computed [131] as follows:

1. Two samples (*S1* and *S2*) with *n1* and *n2* observations, respectively, are compared. These observations are combined and ranked in ascending order.

2. The sum of the ranks of the *S1* observations is calculated (*R1*). The sum of the ranks of the *S2* observations is now determined (*R2*), as the total size of the combined sample

$(n = n1 + n2)$ is known:

$$R2 = \frac{n \times (n+1)}{2} - R1. \tag{A.2.1}$$

3. Test statistic U is calculated as:

$$U1 = R1 - \frac{n1 \times (n1+1)}{2}, \tag{A.2.2}$$

$$U2 = R2 - \frac{n2 \times (n2+1)}{2}. \tag{A.2.3}$$

4. Distribution of U under the null hypothesis is known. It is tabulated for small sample sizes. For n $>$ 20 it is well approximated by the normal distribution. The $min(U1, U2)$ is used to obtain the p-value.

I performed WRSTs using the statistical software R [132].

B. Tables

Dataset	Workflow	Descriptor									
		X	Y	Z	A	B	C	D	E	F	Variance
TC-1-C/C	Structure	3.0 e-2	2.3 e-1	4.4 e-1	8.0 e-2	4.0 e-1	6.0 e-1	4.0 e-1	4.0 e-1	2.0 e-2	2.0 e-2
	Random lines	4.4 e-1	3.7 e-1	2.1 e-1	4.1 e-1	2.7 e-1	2.0 e-1	2.4 e-1	1.3 e-1	3.0 e-2	6.6 e-1
TC-2-C/C	Structure	5.7 e-1	4.2 e-1	5.8 e-1	6.5 e-1	5.7 e-1	3.4 e-1	6.4 e-1	5.2 e-1	4.0 e-2	1.2 e-1
	Random lines	3.2 e-2	7.4 e-2	3.5 e-3	4.4 e-2	1.3 e-2	1.2 e-1	3.2 e-3	8.2 e-3	1.6 e-3	5.3 e-1
TC-3-C/C	Structure	3.0 e-2	4.7 e-1	5.8 e-1	3.4 e-1	5.2 e-1	2.2 e-2	2.5 e-2	9.8 e-3	2.2 e-1	5.4 e-1
	Random lines	1.9 e-1	4.7 e-1	6.3 e-1	3.0 e-1	5.0 e-1	6.2 e-1	5.6 e-1	6.7 e-1	3.9 e-1	4.5 e-1
TC-4-C/C	Structure	2.1 e-4	8.8 e-7	1.4 e-5	8.9 e-6	1.4 e-6	6.1 e-4	1.1 e-1	1.0 e-2	9.3 e-11	1.9 e-2
	Random lines	1.5 e-4	1.6 e-3	1.4 e-3	9.5 e-4	9.3 e-4	4.0 e-2	2.7 e-1	8.5 e-2	7.2 e-4	8.1 e-2
BDL-2-C / BDL-3-C	Manual	7.3 e-2	8.6 e-4	6.1 e-3	3.6 e-3	2.0 e-3	3.2 e-3	9.6 e-3	4.3 e-3	8.4 e-2	7.2 e-2
	Structure	3.1 e-4	2.0 e-2	2.1 e-1	4.5 e-3	9.7 e-1	6.1 e-1	2.4 e-1	4.0 e-1	4.8 e-3	5.5 e-1
	Random lines	1.5 e-4	3.5 e-4	4.8 e-3	2.7 e-4	1.2 e-3	3.0 e-2	6.3 e-1	3.1 e-1	1.2 e-1	1.2 e-1
Osmol-1-C/C	Structure	7.0 e-1	7.0 e-2	2.4 e-2	2.7 e-1	3.3 e-2	2.0 e-2	1.3 e-2	9.0 e-3	6.5 e-1	5.0 e-2
	Random lines	8.0 e-1	2.7 e-1	5.7 e-1	5.4 e-1	4.1 e-1	1.2 e-1	5.0 e-1	2.1 e-1	3.2 e-3	2.2 e-1
Osmol-2-C/C	Structure	7.6 e-1	4.2 e-1	6.8 e-1	6.9 e-1	6.0 e-1	3.0 e-1	6.6 e-1	5.8 e-1	7.1 e-1	3.0 e-1
	Random lines	1.9 e-1	3.8 e-1	3.1 e-1	2.8 e-1	2.9 e-1	6.4 e-1	7.0 e-1	6.8 e-1	3.0 e-2	6.9 e-1
Osmol-3-C/C	Structure	1.0 e-1	6.6 e-1	5.1 e-1	3.1 e-1	5.0 e-1	4.1 e-1	6.8 e-1	6.6 e-1	3.2 e-2	6.8 e-1
	Random lines	7.2 e-1	6.6 e-1	6.5 e-1	7.3 e-1	6.3 e-1	4.9 e-1	5.5 e-1	5.2 e-1	2.0 e-2	1.2 e-1

Table B.0.1.: Significance of translocation detected in control-control datasets for the 1D-Can process by the automated workflows (structure-based and random lines-based, 20 top-ranked profiles) and by the manual workflow (10 profiles). Statistically significant translocation effects (p-value ≤ 0.05) are highlighted red.

Dataset	Workflow	Descriptor									
		X	Y	Z	A	B	C	D	E	F	Variance
Osmol-1-C/T	Manual	1.3 e-3	1.6 e-21	3.3 e-24	2.0 e-10	5.7 e-24	3.7 e-26	4.3 e-27	1.1 e-27	6.2 e-14	3.4 e-17
	Structure	6.8 e-5	3.9 e-20	7.3 e-23	6.9 e-11	2.2 e-22	2.5 e-25	9.0 e-27	1.5 e-27	1.7 e-17	1.3 e-17
	Random lines	1.6 e-2	2.4 e-6	1.3 e-9	5.7 e-4	4.0 e-8	4.5 e-9	6.0 e-14	1.2 e-12	6.1 e-6	1.0 e-17
Osmol-2-C/T	Structure	8.1 e-17	5.1 e-17	1.9 e-16	1.2 e-15	8.9 e-18	2.4 e-11	1.5 e-4	1.0 e-7	1.7 e-23	8.7 e-11
	Random lines	1.1 e-9	4.2 e-8	3.5 e-9	4.5 e-8	1.2 e-8	1.6 e-5	4.3 e-5	8.6 e-6	2.0 e-9	2.4 e-10
BDL-1-C/T	Manual	1.9 e-1	4.8 e-4	8.9 e-5	2.9 e-1	8.5 e-5	1.3 e-6	9.3 e-8	4.0 e-8	6.5 e-1	6.8 e-10
	Structure	6.2 e-2	4.7 e-3	3.5 e-4	5.4 e-1	9.3 e-4	5.1 e-10	1.2 e-12	9.0 e-13	4.7 e-4	3.1 e-9
	Random lines	1.3 e-1	2.9 e-1	5.0 e-2	6.3 e-1	1.2 e-1	6.0 e-5	4.0 e-6	6.6 e-1	4.0 e-6	4.0 e-6
BDL-2-C/T	Manual	3.9 e-1	4.1 e-9	4.1 e-4	5.7 e-3	1.7 e-6	1.8 e-15	9.9 e-7	1.5 e-10	1.9 e-11	3.8 e-9
	Structure	7.6 e-4	4.6 e-7	3.9 e-5	1.4 e-5	3.4 e-6	2.6 e-7	2.2 e-3	1.8 e-5	7.9 e-10	2.9 e-3
	Random lines	9.5 e-4	2.5 e-6	7.0 e-7	7.0 e-5	4.0 e-7	1.0 e-6	3.0 e-6	2.7 e-7	3.5 e-6	1.4 e-7
BDL-3-C/T	Manual	2.0 e-1	1.9 e-17	6.3 e-14	8.4 e-9	2.5 e-16	3.9 e-21	8.1 e-15	1.3 e-18	5.5 e-1	8.9 e-15
	Structure	5.2 e-6	7.3 e-16	1.1 e-12	2.4 e-11	6.1 e-15	1.3 e-13	1.0 e-8	1.1 e-11	1.6 e-4	5.0 e-7
	Random lines	2.3 e-7	4.2 e-11	6.0 e-10	8.8 e-9	6.4 e-11	1.8 e-11	1.0 e-7	3.8 e-10	9.0 e-2	4.9 e-10
TCDC-1-C/T	Structure	1.8 e-2	6.0 e-1	2.7 e-1	3.3 e-1	3.8 e-1	4.3 e-2	5.8 e-3	6.4 e-3	2.2 e-2	5.8 e-1
	Random lines	7.2 e-1	3.4 e-1	7.9 e-2	5.7 e-1	1.5 e-1	4.4 e-2	2.5 e-3	4.5 e-3	2.7 e-1	5.5 e-1
TCDC-2-C/T	Structure	7.6 e-3	2.2 e-5	4.4 e-4	1.4 e-3	5.0 e-5	1.1 e-4	3.8 e-2	1.7 e-3	7.7 e-11	3.9 e-3
	Random lines	5.9 e-1	8.7 e-2	1.9 e-1	3.1 e-1	1.3 e-1	1.8 e-2	5.7 e-2	1.9 e-2	1.2 e-1	3.8 e-3
TCDC-3-C/T	Structure	1.9 e-1	4.4 e-3	1.6 e-4	4.7 e-1	5.4 e-4	4.9 e-6	1.2 e-7	1.9 e-7	2.2 e-6	4.3 e-7
	Random lines	3.0 e-1	4.0 e-3	1.0 e-3	6.7 e-2	1.3 e-3	3.6 e-4	8.0 e-4	2.2 e-4	7.9 e-4	6.0 e-5
TCDC-4-C/T	Structure	5.0 e-1	1.5 e-7	7.4 e-12	1.5 e-2	3.5 e-10	3.0 e-11	1.8 e-14	1.1 e-13	1.9 e-8	2.3 e-11
	Random lines	3.2 e-2	7.4 e-2	3.5 e-3	4.4 e-2	1.3 e-2	1.2 e-1	3.2 e-3	8.2 e-3	1.6 e-4	5.3 e-1

Table B.0.2.: Significance of translocation detected in control-test datasets for the 1D-Can process by the automated workflows (structure-based and random lines-based, 20 top-ranked profiles) and by the manual workflow (10 profiles). Translocation effects that were not found statistically significant (p-value > 0.05) are highlighted blue.

Dataset	Canalicular width	Descriptor									
		X	Y	Z	A	B	C	D	E	F	Variance
Control-test datasets											
Osmol-1-C/T	All widths	6.8 e-5	3.9 e-20	7.3 e-23	6.9 e-11	2.2 e-22	2.5 e-25	9.0 e-27	1.5 e-27	1.7 e-17	1.3 e-17
	1.2 – 2.2 μm	3.7 e-4	3.4 e-15	4.7 e-21	1.1 e-7	4.8 e-19	1.3 e-20	4.4 e-24	6.5 e-25	2.9 e-17	2.7 e-16
	1.4 – 2.0 μm	1.8 e-3	4.1 e-15	2.1 e-19	4.9 e-7	5.6 e-18	2.4 e-24	1.0 e-23	6.5 e-26	2.5 e-20	1.9 e-17
BDL-3-C/T	All widths	5.2 e-6	7.3 e-16	1.1 e-12	2.4 e-11	6.1 e-15	1.3 e-13	1.0 e-8	1.1 e-11	1.6 e-4	5.0 e-7
	1.2 – 2.2 μm	1.0 e-7	1.7 e-13	3.2 e-13	1.3 e-10	1.7 e-14	9.2 e-12	5.5 e-8	7.6 e-11	3.0 e-9	1.2 e-8
	1.4 – 2.0 μm	1.3 e-6	8.0 e-13	1.1 e-14	1.4 e-9	3.6 e-15	1.7 e-12	1.7 e-10	2.4 e-13	7.2 e-12	2.4 e-10
TCDC-3-C/T	All widths	1.9 e-1	4.4 e-3	1.6 e-4	4.7 e-1	5.4 e-4	4.9 e-6	1.2 e-7	1.9 e-7	2.2 e-6	4.3 e-7
	1.2 – 2.2 μm	6.8 e-1	2.6 e-3	1.6 e-5	2.5 e-1	1.3 e-4	3.8 e-6	2.8 e-7	1.1 e-7	2.8 e-7	1.2 e-7
	1.4 – 2.0 μm	3.7 e-1	2.3 e-5	2.7 e-7	2.0 e-2	5.6 e-7	1.5 e-7	2.6 e-8	3.7 e-9	1.2 e-9	2.0 e-8
TCDC-4-C/T	All widths	5.0 e-1	1.5 e-7	7.4 e-12	1.5 e-2	3.5 e-10	3.0 e-11	1.8 e-14	1.1 e-13	1.9 e-8	2.3 e-11
	1.2 – 2.2 μm	5.1 e-1	5.9 e-8	2.1 e-14	3.8 e-2	8.4 e-12	1.3 e-11	2.2 e-14	1.8 e-14	2.3 e-16	4.6 e-12
	1.4 – 2.0 μm	2.7 e-1	3.1 e-10	5.0 e-15	6.0 e-3	3.8 e-13	1.1 e-12	1.4 e-13	5.7 e-14	2.2 e-15	6.8 e-11
Control-control datasets											
TC-1-C/C	All widths	3.0 e-2	2.3 e-1	4.4 e-1	8.0 e-2	4.0 e-1	6.0 e-1	4.0 e-1	4.0 e-1	2.0 e-2	2.0 e-2
	1.2 – 2.2 μm	1.8 e-2	4.0 e-1	4.7 e-1	9.7 e-2	4.5 e-1	3.2 e-1	2.3 e-1	1.8 e-1	4.4 e-3	2.5 e-3
	1.4 – 2.0 μm	9.0 e-2	3.5 e-1	3.8 e-1	2.1 e-1	3.6 e-1	6.4 e-1	6.5 e-1	6.0 e-1	2.0 e-1	1.4 e-1
TC-2-C/C	All widths	5.7 e-1	4.2 e-1	5.8 e-1	6.5 e-1	5.7 e-1	3.4 e-1	6.4 e-1	5.2 e-1	4.0 e-2	1.2 e-1
	1.2 – 2.2 μm	6.2 e-1	2.5 e-1	6.0 e-1	6.5 e-1	3.9 e-1	1.8 e-1	6.5 e-1	4.4 e-1	2.1 e-4	4.2 e-1
	1.4 – 2.0 μm	6.8 e-1	4.4 e-2	7.5 e-2	3.5 e-1	1.6 e-2	3.7 e-2	2.6 e-1	6.6 e-2	1.0 e-2	5.3 e-1
Osmol-2-C/C	All widths	7.6 e-1	4.2 e-1	6.8 e-1	6.9 e-1	6.0 e-1	3.0 e-1	6.6 e-1	5.8 e-1	7.1 e-1	3.0 e-1
	1.2 – 2.2 μm	2.7 e-1	6.2 e-1	6.0 e-1	4.7 e-1	6.3 e-1	3.7 e-1	1.9 e-1	2.5 e-1	5.1 e-2	7.3 e-1
	1.4 – 2.0 μm	2.6 e-1	6.8 e-1	5.6 e-1	4.0 e-1	6.9 e-1	4.6 e-1	1.6 e-1	1.9 e-1	3.8 e-1	1.2 e-1

Table B.0.3.: Influence of canalicular width restriction on significance of the detected translocation (1D-Can process evaluated by the structure-based workflow). Statistically significant translocations in control-control datasets and not statistically significant translocations in control-test datasets are highlighted red and blue, respectively.

Dataset	Workflow	Descriptor					
		L	M	N	O	P	Variance
TC-1-C/C	Manual	2.3 e-2	3.5 e-2	1.8 e-4	5.5 e-4	5.4 e-4	1.0 e-3
	Structure	5.5 e-1	5.8 e-1	5.3 e-1	5.2 e-1	5.9 e-1	1.6 e-1
	Random lines	3.4 e-1	1.2 e-1	5.3 e-1	3.2 e-1	6.2 e-1	7.1 e-1
TC-2-C/C	Manual	9.6 e-2	9.7 e-1	1.5 e-1	4.4 e-1	4.9 e-2	7.3 e-2
	Structure	5.8 e-2	2.3 e-2	1.7 e-3	1.7 e-3	1.3 e-2	2.0 e-2
	Random lines	2.1 e-2	1.6 e-4	1.7 e-5	8.6 e-4	1.2 e-4	1.7 e-2
TC-3-C/C	Manual	5.9 e-1	8.7 e-1	6.3 e-1	7.0 e-1	6.3 e-1	4.3 e-4
	Structure	4.5 e-1	1.8 e-1	3.0 e-1	5.7 e-2	7.1 e-1	4.9 e-1
	Random lines	5.9 e-1	7.9 e-2	2.7 e-1	1.8 e-1	4.4 e-1	3.1 e-1
TC-4-C/C	Manual	3.0 e-2	3.0 e-3	5.5 e-3	1.6 e-3	9.7 e-2	1.9 e-1
	Structure	5.2 e-1	3.7 e-1	6.3 e-1	6.7 e-1	4.8 e-1	1.5 e-1
	Random lines	2.9 e-1	1.7 e-1	5.8 e-2	5.8 e-2	9.3 e-1	5.1 e-1
TC-5-C/C	Manual	1.1 e-2	7.9 e-1	7.2 e-3	1.0 e-1	1.5 e-3	3.9 e-2
	Structure	6.9 e-1	4.3 e-1	3.3 e-1	3.3 e-1	2.0 e-1	4.7 e-1
	Random lines	8.2 e-4	1.5 e-2	1.5 e-5	1.2 e-4	3.0 e-5	7.6 e-2
TC-6-C/C	Manual	9.9 e-1	7.8 e-1	3.8 e-1	8.0 e-1	1.1 e-1	2.9 e-1
	Structure	2.6 e-1	1.5 e-1	3.6 e-1	5.4 e-2	6.3 e-2	3.5 e-1
	Random lines	7.0 e-1	6.3 e-1	6.1 e-1	6.0 e-1	6.2 e-1	4.5 e-1
TC-7-C/C	Structure	2.6 e-1	1.5 e-1	8.6 e-2	5.4 e-2	6.2 e-2	3.5 e-1
	Random lines	2.8 e-1	5.3 e-1	5.7 e-1	6.4 e-1	5.1 e-1	4.3 e-1

Table B.0.4.: Significance of translocation detected in control-control datasets for the 1D-Bas process by the automated workflows (structure-based and random lines-based, 20 top-ranked profiles) and by the manual workflow (10 profiles). Statistically significant translocation effects (p-value ≤ 0.05) are highlighted red.

Dataset	Workflow	Descriptor					
		L	M	N	O	P	Variance
TCDC-1-C/T	Manual	1.3 e-2	2.0 e-6	6.0 e-6	4.2 e-7	3.9 e-3	7.5 e-10
	Structure	1.3 e-5	3.6 e-10	4.5 e-12	5.8 e-13	7.0 e-8	4.5 e-8
	Random lines	1.5 e-3	5.9 e-4	1.4 e-7	1.8 e-6	3.7 e-7	5.8 e-9
TCDC-2-C/T	Manual	1.1 e-2	1.2 e-6	6.5 e-8	5.0 e-8	9.0 e-6	5.7 e-7
	Structure	2.4 e-4	2.0 e-7	9.9 e-12	5.3 e-11	6.8 e-10	4.4 e-4
	Random lines	1.6 e-1	7.7 e-7	7.4 e-5	2.0 e-6	3.9 e-3	2.6 e-6
TCDC-3-C/T	Manual	3.8 e-2	8.0 e-12	1.5 e-10	6.0 e-12	1.2 e-5	1.6 e-4
	Structure	6.7 e-1	4.6 e-2	3.2 e-2	1.8 e-2	1.3 e-1	9.6 e-3
	Random lines	6.4 e-1	3.8 e-3	1.6 e-2	2.1 e-3	1.4 e-1	1.8 e-1
TCDC-4-C/T	Manual	1.2 e-3	3.8 e-1	7.8 e-2	5.7 e-2	7.5 e-4	1.1 e-3
	Structure	5.2 e-1	2.7 e-2	3.1 e-2	1.2 e-2	1.2 e-1	1.6 e-5
	Random lines	5.7 e-1	1.0 e-1	1.8 e-1	1.3 e-1	3.3 e-1	8.1 e-4
TCDC-5-C/T	Manual	9.0 e-1	5.0 e-2	9.0 e-2	5.0 e-2	1.4 e-1	3.0 e-3
	Structure	7.2 e-1	6.3 e-1	3.4 e-1	4.2 e-1	3.0 e-1	5.2 e-1
	Random lines	3.5 e-1	5.8 e-1	5.0 e-1	5.3 e-1	5.1 e-1	6.5 e-1
TCDC-6-C/T	Structure	3.0 e-4	2.1 e-5	8.9 e-7	1.6 e-7	1.6 e-7	1.9 e-4
	Random lines	3.7 e-1	5.3 e-3	4.4 e-3	1.9 e-3	2.6 e-2	4.4 e-5

Table B.0.5.: Significance of translocation detected in control-test datasets for the 1D-Bas process by the automated workflows (structure-based and random lines-based, 20 top-ranked profiles) and by the manual workflow (10 profiles). Translocation effects that were not found statistically significant (p-value > 0.05) are highlighted blue.

Dataset	Workflow	Glucose g/L											
		0	0	1	1	2	2	3	3	4	4	4.5	4.5
Gluc-1	Structure	2.60	3.41	2.46	2.25	2.43	2.38	2.10	2.24	1.75	1.69	1.67	1.44
	Random lines	3.01	3.59	2.63	2.39	2.37	2.44	2.14	2.22	1.69	1.72	1.73	1.46
GlucDMSO-1	Structure	2.72	3.23	1.97	2.34	1.37	1.21	1.30	1.16	1.19	1.13	1.10	1.14
	Random lines	2.71	3.04	1.86	2.41	1.42	1.21	1.30	1.18	1.20	1.16	1.10	1.14
GlucIns-1	Structure	3.12	2.65	1.79	1.88	2.18	2.13	1.68	2.54	1.63	1.44	1.72	1.87
	Random lines	2.95	2.64	1.84	1.83	2.13	2.11	1.78	2.49	1.61	1.49	1.89	2.00
GlucIns-2	Structure	2.24	2.78	2.29	1.80	1.50	1.40	1.20	1.44	1.02	1.04	1.10	1.20
	Random lines	2.19	2.48	2.14	1.71	1.48	1.54	1.24	1.46	0.99	1.03	1.13	1.22
GlucR-1	Structure	2.26	2.76	1.73	1.90	1.26	1.22	1.02	0.89	0.99	1.11	0.94	0.93
	Random lines	2.44	2.91	1.69	1.81	1.27	1.23	1.05	0.88	0.97	1.14	0.95	0.93
GlucR-2	Structure	4.01	3.47	2.05	2.82	2.61	2.43	2.93	3.16	3.08	3.21	3.08	1.96
	Random lines	4.28	3.69	2.04	2.86	2.88	2.75	3.09	3.13	3.07	3.04	3.04	2.01
GlucR-3	Structure	2.26	2.76	1.73	1.90	1.26	1.22	1.02	0.89	0.99	1.11	0.94	0.93
	Random lines	2.44	2.91	1.69	1.81	1.27	1.23	1.05	0.88	0.97	1.14	0.95	0.93
GlucR-4	Structure	2.06	2.26	1.93	1.91	1.45	1.60	1.24	1.19	1.13	1.14	1.02	1.01
	Random lines	2.13	2.38	1.97	1.86	1.47	1.71	1.29	1.23	1.11	1.14	1.02	1.00
GlucR-5	Structure	2.15	2.66	2.04	1.79	1.75	1.73	1.35	1.49	1.15	1.19	1.23	1.20
	Random lines	2.26	2.58	2.14	1.73	1.77	1.81	1.35	1.53	1.15	1.19	1.20	1.18
Dataset	Workflow	2 g/L glucose, *Reagent6* mol/L											
		0	0	1	1	2	2	5	5	10	10	100	100
GlucR-6	Structure	2.23	2.09	2.31	2.22	2.75	2.71	2.10	1.81	2.12	2.01	1.69	1.84
	Random lines	2.16	2.11	2.35	2.16	2.65	2.69	2.14	1.66	2.10	2.15	1.58	1.82

Table B.0.6.: Nucleus-to-cytoplasm ratios obtained by the random lines-based and the established structure-based workflow for the 2D-Nuc process.

Glossary

1D translocation is a directed translocation of a functional marker to or across the membrane and orthogonally to it.

1D-Bas process is a translocation of Ntcp between basolateral membranes and cytoplasm of hepatocytes.

1D-Can process is a translocation of Bsep between canalicular membranes and cytoplasm of hepatocytes.

2D translocation such translocation processes where biomolecules are included into or excluded from a closed region.

2D-Nuc process is a translocation of glucokinase between nuclei and cytoplasm of hepatocytes.

con-centric profile is a profile centered at the center of an already accepted random profile.

con-centric profile extraction is an extraction of profiles centered at the central positions of the already accepted random profiles. It can be applied to account for roundish regions in e.g. 2D translocation cases.

densitometry is the quantitative analysis of molecule density in images.

descriptor is a numeric parameter calculated from the molecule distribution function. Can be used to quantify molecule distribution in relation to the reference structures.

distribution profiling yields a set of profiles that clearly depict the biological structure studied. It consists of profile extraction and selection.

functional marker is a biomolecule whose translocation has to be quantified.

manual workflow is a translocation quantification workflow based on the manual extraction of molecule distribution profiles.

multi-marker image is a biological image containing at least one labeled functional and one structural marker (e.g. fluorescence microscopy images).

optimal 1D transport model describes the directed translocation of functional marker along the optimal trajectory (concentration gradient) across the membrane and orthogonally to it.

optimal transport trajectory is the most efficient trajectory for translocation of biomolecules which runs along the concentration gradients. The concentration gradients, in turn, frequently run across a membrane orthogonally to it.

profile or density profile, or intensity profile, or distribution profile, is a 1D numeric vector representing concentration (or density, or intensity) of biomolecules extracted along a line.

profile extraction is the extraction of the information on molecule distribution along the lines resulting in profiles (e.g. by line densitometry in biological images).

random lines-based workflow is an automated translocation quantification workflow avoiding structure detection in images. It is based on the profile extraction along numerous random lines (*random profile extraction*).

random profile is a profile extracted along a random line drawn in multi-marker images.

random profile extraction is a profile extraction along random lines in multi-marker images avoiding structure detection by image processing operations.

reference structure is a biological structure that is stable and visible under various experimental conditions. It is helpful to quantify translocation of biomolecules based on images of different biological samples.

structural marker is a biomolecule whose localization in the biological sample is known and is visible under experimental conditions.

structure detection is an image processing operation that localizes objects in images.

structure-based profile is a profile extracted in relation to a biological structure detected in images (e.g. along concentration gradients which run orthogonally to membranes).

structure-based profile extraction is the profile extraction in relation to biological structures detected in images.

structure-based workflow is an automated workflow to quantify translocation of biomolecules, including structure detection in images, distribution profiling and toponomic characterization.

toponomic characterization is a representation of the unknown molecule distribution function by a set of numeric descriptors.

toponomics is the field of research describing the laws of spatial arrangement of molecules.

translocation is a transport process where biomolecules are moving from one compartment to another.

transport is movement of biomolecules from one location to another.

zone model is a set of biologically relevant zones (regions) defined in profiles.

Acronyms and notations

Bsep the bile salt export pump, a bile transporter protein at canalicular membranes of hepatocytes.

control dataset images acquired from tissue sections of the liver incubated under normo-osmolar conditions or perfused with reagents that should not influence the functional marker distribution.

GK glucokinase, the enzyme that catalyses the first step in glucose metabolism.

GKRP glucokinase regulatory protein required to transfer glucokinase into hepatocyte nuclei.

MELC Multi-epitope-ligand cartography, is a prominent technology that performs colocalization analysis of up to 100 biomolecules in a sample.

Mrp2 the conjugate export pump multidrug resistance protein 2, a bile transporter protein at canalicular membranes of hepatocytes.

Na^+-K^+-ATPase a standard structural marker used for the analysis of basolateral membranes of hepatocytes.

Ntcp the sodium taurocholate cotransporting polypeptide, is the major transporter for the bile salt uptake at the basolateral membrane of hepatocytes.

SE structuring element, is a set of a known shape with which the image is probed.

TC taurocholate, a bile salt which does not influence translocation of bile transporter proteins.

TCDC taurochenodeoxycholate, a bile salt which influences translocation of bile transporter proteins.

test dataset images acquired from tissue sections of the liver either incubated under aniso-osmolar conditions, or after the bile duct ligature, or perfused with reagents that are expected to influence distribution of the functional marker.

WRST Wilcoxon Rank Sum Test, a widely used nonparametric test which compares observations from independent sets.

z-stack is a series of images acquired at the same X, Y coordinates while varied Z position.

Zo-1 the Zonula occludens 1 protein, a protein of the tight junction complex at canalicular membranes of hepatocytes.

List of Figures

List of Tables

Bibliography

[1] M Abramoff, P Magelhaes, and S Ram. Image processing with ImageJ. *Biophotonics International*, 11:36–42, 2004.

[2] L Agius. Glucokinase and molecular aspects of liver glycogen metabolism. *Biochemical Journal*, 414(1):1–18, 2008.

[3] Akaike & Tagawa Lab. Liver. http://www.akaike-lab.bio.titech.ac.jp/akaike/english/resarch/index.html. Last visited on 2013-05-20.

[4] B Alberts, A Johnson, J Lewis, and et al. *Molecular Biology of the Cell*. Garland Science, New York, NY, USA, 4th edition, 2002.

[5] W Alrefai and R Gill. Bile acid transporters: structure, function, regulation and pathophysiological implications. *Pharmaceutical Research*, 24:1803–1823, 2007.

[6] Amira. 3D Analysis Software for Life Sciences. http://www.amira.com/. Last visited on 2013-05-20.

[7] M Ashwin and B Dinesh. A new sequential thinning algorithm to preserve topology and geometry of the image. *International Journal of Mathematics Trends and Technology*, 2 (2), 2011.

[8] X Bai, L Latecki, and W Liu. Skeleton Pruning by Contour Partitioning with Discrete Curve Evolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29 (3):449–462, 2007.

[9] S Becker, R Reinehr, D Graf, S vom Dahl, and D Häussinger. Hydrophobic bile salts induce hepatocyte shrinkage via NADPH oxidase activation. *Cellular Physiology and Biochemistry*, 19(1-4):89–98, 2007.

[10] BIO Imaging. Fluorescence microscopy. http://www.scienceinyoureyes.com/index.php?id=79. Last visited on 2013-05-20.

[11] Bitplane. Scientific software. http://www.bitplane.com/. Last visited on 2013-05-20.

[12] M Bode, M Irmler, M Friedenberger, C May, K Jung, C Stephan, H Meyer, C Lach, R Hillert, A Krusche, J Beckers, K Marcus, and W Schubert. Interlocking transcriptomics, proteomics and toponomics technologies for brain tissue analysis in murine hippocampus. *Proteomics*, 8(6):1170–1178, 2008.

[13] S Bolte and F Cordelieres. A guided tour into subcellular colocalization analysis in light microscopy. *Journal of Microscopy*, 224(Pt 3):213–232, 2006.

[14] B Bonnekoh, H Gollnick, L Philipsen, A Pommer, and R Böckelmann. Skin topoproteome analysis by multi-epitope ligand cartography for identification of pathogenetic and therapeutic targets. *Drug Discovery*, pages 32–34, 2007.

[15] J Bradley, A Nofal, I El Naqa, W Lu, J Liu, J Hubenschmidt, D Low, R Drzymala, and D Khullar. Comparison of helical, maximum intensity projection (MIP), and averaged intensity (AI) 4D CT imaging for stereotactic body radiation therapy (SBRT) planning in lung cancer. *Radiotherapy & Oncology*, 81(3):264–268, 2006.

[16] E Bromage and S Kaattari. Simultaneous quantitative analysis of multiple protein species within a single sample using standard scanning densitometry. *Journal of Immunological Methods*, 323(2):109–113, 2007.

[17] J Brosnan. Glutamate, at the interface between amino acid and carbohydrate metabolism. *The Journal of Nutrition*, 130(4S Suppl):988S–990S, 2000.

[18] L Bull and et al. Genetic and morphological findings in progressive familial intrahepatic cholestasis (Byler disease [PFIC-1] and Byler syndrome): evidence for heterogeneity. *Hepatology*, 26(1):155–164, 1997.

[19] M Calandrella, D Matteucci, P Mazzetti, and A Poli. Densitometric analysis of western blot assays for feline immunodeficiency virus antibodies. *Veterinary Immunology and Immunopathology*, 79(3-4):261–271, 2001.

[20] M Cantore, R Reinehr, A Sommerfeld, M Becker, and D Häussinger. The Src family kinase Fyn mediates hyperosmolarity-induced Mrp2 and Bsep retrieval from canalicular membrane. *The Journal of Biological Chemistry*, 286(52):45014–45029, 2011.

[21] L Cardelli. Abstract machines of systems biology. In C Priami, E Merelli, P Gonzalez, and A Omicini, editors, *Transactions on Computational Systems Biology*, volume 3737 of *Lecture Notes in Computer Science*, pages 145–168. Springer, New York, NY, USA, 2005.

[22] A Carpenter. Extracting rich information from images. *Methods in Molecular Biology*, 486:193–211, 2009.

[23] A Carpenter, T Jones, M Lamprecht, C Clarke, I Kang, O Friman, D Guertin, J Chang, R Lindquist, J Moffat, P Golland, and D Sabatini. Cellprofiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology*, 7(10):R100, 2006.

[24] K Chun. Bone densitometry. *Seminars in Nuclear Medicine*, 41(3):220–228, 2011.

[25] B Cohen. Biological imaging: Beyond fluorescence. *Nature*, 467(7314):407–408, 2010.

[26] M Collins. Generating 'omic knowledge': the role of informatics in high content screening. *Comb Chem High Throughput Screen.*, 12(9):917–925, 2009.

[27] J A Conchello and J W Lichtman. Optical sectioning microscopy. *Nature Methods*, 2: 920–931, 2005.

[28] S Costes, D Daelemans, E Cho, Z Dobbin, G Pavlakis, and S Lockett. Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophysical Journal*, 86(6):3993–4003, 2004.

[29] N Crabtree and K Ward. Bone densitometry: current status and future perspectives. *Endocrine Development*, 16:58–72, 2009.

[30] C Daly and J McGrath. Fluorescent ligands, antibodies, and proteins for the study of receptors. *Pharmacology & Therapeutics*, 100(2):101–118, 2003.

[31] M Davidson and M Abramowitz. Optical Microscopy. *Encyclopedia of Imaging Science and Technology*, pages 1–41, 2002.

[32] A Davit-Spraul, E Gonzales, C Baussan, and E Jacquemin. Progressive familial intrahepatic cholestasis. *Orphanet Journal of Rare Diseases*, 8(4), 2009.

[33] N de la Iglesia, M Veiga-da Cunha, E Van Schaftingen, J Guinovart, and J Ferrer. Glucokinase regulatory protein is essential for the proper subcellular localisation of liver glucokinase. *FEBS Letters*, 456(2):332–338, 1999.

[34] A Deptala, E Bedner, W Gorczyca, and Z Darzynkiewicz. Activation of nuclear factor kappa B (NF-kappaB) assayed by laser scanning cytometry (LSC). *Cytometry*, 33(3): 376–382, 1998.

[35] O Domanova, S Borbe, S Mühlfeld, M Becker, R Kubitz, D Häussinger, and T Berlage. Toponomics method for the automated quantification of membrane protein translocation. *BMC Bioinformatics*, 12:370, 2011.

[36] F Dombrowski, R Kubitz, A Chittattu, M Wettstein, N Saha, and D Häussinger. Electron-microscopic demonstration of multidrug resistance protein 2 (Mrp2) retrieval from the canalicular membrane in response to hyperosmolarity and lipopolysaccharide. *Biochemical Journal*, 348(Pt 1):183–188, 2000.

[37] G Donnan. Translational research: a critical point in time. *International Journal of Stroke*, 7(5):367, 2012.

[38] M Donner and D Keppler. Up-regulation of basolateral multidrug resistance protein 3 (Mrp3) in cholestatic rat liver. *Hepatology*, 34:351–359, 2001.

[39] E Dougherty and R Lotufo. *Hands-on morphological image processing*. Spie Press Series, Bellingham, 2003.

[40] J Dranoff, M McClure, A Burgstahler, L Denson, A Crawford, J Crawford, S Karpen, and M Nathanson. Short-term regulation of bile acid uptake by microfilament-dependent translocation of rat ntcp to the plasma membrane. *Hepatology*, 30(1):223–229, 1999.

[41] E Drelie Gelasca, B Obara, D Fedorov, K Kvilekval, and B Manjunath. A biosegmentation benchmark for evaluation of bioimage analysis methods. *BMC Bioinformatics*, 10:368, 2009.

[42] K Dunn, M Kamocka, and J McDonald. A practical guide to evaluating colocalization in biological microscopy. *American Journal of Physiology Cell Physiology*, 300(4):C723–C742, 2011.

[43] T Duong, B Goud, and K Schauer. Closed-form density-based framework for automatic detection of cellular morphology changes. *Proceedings of the National academy of Sciences of the United States of America*, 109(22):8382–8387, 2012.

[44] R Eils and C Athale. Computational imaging in cell biology. *Journal of Cell Biology*, 161 (3):477–481, 2003.

[45] R Elferink and A Groen. Genetic defects in hepatobiliary transport. *Biochimica et Biophysica Acta*, 1586(2):129–145, 2002.

[46] K Eliceiri, M Berthold, I Goldberg, L Ibez, B Manjunath, M Martone, R Murphy, H Peng, A Plant, B Roysam, N Stuurmann, J Swedlow, P Tomancak, and A Carpenter. Biological imaging software tools. *Nature Methods*, 9(7):697–710, 2012.

[47] A Esteller. Physiology of bile secretion. *World Journal of Gastroenterology*, 14(37):5641–5649, 2008.

[48] F Fang and A Casadevall. Lost in translation–basic science in the era of translational research. *Infection and Immunity*, 78(2):563–566, 2010.

[49] R Ferguson, H Carroll, A Harris, E Maher, P Selby, and R Banks. Housekeeping proteins: a preliminary study illustrating some limitations as useful references in protein expression studies. *Proteomics*, 5(2):566–571, 2005.

[50] R Fischer, Y Wu, P Kanchanawong, H Shroff, and C Waterman. Microscopy in 3D: a biologist's toolbox. *Trends in Cell Biology*, 21(12):682–691, 2011.

[51] J Frank, T Wagenknecht, B McEwen, M Marko, C Hsieh, and C Mannella. Three-dimensional imaging of biological complexity. *Journal of Structural Biology*, 138(1-2): 85–91, 2002.

[52] M Friedman. *Principles and Models of Biological Transport*. Springer, 2nd edition, 2008.

[53] Y Fu, C Lin, G Enikolopov, E Sibley, A Chiang, and S Tang. Microtome-free 3-dimensional confocal imaging method for visualization of mouse intestine with subcellular-level resolution. *Gastroenterology*, 137:453–465, 2009.

[54] C Galbraith and J Galbraith. Super-resolution microscopy at a glance. *Journal of Cell Science*, 124:1607–1611, 2011.

[55] M Gassmann, B Grenacher, B Rohde, and J Vogel. Quantifying western blots: pitfalls of densitometry. *Electrophoresis*, 30(11):1845–1855, 2009.

[56] V Geenes and C Williamson. Intrahepatic cholestasis of pregnancy. *World Journal of Gastroenterology*, 15(17):2049–2066, 2009.

[57] A Geier, P Fickert, and M Trauner. Mechanisms of disease: mechanisms and clinical implications of cholestasis in sepsis. *Nature Clinical Practice Gastroenterology & Hepatology*, 3(10):574–585, 2006.

[58] D Gerlich, J Beaudouin, M Gebhard, J Ellenberg, and R Eils. Four-dimensional imaging and quantitative reconstruction to analyse complex spatiotemporal processes in live cells. *Nature Cell Biology*, 3(9):852–855, 2001.

[59] T Gerloff, B Stieger, B Hagenbuch, J Madon, L Landmann, J Roth, A F Hofmann, and P J Meier. The sister of P-glycoprotein represents the canalicular bile salt export pump of mammalian liver. *The Journal of Biological Chemistry*, 273(16):10046–10050, 1998.

[60] I Ghiran. Introduction to fluorescence microscopy. *Methods in Molecular Biology*, 689: 93–136, 2011.

[61] E Glory and R Murphy. Automated subcellular location determination and high-throughput microscopy. *Developmental Cell*, 12(1):7–16, 2007.

[62] Glucokinase on ExPASy, the SIB Bioinformatics Resource Portal. http://enzyme.expasy.org/EC/2.7.1.2. Last visited on 2013-05-20.

[63] Glucokinase on UniProtKB/Swiss-Prot. http://www.uniprot.org/uniprot/P35557. Last visited on 2013-05-20.

[64] R Gonzalez and R Woods. *Digital Image Processing*. Prentice-Hall, Upper Saddle River, NJ, USA, 3rd edition, 2006.

[65] M Gottesman and I Pastan. Biochemistry of multidrug resistance mediated by the multidrug transporter. *Annual Review of Biochemistry*, 62:385–427, 1993.

[66] N Grabe. Reconstructing protein networks of epithelial differentiation from histological sections. *Bioinformatics*, 23(23):3200–3208, 2007.

[67] R Haralick, K Shanmugam, and I Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6):610–621, 1973.

[68] A Hattersley and E Pearson. Minireview: pharmacogenetics and beyond: the interaction of therapeutic response, beta-cell physiology, and genetics in diabetes. *Endocrinology*, 147 (6):2657–2663, 2006.

[69] D Häussinger, M Schmitt, O Weiergräber, and R Kubitz. Short-term regulation of canalicular transport. *Seminars in Liver Diseases*, 20:307–321, 2000.

[70] D Häussinger, R Kubitz, R Reinehr, JG Bode, and F Schliess. Molecular aspects of medicine: from experimental to clinical hepatology. *Molecular Aspects of Medicine*, 25 (3):221–360, 2004.

[71] C Hedvat. Digital microscopy: past, present, and future. *Archives of Pathololgy & Laboratory Medicine*, 134(11):1666–1670, 2010.

[72] J Helmuth, G Paul, and I Sbalzarini. Beyond co-localization: inferring spatial interactions between sub-cellular structures from microscopy images. *BMC Bioinformatics*, 11(1):372, 2010.

[73] B Herman, R Krishnan, and V Centonze. Microscopic analysis of fluorescence resonance energy transfer (FRET). *Methods in Molecular Biology*, 261:351–370, 2004.

[74] T Hori, J Nguyen, and S Uemoto. Progressive familial intrahepatic cholestasis. *Hepatobiliary & Pancreatic Diseases International*, 9(6):570–578, 2010.

[75] J Hornak. The basics of MRI. http://www.cis.rit.edu/htbooks/mri/. Last visited on 2013-05-20.

[76] E Hunter, J Price, and D Gough. Optimal filters for segmentation of microscope images. In *Proceedings of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 487–488, San Diego, CA, USA, 1993.

[77] Image Pro Plus. Powerful 2D and 3D image processing. http://www.mediacy.com/index.aspx?page=IPP. Last visited on 2013-05-20.

[78] Imaris. 3D and 4D Real-Time Interactive Data Visualization. http://www.bitplane.com/go/products/imaris. Last visited on 2013-05-20.

[79] P Iynedjian. Molecular physiology of mammalian glucokinase. *Cellular and Molecular Life Sciences*, 66(1):27–42, 2009.

[80] J Janáček, E Cvetko, L Kubínová, L Travnik, and I Eržen. A novel method for evaluation of capillarity in human skeletal muscles from confocal 3D images. *Microvasc Res*, 81(2): 231–238, 2011.

[81] T Jones, A Carpenter, and P Golland. Voronoi-based segmentation of cells on image manifolds. In Y Liu, T Jiang, and C Zhang, editors, *Computer Vision for Biomedical Image Applications*, volume 3765 of *Lecture Notes in Computer Science*, pages 535–543, Berlin, 2005. Springer.

[82] P Kankaanpää, K Pahajoki, V Marjomäki, J Heino, and D White. BioImageXD New Open Source Free Software for the Processing, Analysis and Visualization of Multidimensional Microscopic Images. *Microscopy Today*, 14(3):12–16, 2006.

[83] P Kankaanpää, L Paavolainen, S Tiitta, M Karjalainen, J Päivärinne, J Nieminen, V Marjomäki, and D White. BioImageXD: an open, general-purpose and high-throughput image-processing platform. *Nature Methods*, 9(7):683–689, 2012.

[84] H Kauczor, C Zechmann, B Stieltjes, and M Weber. Functional magnetic resonance imaging for defining the biological target volume. *Cancer Imaging*, 6(1):51–55, 2006.

[85] D Keppler and J König. Hepatic secretion of conjugated drugs and endogenous substances. *Seminars in Liver Diseases*, 20(3):265–272, 2000.

[86] A Kherlopian, T Song, Q Duan, M Neimark, M Po, J Gohagan, and A Laine. A review of imaging techniques for systems biology. *BMC Systems Biology*, 2:74, 2008.

[87] I Kim and et al. Differential regulation of bile acid homeostasis by the farnesoid x receptor in liver and intestine. *Journal of Lipid Research*, 48(12):2664–2672, 2007.

[88] Klinische Forschergruppe 217. "Hepatobiliärer Transport und Lebererkrankungen". http://www.dfg.de/foerderung/programme/listen/projektdetails/index.jsp?id=101434388. Last visited on 2013-05-20.

[89] M Kociński, A Klepaczko, A Materka, M Chekenya, and A Lundervold. 3D image texture analysis of simulated and real-world vascular trees. *Computer Methods and Programs in Biomedicine*, 107(2):140–154, 2012.

[90] J Kondrackiene and L Kupcinskas. Intrahepatic cholestasis of pregnancy-current achievements and unsolved problems. *World Journal of Gastroenterology*, 14(38):5781–5788, 2008.

[91] M Kozubek and P Matula. An efficient algorithm for measurement and correction of chromatic aberrations in fluorescence microscopy. *Journal of Microscopy*, 200(Pt 3):206–217, 2000.

[92] R Kubitz, D D' Urso, D Keppler, and D Häussinger. Osmodependent dynamic localization of the multidrug resistance protein 2 in the rat hepatocyte canalicular membrane. *Gastroenterology*, 113(5):1438–1442, 1997.

[93] R Kubitz, M Wettstein, U Warskulat, and D Häussinger. Regulation of the multidrug resistance protein 2 in the rat liver by lipopolysaccharide and dexamethasone. *Gastroenterology*, 116(2):401–410, 1999.

[94] R Kubitz, C Huth, M Schmitt, A Horbach, G Kullak-Ublick, and D Häussinger. Protein kinase C-dependent distribution of the Multidrug Resistance Protein 2 from the canalicular to the basolateral membrane in human HepG2 cells. *Hepatology*, 34(2):340–350, 2001.

[95] R Kubitz, G Sütfels, T Kühlkamp, R Kölling, and D Häussinger. Trafficking of the bile salt export pump from the Golgi to the canalicular membrane is regulated by the p38 MAP kinase. *Gastroenterology*, 126(2):541–553, 2004.

[96] G Kullak-Ublick. Drug-Induced Cholestatic Liver Disease. http://www.ncbi.nlm.nih.gov/books/NBK6102/. Last visited on 2013-05-20.

[97] A Kurz, D Graf, M Schmitt, S vom Dahl, and D Häussinger. Tauroursodesoxycholate-induced choleresis involves p38(MAPK) activation and translocation of the bile salt export pump in rats. *Gastroenterology*, 121(2):407–419, 2001.

[98] M Lamprecht, D Sabatini, and A Carpenter. Cellprofiler: free, versatile software for automated biological image analysis. *Biotechniques*, 42(1):71–75, 2007.

[99] R Larsen and M Marx. *An Introduction to Mathematical Statistics and Its Applications*. Prentice Hall, 3rd edition, 2000.

[100] Q Li and Z Deng. A Surface-based 3D Dendritic Spine Detection Approach from Confocal Microscopy Images. *IEEE Transactions on Image Processing*, 21(3):1223–1230, 2012.

[101] V Ljosa and A Carpenter. Introduction to the quantitative analysis of two-dimensional fluorescence microscopy images for cell-based screening. *PLoS Computational Biology*, 5 (12):e1000603, 2009.

[102] M Maeda, C Di Loreto, N Shirata, L Shih, M Cavaliere, F Longatto, and V Alves. Image analysis of nuclear/cytoplasmic ratio in cervical smears to discriminate three grades of cervical intraepithelial neoplasia. *Acta Cytolologica*, 41(3):744–748, 1997.

[103] E Manders, F Verbeek, and J Aten. Measurement of co-localization of objects in dual-colour confocal images. *Journal of Microscopy*, 169(3):375–382, 1993.

[104] M Massa, J Gagliardino, and F Francini. Liver glucokinase: An overview on the regulatory mechanisms of its activity. *International Union of Biochemistry and Molecular Biology Life*, 63(1):1–6, 2011.

[105] Y Meng, C Shaw, X Liu, M Altunbas, T Wang, L Chen, S Tu, S Kappadath, and C Lai. Comparison of two detector systems for cone beam CT small animal imaging - a preliminary study. In *Proceedings of the Society of Photo-Optical Instrumentation Engineers*, volume 6142, page 6142451, San Diego, CA, USA, 2006.

[106] P Mitchell. Clinical and translational research: Nursing scientists at the core. *Heart & Lung: the journal of critical care*, 41(5):435, 2012.

[107] T Mitchell. *Machine Learning*. McGraw-Hill Science/Engineering/Math, 1st edition, 1997.

[108] C Montero. The antigen-antibody reaction in immunohistochemistry. *Journal of Histochemistry and Cytochemistry*, 51(1):14, 2003.

[109] M Morelock, E Hunter, T Moran, S Heynen, C Laris, M Thieleking, M Akong, I Mikic, S Callaway, R DeLeon, A Goodacre, D Zacharias, and J Price. Statistics of assay validation in high throughput cell imaging of nuclear factor kappaB nuclear translocation. *Assay and Drug Development Technologies*, 3(5):483–499, 2005.

[110] A Mühlfeld, R Kubitz, O Dransfeld, D Häussinger, and M Wettstein. Taurine supplementation induces Mrp2 and Bsep expression in rats and prevents endotoxin-induced cholestasis. *Archives of Biochemistry and Biophysics*, 413(1):32–40, 2003.

[111] S Mühlfeld, O Domanova, T Berlage, C Stross, A Helmer, V Keitel, D Häussinger, and R Kubitz. Short-term feedback regulation of bile salt uptake by bile salts in rat liver. *Hepatology*, 56(6):2387–2397, 2012.

[112] D Murphy and M Davidson. *Fundamentals of light microscopy and electronic imaging*. Hoboken, N.J. Wiley-Blackwell, 2013.

[113] R Murphy. Putting proteins on the map. *Nature Biotechnology*, 24(10):1223–1224, 2006.

[114] A Nischwitz, M Fischer, and P Haberäcker. *Computergrafik und Bildverarbeitung*. Vieweg, Wiesbaden, 2007.

[115] J Noe, B Stieger, and P Meier. Functional expression of the canalicular bile salt export pump of human liver. *Gastroenterology*, 123(5):1659–1666, 2002.

[116] M Noursadeghi, J Tsang, T Haustein, R Miller, B Chain, and D Katz. Quantitative imaging assay for NF-kappaB nuclear translocation in primary human macrophages. *Journal of Immunological Methods*, 329(1-2):194–200, 2008.

[117] Obstetric Cholestasis. http://www.patient.co.uk/health/obstetric-cholestasis. Last visited on 2013-05-20.

[118] Institute of Laboratory Animal Resources. Guide for the Care and Use of Laboratory Animals. (NIH) 86-23, 1985. Department of Health, Education, and Welfare, Public Health Service, National Institutes of Health.

[119] A Olokoba, O Obateru, and L Olokoba. Type 2 Diabetes Mellitus: A Review of Current Trends. *Oman Medical Journal*, 27(4):269–273, 2012.

[120] OMIM Database. Maturity-onset diabetes of the young. http://www.omim.org/entry/606391. Last visited on 2013-05-20.

[121] N Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

[122] A Papoulis. *Probability, Random Variables, and Stochastic Processes*. Mc-Graw Hill, 1984.

[123] R Pautler and S Fraser. The year(s) of the contrast agent - micro-mri in the new millennium. *Current Opinion in Immunology*, 15(4):385–392, 2003.

[124] J Pawley. *Handbook of Biological Confocal Microscopy*. Springer, New York, NY, USA, 3rd edition, 2006.

[125] S Pierre and K Scholich. Toponomics: studying protein-protein interactions and protein networks in intact tissue. *Molecular Biosystems*, 6(4):641–647, 2010.

[126] S Pierre, C Maeurer, O Coste, W Becker, A Schmidtko, S Holland, C Wittpoth, G Geisslinger, and K Scholich. Toponomics analysis of functional interactions of the ubiquitin ligase PAM (Protein Associated with Myc) during spinal nociceptive processing. *Molecular & Cellular Proteomics*, 7(12):2475–2485, 2008.

[127] A Pinidiyaarachchi and C Wählby. Seeded Watersheds for Combined Segmentation and Tracking of Cells Image Analysis and Processing ICIAP 2005. volume 3617 of *Lecture Notes in Computer Science*, chapter 41, pages 336–343. Springer Berlin / Heidelberg, Berlin, Heidelberg, 2005.

[128] I Pitas. *Digital Image Processing Algorithms and Applications*. John Wiley & Sons, New York, NY, USA, 1st edition, 2000.

[129] T Pommerencke, T Steinberg, H Dickhaus, P Tomakidi, and N Grabe. Nuclear staining and relative distance for quantifying epidermal differentiation in biomarker expression profiling. *BMC Bioinformatics*, 9:473, 2008.

[130] W Pratt. *Digital Image Processing*. Wiley-Interscience, New York, NY, USA, 1978.

[131] W Press, S Teukolsky, W Vetterling, and B Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, UK, 2nd edition, 1992.

[132] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008.

[133] E Rexhepaj, K Jirstrom, D O'Connor, S O'Brien, G Landberg, M Duffy, D Brennan, and W Gallagher. Validation of cytoplasmic-to-nuclear ratio of survivin as an indicator of improved prognosis in breast cancer. *BMC Cancer*, 10:639, 2010.

[134] X Ronot and Y Usson. Imaging of nucleic acids and quantitation in photonic microscopy. In *Methods in Visualization*, volume 5, New York, NY, USA, 2001. CRC Press.

[135] D Rost, J Kartenbeck, and D Keppler. Changes in the localization of the rat canalicular conjugate export pump Mrp2 in phalloidin-induced cholestasis. *Hepatology*, 29(3):814–821, 1999.

[136] P Ruusuvuori, T Aijö, S Chowdhury, C Garmendia-Torres, J Selinummi, M Birbaumer, A Dudley, L Pelkmans, and O Yli-Harja. Evaluation of methods for detection of fluorescence labeled subcellular objects in microscope images. *BMC Bioinformatics*, 11:248, 2010.

[137] L Saba and G Mallarini. Differences between MIP and MPR techniques in the carotid artery stenosis degree measurement. Evaluation using multi-detector-row CT angiograph. *Minerva Cardioangiologica*, 56(1):21–27, 2008.

[138] I Sbalzarini and P Koumoutsakos. Feature point tracking and trajectory analysis for video imaging in cell biology. *Journal of Structural Biology*, 151(2):182–195, 2005.

[139] K Schauer, T Duong, K Bleakley, S Bardin, M Bornens, and B Goud. Probabilistic density maps to study global endomembrane organization. *Nature Methods*, 7(7):560–566, 2010.

[140] S Scheede, A Herpens, F Burmeister, B Oltrogge, K Saenger, T Schmidt-Rose, V Schreiner, H Wenck, T Knieps, and T Berlage. Qualification of a new and precise automatic tool for the assessment of hair diameters in phototrichograms. *Skin Research and Technology*, 17(2):186–195, 2011.

[141] M Schmitt, R Kubitz, M Wettstein, S vom Dahl, and D Häussinger. Retrieval of the mrp2 gene encoded conjugate export pump from the canalicular membrane contributes to cholestasis induced by tert-butyl hydroperoxide and chloro-dinitrobenzene. *Biological Chemistry*, 381(5-6):487–495, 2000.

[142] M Schmitt, R Kubitz, S Lizun, M Wettstein, and D Häussinger. Regulation of the dynamic localization of the rat Bsep gene-encoded bile salt export pump by anisoosmolarity. *Hepatology*, 33(3):509–518, 2001.

[143] D Schmolze, C Standley, K Fogarty, and A Fischer. Advances in microscopy techniques. *Archives of Pathology & Laboratory Medicine*, 135(2):255–263, 2011.

[144] C Schneider, W Rasband, and K Eliceiri. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7):671–675, 2012.

[145] W Schubert. Topological proteomics, toponomics, MELK technology. *Advances in Biochemical Engineering/Biotechnology*, 83:189–209, 2003.

[146] W Schubert, Bonnekoh B, A Pommer, L Philipsen, R Böckelmann, Y Malykh, H Gollnick, M Friedenberger, M Bode, and A Dress. Analyzing proteome topology and function by automated multidimensional fluorescence microscopy. *Nature Biotechnology*, 24(10):1270–1278, 2006.

[147] R Schwarz, T Berlage, and T Knieps. Trainierbare Bildanalyse für High-Content-Screening. *LaborPraxis*, 30(3):22–25, 2006.

[148] D Shaken and A Bruckstein. Pruning medial axes. *Computer Vision and Image Understanding*, 69(2):156–169, 1998.

[149] A Shariff, J Kangas, L Coelho, S Quinn, and R Murphy. Automated image analysis for high-content screening and analysis. *Journal of Biomolecular Screening*, 15(7):726–734, 2010.

[150] M Shils, M Shike, A Ross, B Caballero, and R Cousins, editors. *Modern Nutrition in Health and Disease*. Lippincott Williams & Wilkins, 10th edition, 2005.

[151] L Sklar, M Carter, and B Edwards. Flow cytometry for drug discovery, receptor pharmacology and high-throughput screening. *Current Opinion in Pharmacology*, 7(5):527–534, 2007.

[152] W Snyder and H Qi. *Machine vision*. Cambridge University Press, Cambridge, 2004.

[153] P Soille. *Morphological Image Analysis: Principles and Applications*. Springer-Verlag, New York, NY, USA, 2003.

[154] H Soleimani and M Khosravifard. Reducing interpolation artifacts for mutual information based image registration. *Journal of Medical Signals and Sensors*, 1(3):177–183, 2011.

[155] Solvo Biotechnology. Liver. http://www.solvo.jp/Solvo%20Solutions/liver.html?lang=jp. Last visited on 2013-05-20.

[156] M Sonka, V Hlavac, and R Boyle. *Image Processing: Analysis and Machine Vision*. Chapman & Hall, 2nd edition, 1998.

[157] D Stoffler, M Steinmetz, and U Aebi. Imaging biological matter across dimensions: from cells to molecules and atoms. *FASEB Journal*, 13(Suppl 2):S195–S200, 1999.

[158] C Stross, A Helmer, K Weissenberger, B Görg, V Keitel, D Häussinger, and R Kubitz. Protein kinase C induces endocytosis of the sodium taurocholate cotransporting polypeptide. *American Journal of Physiology - Gastrointestinal and Liver Physiology*, 299(2): G320–328, 2010.

[159] N Stuurman and J Swedlow. Software tools, data structures, and interfaces for microscope imaging. *Cold Spring Harbor Protocols*, 2012(1):50–61, 2012.

[160] M Sullivan-Gunn, E Hinch, V Vaughan, and P Lewandowski. Choosing a stable housekeeping gene and protein is essential in generating valid gene and protein expression results. *British Journal of Cancer*, 104(6):1055, 2011.

[161] H Suzuki and Y Sugiyama. Transport of drugs across the hepatic sinusoidal membrane: sinusoidal drug influx and efflux in the liver. *Seminars in Liver Diseases*, 20(3):251–263, 2000.

[162] D Taatjes, M Wadsworth, A Quinn, J Rand, E Bovill, and B Sobel. Imaging aspects of cardiovascular disease at the cell and molecular level. *Histochemistry and Cell Biology*, 130(2):235–245, 2008.

[163] A Tarca, V Carey, X Chen, R Romero, and S Drăghici. Machine Learning and Its Applications to Biology. *PLoS Computational Biology*, 3(6):e116, 2007.

[164] N Tavoloni. The intrahepatic biliary epithelium: an area of growing interest in hepatology. *Seminars in Liver Diseases*, 7(4):280–292, 1987.

[165] TopoScan. From platform to systems. http://www.fit.fraunhofer.de/en/fb/life/projects/toposcan2.html. Last visited on 2013-05-20.

[166] Translational Research Working Group. http://www.cancer.gov/researchandfunding/trwg/TRWG-definition-and-TR-continuum. Last visited on 2013-05-20.

[167] M Trauner, P Meier, and J Boyer. Molecular pathogenesis of cholestasis. *The New England Journal of Medicine*, 339(17):1217–1227, 1998.

[168] J Tukey. *Exploratory Data Analysis*. Addison-Wesley, 1977.

[169] B Uhal and K Roehrig. Effect of dietary state on hepatocyte size. *Bioscience Reports*, 2 (12):1003–1007, 1982.

[170] M Vaxillaire and P Froguel. Monogenic diabetes in the young, pharmacogenetics and relevance to multifactorial forms of type 2 diabetes. *Endocrine Reviews*, 29(3):254–264, 2008.

[171] F Watanabe and E Furuya. Quantitative image analysis reveals that phosphorylation of liver-type isozyme of fructose-6-phosphate 2-kinase/fructose-2,6-bisphosphatase does not affect nuclear translocation of glucokinase in rat primary hepatocytes. *Journal of Biochemistry*, 148(6):713–719, 2010.

[172] Weltchronik. Albert Einstein Biografie. http://www.weltchronik.de/ws/bio/e/einsteinA/ea01955a-EinsteinAlbert-18790314b-19550418d.htm. Last visited on 2013-05-29.

[173] What-when-how. In Depth Tutorials and Information. http://what-when-how.com/artificial-intelligence/morphological-filtering-principles-artificial-intelligence/. Last visited on 2013-05-20.

[174] F Wilcoxon. Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, 1(6): 80–83, 1945.

[175] T Wilson. Optical sectioning in fluorescence microscopy. *Journal of Microscopy*, 242(2): 111–116, 2011.

[176] H Witten and E Frank. *Data mining: Practical machine learning tools and techniques*. Elsevier, Amsterdam, 2nd edition, 2008.

[177] Zeta. Trainable image analysis for high content analysis. http://www.fit.fraunhofer.de/de/fb/life/projects/zeta.html. Last visited on 2013-05-20.

[178] Y Zhang, X Zhou, J Lu, J Lichtman, D Adjeroh, and S Wong. 3d axon structure extraction and analysis in confocal fluorescence microscopy images. *Neural Computation*, 20:1899–1927, 2008.

[179] J Zhou and H Peng. Automatic recognition and annotation of gene expression patterns of fly embryos. *Bioinformatics*, 23(5):589–596, 2007.

[180] M Zhou and T Veenstra. Multiplexed immunofluorescence for the interrogation of cellular protein complexes. *Expert Review of Proteomics*, 3(6):581–583, 2006.

[181] J Zock. Applications of high content screening in life science research. *Combinatorial Chemistry & High Throughput Screening*, 12(9):870–876, 2009.