universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Two methods of automized quantification of peroxisomal matrix protein import by fluorescence image analysis“

verfasst von / submitted by

Maximilian Bauer, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2020 / Vienna, 2020

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 834
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Molekulare Biologie UG2002
Betreut von / Supervisor:	Univ.Prof. Dr. Johannes Berger

Inhalt

1 Abstract

Peroxisomes are single membrane bound organelles enclosing various important metabolic pathways. The peroxisomal matrix protein import is mediated by two targeting signals namely peroxisomal targeting signal 1 (PTS1, C-terminal) and peroxisomal targeting signal 2 (PTS2, N-terminal). The PTS1 is bound by the receptor PEX5 and PTS2 is bound by a dimer consisting of the receptor PEX7 and co-receptor PEX5 which are then able to cross the peroxisomal membrane.

The efficiency of peroxisomal matrix protein import depends on the quality of the PTS. To determine the quality of individual PTS, a visual evaluation of import efficiency in cells expressing reporter proteins has originally been used. In this study, we want to present a first step into the direction of automated image analysis of fluorescence images. On the one hand, peroxisomes were first identified by a segmentation algorithm called Trainable Weka Segmentation and the average of the intensity maxima of these areas was compared to the average intensities of the non-peroxisomal cellular background. On the other hand, the analysis of the intensity distribution of all pixels of each cell was used to discriminate cells presenting with different import efficiencies of the reporter protein, independently of the different shapes of cells.

The results of the segmentation approach demonstrate significant differences between the mean intensities of peroxisomal and non-peroxisomal areas within cells. The results of the analysis of the pixel-based intensity distribution suggest that the shape of the distribution curves reflect the visual evaluation of import efficiency. Moreover, the total amount of the reporter protein has a great impact on the evaluation of import efficiency.

In summary, both methods presented in this thesis show promising results and are good candidates for further testing and improvement.

2 Introduction

2.1 Peroxisomes

Peroxisomes are small vesicles surrounded by a single membrane and were first reported in a light microscopy study on kidney tissue (Rhodin, 1954). Peroxisomes (mammals, fungi), glycosomes (protozoa) and glyoxysomes (plants) are organelles, which belong to the microbody family. They are ubiquitous subcellular organelles which lack nucleic acids.

2.2 Metabolism

Peroxisomal matrix protein import is essential for the functioning of the many metabolic pathways occurring in peroxisomes such as detoxification of glyoxylate or the biosynthesis of ether phospholipids. Here, I want to shortly describe two of the peroxisomal pathways, the fatty acid alpha and beta oxidation.

Peroxisomal fatty acid alpha-oxidation was first described studying skin fibroblasts (Singh *et al.*, 1992). The primary substrate which is used in the peroxisomal fatty acid alpha-oxidation is phytanic acid. Phytanic acid has to be activated to phytanoyl-CoA prior to alpha-oxidation (Watkins, Howard & Mihalik, 1994). The phytanic acid pathway is predominantly, if not exclusively, peroxisomal (Jansen & Wanders, 2006).

In plants and yeast, peroxisomes are the sole place of **fatty acid beta-oxidation**, whereas in humans it occurs both in peroxisomes and mitochondria. In humans, the substrates for the fatty acid beta oxidation in peroxisomes and mitochondria are medium and long chain fatty acids. The very long chain fatty acids, however, can only be degraded in peroxisomes (Inestrosa, Bronfman & Leighton, 1979; Waterham, Ferdinandusse & Wanders, 2016; Poirier *et al.*, 2006).

2.3 Peroxisomal life cycle

Since the discovery of peroxisomes, there has been a debate whether peroxisomes are constantly formed de novo or are maintained by growth and division. Present, a hybrid of both models is widely accepted (Figure 1). Genes encoding proteins important for the biogenesis and function of peroxisomes are called peroxisome biogenesis factors (PEX) (Fujiki *et al.*, 2014).

Physiological regulation of de novo peroxisome biogenesis remains unclear, but it is suggested that the endoplasmic reticulum is involved (Smith & Aitchison, 2013). The other model, based on growth and division, states that peroxisomes form

by a process involving elongation, constriction and fission of pre-existing peroxisomes. The most important peroxin within this model is PEX11, as an overexpression of PEX11 leads to an increased amount of smaller peroxisomes compared to the wild-type (Honsho, Yamashita & Fujiki, 2016). Part of the peroxisome homeostasis is the degradation system called pexophagy, which is a special form of autophagy that is selective for peroxisomes. The currently knowledge is that ubiquitin is involved (Cho *et al.*, 2018; Kim *et al.*, 2008).

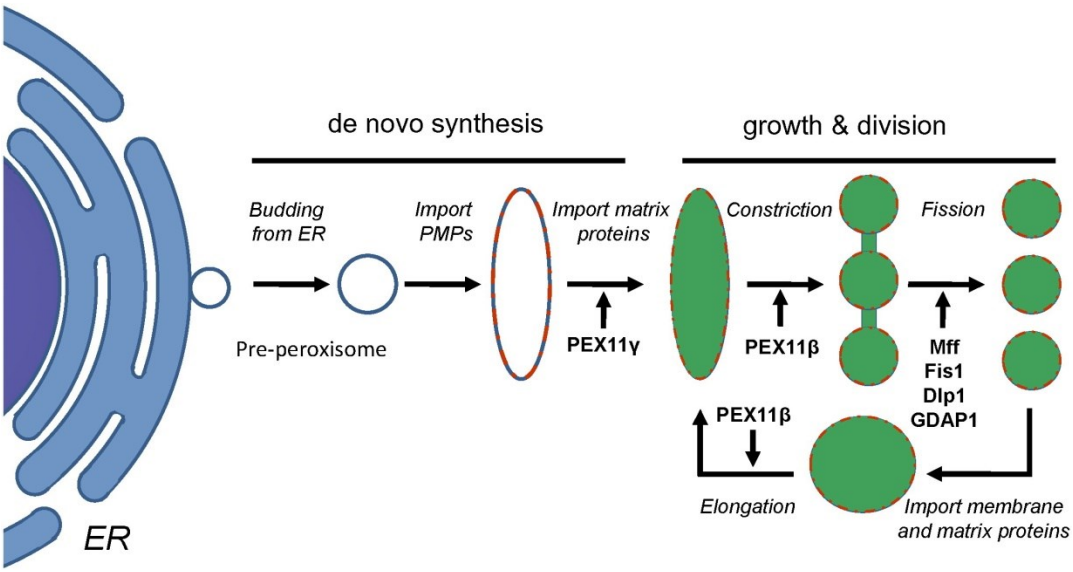


Figure 1. Peroxisomal life cycle. Two models, the de novo synthesis and the growth and division model, were combined to form a hybrid model. The first model states that peroxisomes can form de novo from vesicles that originate from the endoplasmic reticulum (ER). The latter model includes the growth and division of peroxisomes by the three steps of elongation, constriction and fission (Waterham, Ferdinandusse & Wanders, 2016).

2.4 Peroxisomal matrix protein import

2.4.1 Peroxisomal targeting signals and receptors

Peroxisomal matrix protein import is initiated by the recognition of a peroxisomal targeting signal (PTS) by a receptor protein. There are two types of targeting signals for peroxisomal matrix proteins: the peroxisomal targeting signal 1 (PTS1) and the peroxisomal targeting signal 2 (PTS2). PTS1 is located at the extreme C-terminus (Gould, Keller & Subramani, 1987; Miyazawa *et al.*, 1989) of the encoding proteins while PTS2 is located near the N-terminus of proteins (Osumi, Tsukamoto & Hata, 1992; Swinkels *et al.*, 1991).

The early consensus sequence for PTS1 was (S/A/C)-(K/H/R)-(L/M) (Gould, Keller & Subramani, 1987). Later, a broader number of tripeptides were found to act

as targeting signals. Additionally, the upstream sequence of the PTS1 has been found to modulate binding strength with its import receptor, PEX5 (Lametschwandtner *et al.*, 1998; Brocard & Hartig, 2006).

The earliest consensus sequence found for the PTS2 was (R/K)(L/V/I)X_5(Q/H)(L/A) (Glover *et al.*, 1994). Bioinformatics approaches suggest R-(L/V/I/Q)-X-X-(L/V/I/H)-(L/D/G/A)-X-(H/Q)-(L/A) as a consensus sequence for the most common PTS2 variants and (R/K)-(L/V/I/Q)-X-X-(L/V/I/H/Q)-(L/S/G/A/K)-X-(H/Q)-(L/A/F) including essentially all naturally occurring variants of PTS2 (Petriv *et al.*, 2004).

The receptors responsible for the peroxisomal import are PEX5 for proteins with PTS1 (Van der Leij *et al.*, 1993; Brocard *et al.*, 1994) and PEX7 (Marzioch *et al.*, 1994; Braverman *et al.*, 1997) together with a co-receptor, which in most multicellular eukaryotes is the PTS1 receptor PEX5, for proteins with PTS2. There are two splicing variants of PEX5, PEX5S and PEX5L, but only the latter interacts with PEX7 (Dodt *et al.*, 1995).

In PTS2 mediated protein transport, the additional binding of the co-receptor PEX5L to the receptor PEX7 and the cargo protein leads to an increase in interaction strength between all three binding partners (Figure 2, (Kunze *et al.*, 2015)). The binding of PEX5L to the dimeric PEX7-PTS2 complex initiates the transportation of the cargo-receptor complex to the peroxisomal membrane (Kunze *et al.*, 2015).

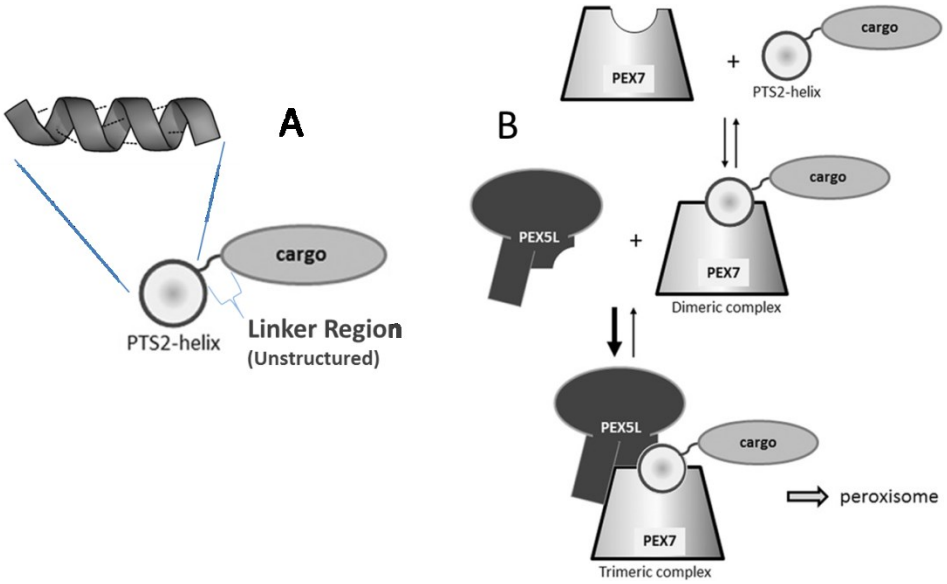


Figure 2. Model of PTS2-mediated peroxisomal protein import. Schematic of the PTS2 which is depicted in a frontal view on the alpha helix, which is connected to the core protein via an unstructured linker region (A). The interaction between PTS2, PEX7 and PEX5L is necessary for protein import (B) (adjusted from (Kunze *et al.*, 2015)).

2.4.2 Import of peroxisomal protein-receptor complexes

Peroxisomal proteins can be imported in a fully folded state through translocons, whereas mitochondrial proteins can only be transported unfolded through gated channels within the membrane (Figure 3). The transport of matrix proteins across the peroxisomal membrane starts with the docking of PTS1 carrying proteins to PEX5 or PTS2 carrying proteins to PEX7 and PEX5L. The dimeric or trimeric complex can then dock to the docking complex, which consists of PEX13 and PEX14 (Albertini *et al.*, 1997; Girzalsky *et al.*, 1999; Waterham & Ebberink, 2012). The exact translocation mechanism remains elusive, but one model describes a dynamic transient pore that disassembles after translocation (Erdmann & Schliebs, 2005). Following translocation of the cargo across the membrane by the docking complex, the cargo is released inside the peroxisome. The receptors PEX5S, PEX7 and PEX5L are then exported back to the cytosol in an ATP- and ubiquitin dependent process (Figure 3).

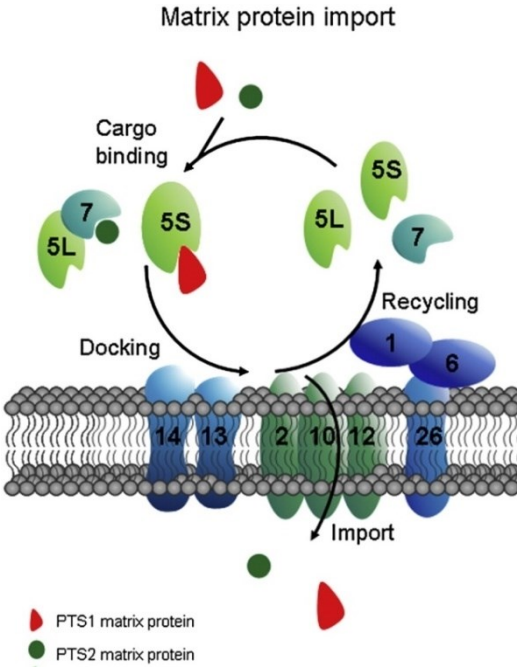


Figure 3. Schematic representation of peroxisomal matrix protein import. Steps of protein import: (i) recognition of a peroxisomal import signal by a receptor (cargo binding), (ii) docking of the cargo-receptor complex (docking), (iii) import of cargo protein (import) and (iv) recycling of receptors (recycling) (acquired from (Waterham & Ebberink, 2012)).

2.5 Matrix protein import efficiency as a proxy for the quality of peroxisomal targeting signals

The quality of peroxisomal targeting signals in living cells can currently not be measured directly. Therefore, an indirect approach such as peroxisomal matrix protein efficiency is needed. The hypothesis states that the higher the efficiency of peroxisomal matrix protein import, the higher the quality of the peroxisomal targeting signal. While an efficient targeting signal will lead to almost exclusive localization of proteins to peroxisomes, weaker signals might lead to an accumulation of peroxisomal matrix proteins in the cytoplasm. This is partially due to hypothesized saturation of proteins within peroxisomes and partially due to competition with proteins with stronger import signals. In this study, we used a peroxisomal targeting signal fused to EGFP for analyzing the intensity distribution within cells by fluorescence microscopy.

2.6 Fluorescence microscopy

2.6.1 Epi-fluorescence microscopy as a technology still valuable today

The microscope of choice was an inverted epi-fluorescence microscope specified in the materials and methods section. Although newer technologies are available, the epi-fluorescence microscope remains valuable at the current moment. In comparison with a laser scanning confocal microscope, the advantages of the epi-fluorescence microscope are that it is easier to use, cheaper and structurally less complicated with faster imaging time. Due to the lack of a pinhole, the epi-fluorescence microscope can gather more light, also leading to better results for low light applications (Webb & Brown, 2013). This in term leads to a reduction in bleaching of the specimen at the same light source output level and exposure time. The lack of the pinhole is also the greatest disadvantage of the epi-fluorescence microscope. This means that while the point-spread-function of the x and y directions are comparable between the two microscope types, the z direction spread is increased when using epi-fluorescence microscopy. Therefore, the epi-fluorescence microscope works best with thin specimen like COS7 cells (<20-30 µm in the z-axis) (Sagi, Basser & Assaf, 2009).

2.6.2 Signal to noise ratio

When excitation light from the light source of the fluorescence microscope comes in contact with the right type of fluorophore, light in form of photons is emitted that can then hit the sensor of the camera. The more fluorophores there are in one place, the more photons are emitted, which corresponds to a higher signal, and thus different pixels present with varying intensities in an image. Noise is the by-product of every signal and represents random distributions of intensities throughout an image. The goal of imaging is to keep it as low as possible. The three major types of noise are (i) the result of the variation of arrival rate of photons (photon noise), (ii) the thermal generation of electrons within the camera (dark noise) and (iii) noise can be produced during the conversion of photon signal into voltage signal (read noise). Methods to reduce noise and therefore maximize signal to noise ratio are optimizing the labelling method of the specimen, cooling the camera and increasing the exposure time (Thomas J. Fellers, no date).

2.7 Image processing in general

Image processing stands for the analysis of images regarding their intensity information and can also include separation of different regions within images (image

segmentation). It is a tool used by researchers for decades now, but in recent years advances in computer power and methodology have made this field very powerful. The intensity distribution within the images can be analysed and extracted. The classical approach is to compare the intensities of peroxisomes to the intensities of the cytoplasm to determine a measure for the effectiveness of the peroxisomal matrix protein import.

2.7.1 Advantages and disadvantages of automated image processing

There are several advantages with automated image processing. First, (i) tasks like peroxisome counting are completed faster than manually. Another advantage (ii) is the reproducibility of the results from the automatic analysis, circumventing the subjective aspect of visual evaluation. The automated program will always output the same number of peroxisomes, even when re-run several years later. A manual count will make mistakes due to day-to-day variability of the same experimenter. Additionally, there is no interpersonal variability when using the automated image analysis.

The disadvantages are that the development of a new program can take a long time and mistakes within the program can always happen during that period, which are often difficult to detect. Therefore, internal controls are essential. It is also important not to absolutely trust the results of an automated program without having at least basic understanding of how it works and a basic test system.

2.7.2 Image filters

Image filters can be used to process images in various ways dependent on the desired outcome. Filters can be considered as fields of certain size, which combine the intensity values of a predetermined number of pixels by calculation and replace the original intensity information of said pixel with the result of the calculation. This calculation is then applied to each pixel in the image. Filters can reduce noise (e.g. *Gaussian* filter), increase the visibility of edges (e.g. *Difference of Gaussian* filter) or highlight texture based features of images (e.g. *Median* filter) (Bankhead, 2014).

2.7.3 Classical image segmentation

Image segmentation is the separation of objects of interest from other sections of an image. There are a manifold of different methods to achieve the highlighting of objects of interest, but it greatly depends on the sample at hand which method is best (Kang, Yang & Liang, 2009). Classical methods use comparably low computer

resources and are still valuable today. The simplest method of image segmentation is thresholding. Threshold segmentation divides the image into foreground and background. Thresholding can be done manual or automatic, for example with *Otsu's* method (Otsu, 1979). It is best suited for large objects within images, like stained nuclei of cells. Smaller objects such as peroxisomes and mitochondria are more difficult to separate, because the area for detection is limited and the objects can be clustered. For smaller objects, edge detection filters like *Sobel*, *Difference of Gaussian* or a combination of different filters would be more applicable. The segmentation method introduced in this study combines classical image segmentation filters with the use of machine learning. The algorithm called Trainable Weka Segmentation (TWS) is implemented in ImageJ and uses manual training data to create classifiers used for segmentation (Arganda-Carreras *et al.*, 2017).

2.7.4 Machine learning segmentation

Machine learning stands for a collection of algorithms that perform intelligent predictions based on a data set. Machine learning algorithms, also called models, are in use for a variety of tasks like image analysis or general analysis of data lists. There are two main types of machine learning algorithms, supervised and unsupervised. In this study, we want to focus on supervised learning. Supervised learning describes the process of teaching a model with input data of choice. Most of the imaging models are classifier algorithms that are trained on the objects of interest and tested on images of the same or different origin as the training images (Nichols, Chan & Baker, 2019).

3 Aim of this Study

The aim of this study was to use two different methods to evaluate the efficiency of peroxisomal matrix protein import in eukaryotic cells using automatic quantification of fluorescence microscopic images of cells incubated with antibodies targeting PMP70 or transfected with EGFP-PTS1 or PTS2-EGFP expression plasmids.

To circumvent the problem of different expression level of the proteins of interest, we manually categorized the cells according to their peroxisomal matrix protein efficiency prior to the automatic image analysis. The process of manual categorization, however, is very time-consuming and the categorization by different experimenters can lead to different results.

Therefore, we wanted to increase efficiency and repeatability of the categorizations by using computational methods. Two approaches were chosen for the computational image analysis, one based on pre-classification of peroxisomal and non-peroxisomal areas and the other an unbiased analysis of all pixels composing the images of cells.

The **first** method uses the differences of peroxisomal intensities in comparison to the intensity of the cytoplasm. This method relies on the correct finding of peroxisomes through a process called image segmentation. To segment images, the task was to search for a free-ware segmentation program, to test its reliability in finding peroxisomes and to compare it to other freeware programs.

The **second** method uses the distribution of intensities within images of the whole cell without segmentation. The individual pixel intensities of each cell were ordered from highest to lowest to form a plot. Within this plot, we then hypothesised that pixels with the highest intensities represent the intensities of peroxisomal areas, the centre of the curve represent intensities of cytoplasmic areas and the last part of intensities stem from pixels at the border regions of the cell.

Finally, the results obtained by the two methods should be compared as a final control.

4 Materials and Methods

4.1 Reporter constructs

The PTS2 reporter construct consist of the genetic information to express the first thirty amino acids of the N-terminus of the PTS2 carrying protein 3-ketoacyl-CoA-thiolase of the organism Rattus norwegicus and was fused to EGFP (thioase$_N$-EGFP) (Kunze *et al.*, 2011). The PTS1 reporter construct consists of the genetic information to express EGFP, which is fused to the C-terminal part of the sequence of ACOX-2 (Chong *et al.*, 2019).

4.2 Cell seeding and cell counting

The green monkey kidney cell line COS7 was purchased from ATCC. Cells were cultivated in DMEM (Sigma-Aldrich Art.No. D6429-500ml – high glucose) supplemented with 10% fetal calf serum (FCS), 250 µg/ml amphotericin (Sigma), 50 units/ml penicillin and 100 µg/ml streptomycin (BioWhittaker).

DMEM medium and trypsin were incubated for at least 30 minutes at 37 °C in a water bath. Cells were washed by replacing the medium inside of the T75 cell culture flasks with PBS. After removing the PBS from the flasks, 2 ml of trypsin was added and cells were incubated for at least 5 minutes at 37°C in the CO_2 incubator. Cells were then checked for detachment from the surface of the flask using the light microscope and then the trypsinization reaction was stopped by adding at least 4 ml of DMEM medium. After carefully mixing and taking up the content of the cell culture flask with a pipette, 1 ml of cell suspension was returned into the flask. The remaining cell suspension was either discarded in the day-to-day cell cultivation routine or placed in a 15 ml falcon tube for cell seeding.

For seeding, the cell suspension in the 15 ml falcon tube was centrifuged for 5 minutes. After discarding the supernatant, the pellet was resuspended in 4 ml DMEM medium. Prior to the cell counting, the 150 µm diameter capillary system of the CASY cell sorter was cleaned. To count the cells, 50 µl of suspension were removed and mixed with 10 ml of CASY-ton liquid. After the counting, $5x10^4$ cells were seeded per well (growing area 1.9 cm^2) of 24-well plates (Greiner Bio-One; Art.No.: 662160) and incubated in the CO_2 incubator for 24 hours.

4.3 Transfection

On the next day, we replaced the medium of the cell culture with 400 µl fresh DMEM medium. We then prepared 1 µg of DNA plasmids for each transfection. The DNA

plasmids were then mixed with 100 µl pure DMEM and 1.25 µl TurboFect (ThermoFisher) and the mix was incubated for 15 minutes at room temperature. Next, we added 100 µl of the TurboFect/plasmid mix to each of the wells and incubated the cells for 24 hours in the CO_2 incubator.

For our experiments, we diluted each expression plasmid with mammalian expression plasmid without expression (pcDNA3.1). Different combinations of pcDNA3.1 and expression plasmid for the reporter protein were used: 31 ng of PTS2 expression plasmid or 125 ng of PTS1 expression plasmid mixed with pcDNA3.1 to a total of 1000 ng.

4.4 Cell splitting, chemical fixation and mounting of glass coverslips

Following the incubation time after transfection, cells were washed with PBS and trypsinized (75 µl trypsin). Trypsinization was stopped after 5 minutes by adding 500 µl fresh DMEM medium and the cell suspension was distributed into two wells of 24 well plates containing glass coverslips (diameter ~1.3 cm^2). Cells were then incubated for 24 hours in the CO_2 incubator. Next, cells in 24 wells with glass coverslips were washed with PBS and chemically fixed with 1 ml of 3.7% (v/v) formaldehyde in PBS for 15 minutes. After the fixation, cells were washed twice with PBS to remove excess formaldehyde. Glass coverslips were then mounted on glass slides using 2.5 µl of Mowiol (https://www.nichd.nih.gov/about/org/dir/other-facilities/cores/microscopyandimaging/support/MOWIOL).

4.5 Immunofluorescence staining

After seeding COS7 cells into wells containing glass coverslips, cells were incubated for 24 hours in the CO_2 incubator. Next, cells in 24 wells with glass coverslips were washed with PBS and fixed with 1 ml of 3.7 % formaldehyde in PBS for 15 minutes. After fixation, cells were washed twice with PBS and then the protocol was continued on the same or the next day. Next, cells were incubated in PBS + 0.05 % Triton-X100 for 5 minutes. Subsequently, cells were washed with PBS and the buffer was removed. Next, parafilm was placed in a box and the glass slides were placed onto droplets of 50 µl of blocking solution (1x PBS, 5% FCS, 50ng/ml BSA, 0,02% NaN_3). The glass plates were then placed upside down onto the droplets and incubated for 1 hour at room temperature with low agitation. Next, cells were incubated with primary antibodies (rabbit anti-PMP70 diluted 1:2000, 1:4000 or 1:10000 in blocking solution) the same way as before for 3 hours. Glass coverslips were then washed with PBS three times using empty 12 well plates, incubated with dye labelled secondary

antibodies (donkey anti-mouse-Cy2 and donkey anti-rabbit-Cy3 both diluted 1:400 in blocking solution) for 1 hour the same way as before. Cells were washed three times in PBS and then glass coverslips were then shortly dipped in a beaker filled with dH_2O and fixed to glass slides with a drop of Mowiol (3-5 µl) per glass coverslip.

4.6 Manual categorization of cells according to their peroxisomal matrix protein import efficiency

Cells were sorted into one of four categories A, B1, B2 and C. The categories are as A (best import), B1 (good import), B2 (weak import) to C (no apparent import). Category A marks cells that appear as bright punctate clouds with a dark oval in the middle where the nucleus is located. This is rated as the best peroxisomal import with reporter proteins inside the cells having a functional PTS2 and without an overload of reporter proteins. Category B1 marks cells that show punctate staining as well as a faint cytoplasmic staining with lightly stained nucleus. Reporter proteins inside the cells have either a weaker PTS2 or a slight overload of reporter proteins. Category B2 includes cells which have a punctate staining in combination with a clear cytoplasmic staining and a visible nucleus. Cells are sorted into category B2 when a mutation in the PTS2 reduces import efficiency into peroxisomes or if the cells produce large amounts of reporter proteins. Category C has similar cytoplasmic and nuclear staining as category B2 but does not show visible punctate staining. Cells in category C have in common that the peroxisomes are not detectable anymore either due to a non-functioning PTS2 or a strong overexpression of the reporter proteins.

4.7 Microscopy

For detection of the reporter proteins or labelled antibodies, we used the Olympus inverted epi-fluorescence microscope IX71 (Olympus) equipped with a CCD camera (Olympus CAM-XM10). Additionally to the respective wavelengths used to excite the reporter proteins, appropriate filters were applied. C-M-Cell software (Olympus) was utilized to obtain images. As an objective lens, we used the Olympus 60x PlanApo N objective with the numerical aperture of 1.42 (normal oil).

4.8 Image Acquisition

Images were acquired with constant exposure time set to 300 ms at full dynamic range of the camera and a bit depth of either 12 bits (0-4096 grey scales) or 14 bits (0-16383 grey scales). The microscope focus was set manually for each cell. Each cell was moved to the centre of the screen before image acquisition to avert the

uneven illumination at the edges of the camera. The images were saved locally in the C-M-Cell software database and in the 16-bit TIFF lossless file format before transferring them to a different computer. Pixel size in µm for measurements was taken from the microscope settings determined by the manufacturer.

4.9 Quantitative Image Processing

Images were loaded into Fiji (the abbreviation stands for Fiji is just ImageJ), a version of ImageJ with additional features. Next, images were stacked and one cell per image was marked with a region of interest which was subsequently saved. The next step was to segment the images with the Trainable Weka Segmentation (TWS) Fiji plugin equipped with a previously generated classifier. The output of the TWS was an 8-bit colour image. Next, images were converted to 8-bit grey scale and then the binarize function in ImageJ was applied where peroxisomal areas were assigned the value of one (white) and non-peroxisomal areas the value of zero (black). After that, the watershed algorithm was applied to separate touching objects. Subsequently the region of interest of the whole cell, which was created in the beginning, was projected onto the image. Next, the particle analyser function within ImageJ counted the number of peroxisomes of the cell region of interest and saved each peroxisomal region of interest individually as well as a summary of information of all peroxisomal areas within each cell. Following this step, the peroxisomal regions of interest were applied on the original images and intensities were measured. A region of interest for the cytoplasm was created by subtracting the peroxisomal regions of the whole cell region of interest. Following the manual testing of this process, the steps were integrated into an ImageJ macro code. Last, the baseline of the intensity noise (mean of intensity of areas outside of cells pus three times standard deviation) was then subtracted from the numeric output of the intensity extraction.

4.10 Training of a classifier of the Trainable Weka Segmentation

The first step in training a classifier was to determine its settings. In the main interface in the settings tab, there are five filters that we used as a basis for the classifier, namely Gaussian blur, Sobel filter, Hessian, Difference of Gaussians and Median. The underlying algorithm used was the FastRandomForest. In the field Class names we named Class 1 as peroxisomes and Class 2 as non-peroxisomes.

Next, we used the draw tools implemented in ImageJ to draw either lines or areas into the images which were then subsequently saved by pressing either “add to

Peroxisomes" or "add to non-peroxisomes". For the training success, it is important that the number of chosen areas are similar in all groups.

After enough areas were saved, we pressed the button Train classifier to see which areas the program would highlight as peroxisomes. When the segmentation was satisfactory, the classifier was saved under Save classifier. Additionally, the image data were saved using Save data. To continue the training of a saved classifier, we loaded the data into the TWS segmentation, trained on images and repeated the process of saving classifier and data.

4.11 Intensity plots

In this method, we extracted the intensity of each pixel of an individual cell without image segmentation by using an ImageJ macro code. The resulting list of x and y pixel place coordinates and their respective intensities were copied into an excel sheet.

Next, the list was sorted in a descending order. The baseline of the intensity noise (mean of intensity of areas outside of cells pus three times its standard deviation) was then subtracted from the numeric output of the intensity extraction.

For the normalization of the intensities, we calculated the mean of the top 100 values with the highest intensities and subtracted the baseline of the intensity noise (mean of intensity of areas outside of cells pus three times its standard deviation) from this value. Then, we divided every value by this mean value and multiplied the resulting value by 100 to gain a percentage value.

For the normalization of the pixel number in percent, we divided each pixel number by the total number of pixels within a cell area.

Of these lists, the key values of choice were then extracted and saved separately. To gain a representative intensity plot, we extracted the first value of the list and the following values at each following percentage point of a total of ten cells. The mean and standard deviation was then calculated of each percentage step to form a mean intensity plot.

5 Results

5.1 Quantitative image analysis

5.1.1 Quality of peroxisomal matrix protein import

The quality of a PTS depends on its amino acid sequence. The hypothesis for our experiments was that the better the quality of the PTS, the higher the peroxisomal matrix protein import efficiency. In this study, we wanted to automatize a manual classification of peroxisomal matrix protein import efficiency, previously performed in our lab. The idea behind the categorization system stems from the observation that cells transfected with plasmids expressing PTS2-EGFP fusion proteins can have different fluorescence intensity distribution within cells. As during the transfection individual cells receive different quantities of plasmid by chance, even cells transfected with plasmids expressing well importing fusion proteins can present with high amounts of fluorescence intensities in their cytoplasm. This is usually positively correlated with clear visibility of the nucleus and nucleolus. In this study, the cytosol and the nucleus are referred to as non-peroxisomal areas. Sometimes, increased peroxisomal staining is not visible in these cells, probably because the bright cytoplasmic staining is covering the peroxisomal intensities. When transfecting and observing mutated PTS2-EGFP fusion proteins with lower peroxisomal matrix import efficiency, the quantity of cells having a signal in the cytoplasm was increased (results of my “Großpraktikum”, data not shown). In the following sections I want to elaborate on the use of fluorescence proteins and present a general introduction to later chapters of this study.

5.1.2 Fluorescence labelling within eukaryotic cells

The goal was to study the import efficiency of PTS2-carring proteins. One of the difficulties of this import system is a limit in import capacity causing a broad variability in apparent import efficiency among transfected cells.

Thus, we first establish the method using two fluorescence staining, an α-PMP70 immunostaining using Cy3-labelled secondary antibodies (Figure 4A) and an EGFP-PTS1 (Chong 2019) fusion protein (Figure 4B). The membrane protein PMP70 is classically used as an antibody target to highlight the location of peroxisomes. The antibody-mediated detection of endogenous proteins facilitates a very high signal to noise ratio. In contrast to the antibody-mediated approach, we used an EGFP-PTS1 fusion protein, which is soluble and localizes to peroxisomes (Figure 4B). The PTS1

fusion protein produces a high signal to noise ratio with peroxisomes, but can also show minor levels of the reporter protein located in non-peroxisomal areas. With its expression upon transfection of encoding plasmid, the soluble reporter protein provides a useful tool for the study of transfection-mediated variability. Finally, we used the PTS2-EGFP fusion protein (Kunze 2011). The protein presents with lower import efficiency and a low signal to noise ratio in transfected cells, presenting different import categories, which makes analysis of images of this type of fusion proteins more difficult (Figure 4C-F).

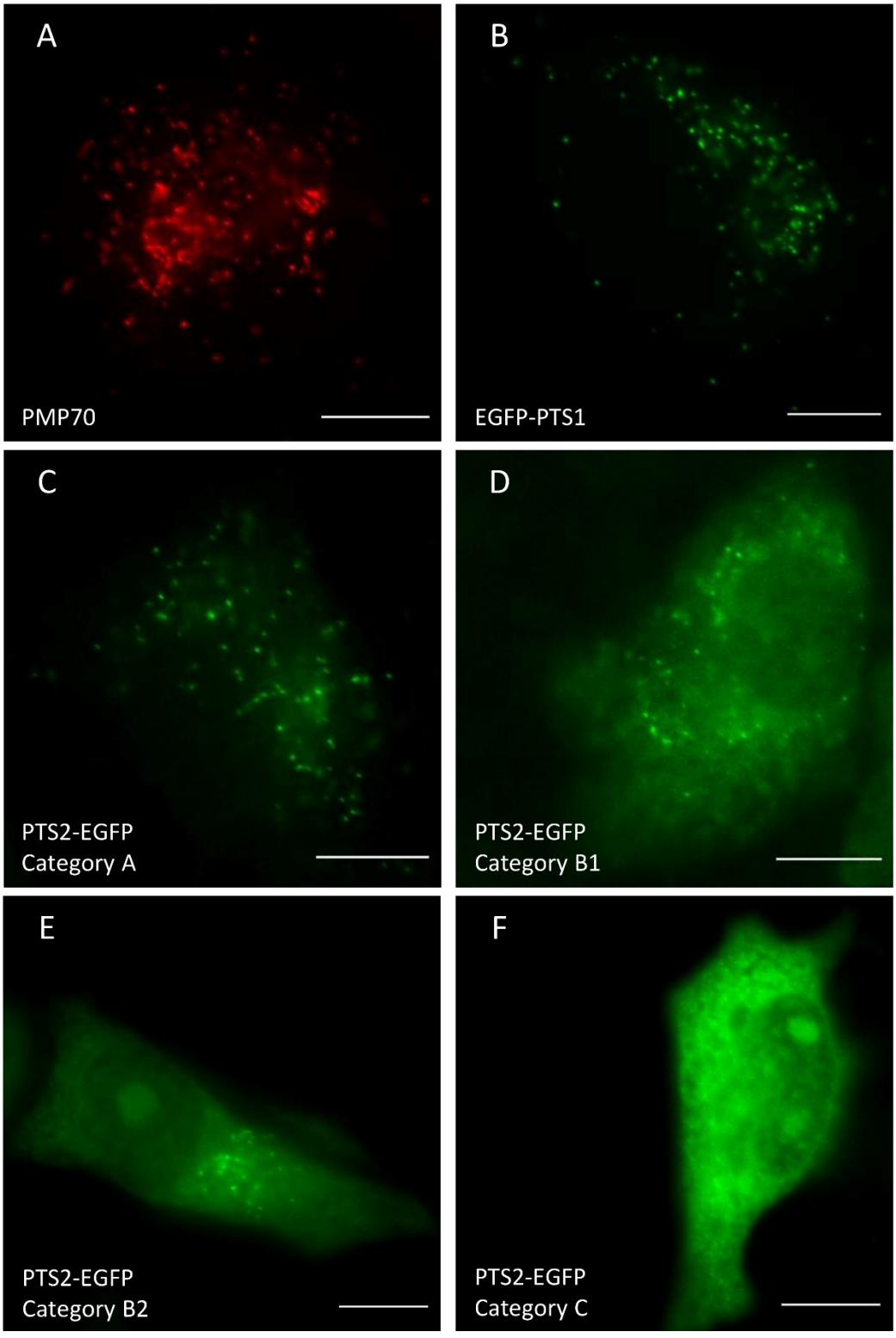


Figure 4. Images of fluorescent labelled cells. Fluorescence signals were created either by using α-PMP70 and Cy3-labelled secondary antibodies (A) or by transfecting plasmids coding for fusion proteins consisting of EGFP and PTS1 (B) or PTS2 and EGFP (C, D, E, F). Peroxisomal matrix protein import efficiency for PTS2-EGFP was categorized into one of four categories namely very good (B, C), good (D), poor (E) and no visible import (F) of peroxisomal matrix proteins. Scale bars shown in the images show a length of 10 µm. Gamma filter was used on all images. The cell type is COS7.

5.1.3 First, second and third dimension of intensities in images

The signals of the labelled secondary antibodies or EGFP fusion proteins can be captured by using a fluorescence microscope. Light signals captured by the camera installed in the fluorescence microscope are converted to electrical currents, stored as binary numbers and displayed with shades of grey in the form of squares. These squares are called pixels and show how much of a signal was captured in a certain area on the camera's sensor. For better visibility of the intensities, a so-called lookup table (LUT) was used to colourize the images where the pixel with the highest intensity is set to 100% (yellow) and the pixel with the lowest intensity is set to 0% (black) (Figure 5). Due to image brightness and gamma filter adjustments for better visibility, the cytoplasm also appears heavily stained. This is not the case, as when looking through the microscope, the cytoplasm is almost not visible.

Within an image, local information can be extracted by creating a so called line graph, which is plotting the intensities against a line in the xy-layer (Figure 5, middle section). Taking the x and y dimensions of a 2 dimensional image and plotting the intensity in the z-axis results in a 3 dimensional plot where the impact of the intensity is more clearly visible (Figure 5, right section). As seen in the 1 dimensional and 3 dimensional plots, the intensities of the pixels belonging to peroxisomes are increased in comparison to the cytosol. Within PMP70 images, only peroxisomes are labelled and the ratio between peroxisomal and cytoplasmic intensity is the highest (Figure 5, A). For the images of cells transfected with EGFP-PTS1 (Figure 5B) or PTS2-EGFP (Figure 5C) expressing plasmids, a fraction of the fusion proteins expressed within the cell remains in the cytoplasm while the majority of fusion proteins is imported into peroxisomes.

The mean intensity signal of peroxisomes is highest with PMP70, intermediate with EGFP-PTS1 and lowest with PTS2-EGFP, leading to a low signal to noise ratio with the latter fusion proteins. Noise does not increase linearly with the signal, therefore the lines in the line graph appear smoother with the PMP70 samples with high signal strength (Figure 5A) than in the PTS2-EGFP sample with low signal strength (Figure 5C). This is the reason why images of the PTS2-EGFP samples are more difficult to process and require a program that can cope with the low signal to noise ratio and is able to find the peroxisomal regions.

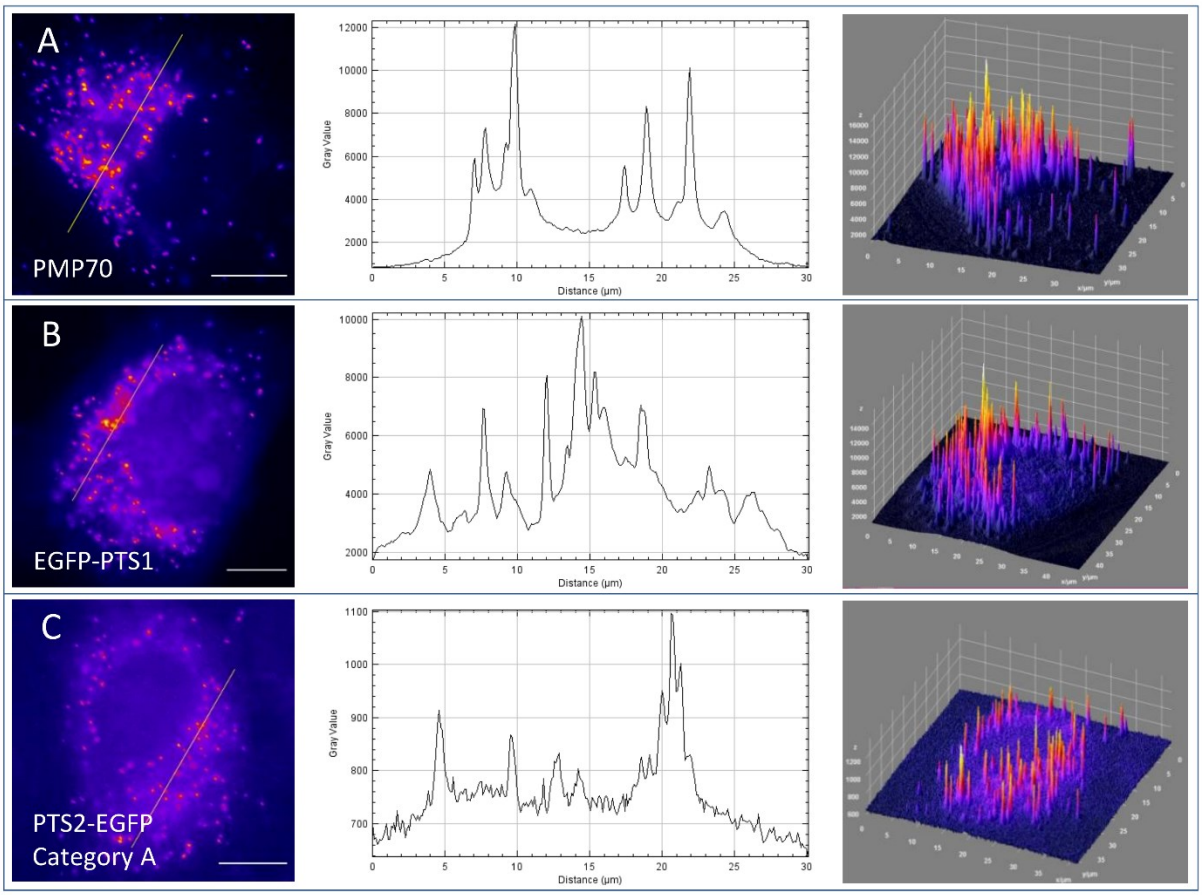


Figure 5. Conversion of signal intensities in one-, two- and three-dimensional depictions. Each row represents three different depictions of the same image obtained for different kind of peroxisomal proteins. In each row, there is a 2 dimensional image of a cell coloured with a so called heat-map lookup table (left), a 1 dimensional graph of a line (middle) drawn into the 2 dimensional image (line in the left picture indicates the position of the line) and a 3 dimensional plot of intensities for the whole 2 dimensional image (right) with false colour. In the upper row there are images of a cell with an antibody label of α-PMP70 and Cy-3 secondary antibodies (A), the middle row represents a cell transfected with a plasmid expressing the fusion protein EGFP-PTS1 (B) and the lower row depicts a cell transfected with plasmids expressing fusion protein PTS2-EGFP with peroxisomal matrix import category A (C). Scale bars shown in the images show a length of 10 µm. Gamma filter was used on all images. The cell type is COS7.

5.1.4 Two methods of image analysis

The central hypothesis in this study is that the import efficiency of peroxisomal matrix proteins can be deduced from analyzing fluorescence microscopy images.

To use the intensity information within the images, we tested two methods of image analysis. One, in which the core tool is the segmentation of the cell into peroxisomal and non-peroxisomal areas, and the other method, in which we use statistical analysis of all pixels of a cell without segmentation.

The first method, automized segmentation, is used to compare the intensity levels of peroxisomes to the intensity levels of the remaining cell as a proxy for the concentration of proteins within compartments.

In this study, we chose the free software Trainable Weka Segmentation (TWS), promoted and implemented by the free image management software ImageJ. It is based on the Weka open source machine learning platform. This software uses supervised input with the argumentation that humans use much more knowledge than what is used by most segmentation methods, which are based on colour and/or intensity information of individual pixels alone. But, putting all human knowledge into image segmentation algorithms would require huge databases which are currently not available. Supervised segmentation methods, however, are the first step into this direction. When compared to more classical segmentation methods, the advantages of the TWS are its versatile settings, the easy to understand graphical user interface and most importantly, the ability to improve the quality of the segmentation results by training. One of the disadvantages is that mistakes during the training process can decrease the quality of the resulting segmentation procedure. As a comparison of this technology, we used the segmentation software Squassh (implemented in ImageJ) and Cell Profiler, which are already in use for image analysis in cell biology.

The second method for image analysis utilizes the whole range of individual pixel intensities in a statistical manner without image segmentation. In this method, we read out the intensity of each pixel and ordered them from highest to lowest to obtain a plot which describes the intensity distribution within cells. Then, we stated the hypothesis that the first part of the plot includes mostly peroxisome pixels, the middle part of the plot encompasses the cytoplasm and nuclear intensities and the last part contains intensities of pixels at the border region of the cell. To compare different import categories to each other, we determined several points within the plot as key values and used them for calculations to obtain a meaningful separation between the categories.

5.2 Part one: Classical image analysis with new technology

5.2.1 Identification of peroxisomes with the Trainable Weka Segmentation

The TWS was our prime tool for identifying peroxisomes in this study. It combines a collection of machine learning algorithms with a set of selected image features to produce pixel-based segmentation. The underlying algorithm that was used was the FastRandomForest. As the name suggests, this is a fast implementation of the random forest algorithm, which uses a multitude of decision trees with the goal of creating a model that predicts the value of a target variable based on several input variables. To create predictive values, the TWS can be trained on images with the aid of image filters (Figure 6B). The predictive values, summarized in a so-called classifier, can then be used to segment different images.

The filters *Sobel* and *Hessian* are responsible for the detection of rapid changes in intensity, so called edges. The *Median* filter is especially useful against certain types of noise while at the same time increasing the highlighting of textures within the image. The *Gaussian blur* is the most common noise reducer used in image analysis. The *Difference of Gaussian* filter is highly effective at detecting peak intensity signals without detecting noisy background, because it subtracts different levels of *Gaussian blur* from one another. Each of these filters creates a new image. These images are then overlaid and a probability is calculated for each pixel to belong either to peroxisomes or cytoplasm.

The resulting probability image can then converted to a binary form (Figure 6C, Binarization). This means that regions in the image are separated in either peroxisomal objects, here in red with the intensity value of one, or non-peroxisomal areas, shown in green with the intensity value of zero (Figure 6C, Binarization). Next, watershed algorithm and a size exclusion algorithm were applied (Figure 6C, watershed). The watershed algorithm is a method based on shape rather than intensity of objects, which leads to a separation of objects connected by only a thin corridor consisting of a few pixels. The size exclusion algorithm precludes objects that are outside of the range of 4 to 300 connected pixels (Figure 6C). This ensured the exclusion of random intensity spikes caused by noise and unspecific large aggregates (Figure 6D).

Finally, peroxisomes were counted within a user defined cell area. After manual highlighting of cells and combining all steps mentioned above in an ImageJ macro language script, we were able to analyse Images automatically.

Next, we tested the segmentation algorithm on a set of 40 images of cells previously transfected with EGFP-PTS1 expressing plasmids (Figure 6E) and compared the automatic count of the new algorithm to a manual peroxisome count (Figure 6F). We found a strong correlation between the numbers of peroxisomes that were counted manually and automatically.

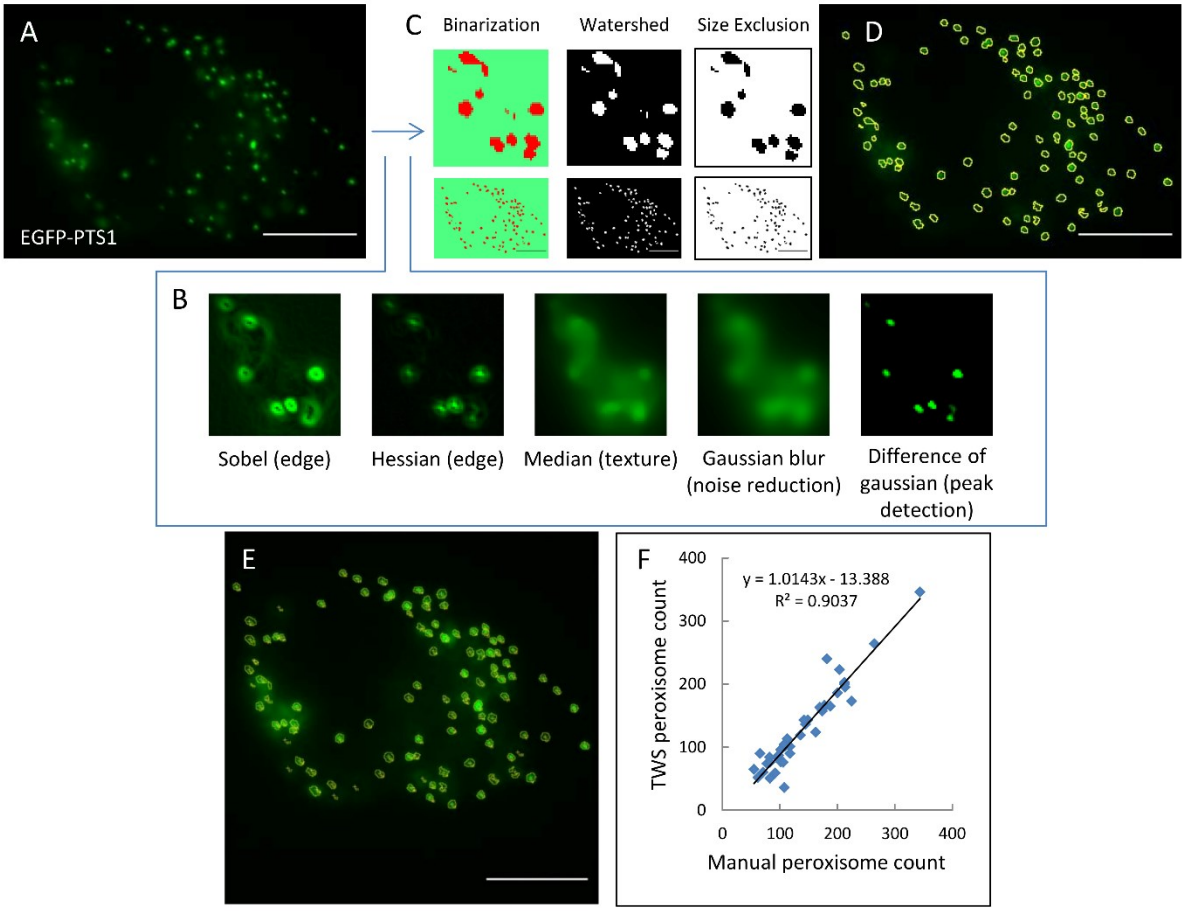


Figure 6. Function and testing of the Trainable Weka Segmentation. (A-D) Steps of segmentation: Original image of a cell transfected with EGFP-PTS1 expressing plasmids (A), applying a combination of user defined filters (B), creating a binary image (peroxisomes red, non-peroxisomes green) and utilizing watershed and size exclusion (C), reading out intensities of original image by overlaying regions of interest (D). Overlay of an automatic peroxisome segmentation and an independent manual segmentation (E) Regression analysis of peroxisome numbers obtained by automatic peroxisome counting versus a manual peroxisome counting, n=40 (F). Scale bars shown in the images show a length of 10 µm. Gamma filter was used on all images. The cell type is COS7.

5.2.2 Watershed algorithm and pixel-based size exclusion increase the robustness of the peroxisome count

To demonstrate the contribution of the watershed algorithm and the pixel-based size exclusion algorithm that followed the segmentation by the TWS, we re-ran the algorithm without using the watershed algorithm (Figure 7C and 7G) or without using the size exclusion (Figure 7D and 7H). Both methods should reduce the errors in the automatic peroxisome count, but in different directions. While the watershed algorithm compensates for a reduction in number caused by clumping of objects, the size exclusion prevents misidentified objects with a certain size to influence the peroxisome count. For this purpose, we selected a set of images (Figure 7A-D and E-H) and compared the results obtained by the final program with the results in the absence of either the watershed algorithm or the size exclusion.

By using watershed, the clump of peroxisomes in the image is separated into many smaller objects. The second set of images (Figure 7E-H) shows a different case. Here, objects are clearly separated, but the cell shows an increased intensity within non-peroxisomal areas which leads to an increase in random local intensity maxima. Without the exclusion of regions of interests based on pixel size, there would be an increase in objects detected that are below 4 pixels in size. In Figure 7 (I) the number of peroxisomes is depicted as bar graphs.

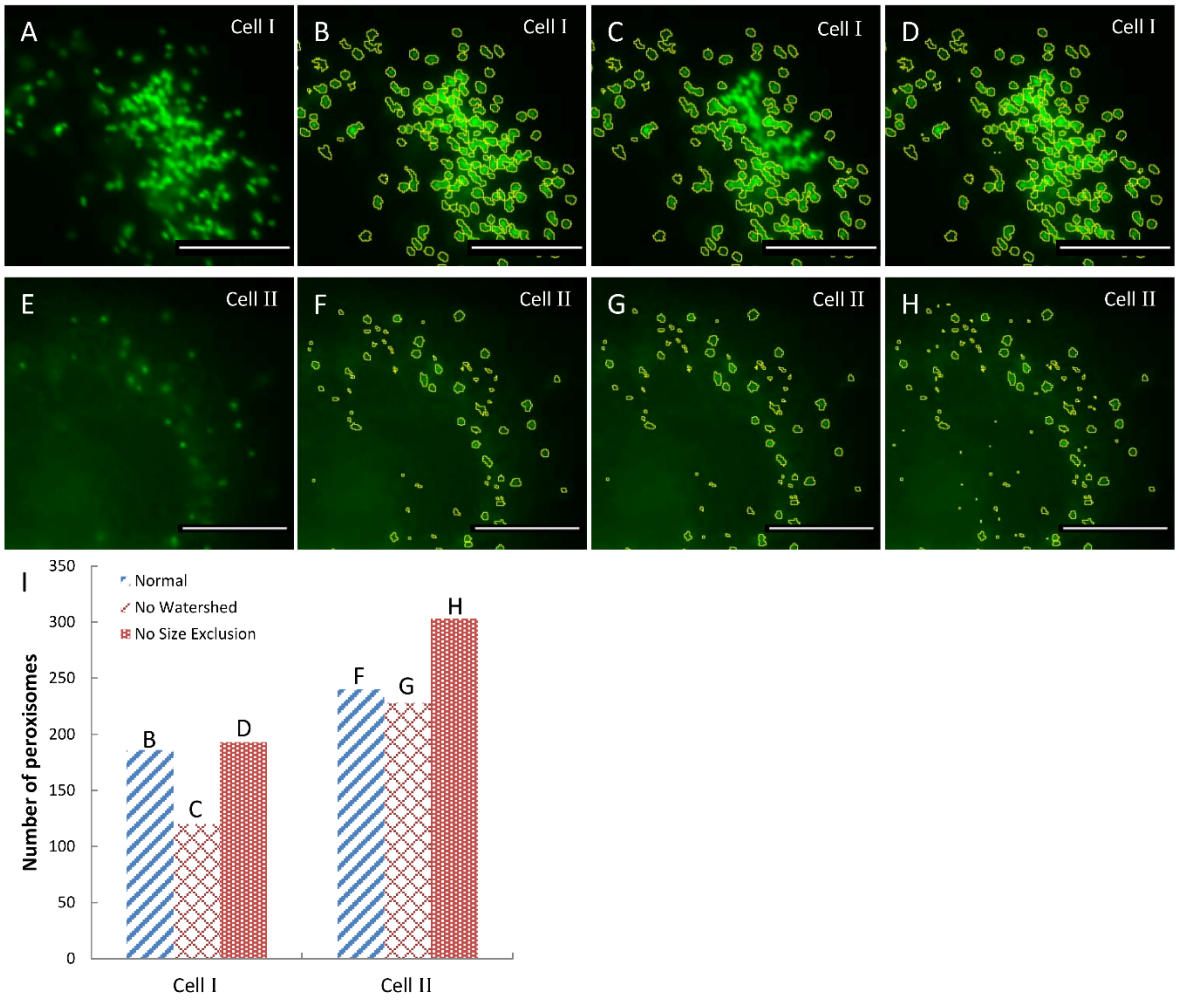


Figure 7. Importance of watershed algorithm and pixel-based size exclusion of regions of interest. The distribution of EGFP-PTS1 reporter proteins in COS7 cells are detected by the trainable WEKA segmentation. We used two cells as examples for the usefulness of the watershed algorithms and the size exclusion (A-D and E-H). From left to right: original false colored images (A, E), segmented images with both watershed and size exclusion (B, F), segmented images without watershed applied but with size exclusion (C, G) and segmented images with watershed applied but without size exclusion (D, H). Depicting the number of peroxisomes as bar graphs (I). The cell type is COS7. Gamma filters were use on all images.

5.2.3 Training of Trainable Weka Segmentation

The training of the TWS is an important asset for its function. We wanted to test whether the number of images used for training influences the resulting number of peroxisomes found and compare the results with a manual peroxisome count. The training of the TWS, described in the materials and methods section, was conducted with images of cells incubated with PMP70 antibodies, transfected with EGFP-PTS1, or transfected with PTS2-EGFP expression plasmids. After each training step of either 2, 3 or 5 images trained, the TWS classifier was applied on sets of images

sharing the same type of fluorescence staining (Figure 8). The number of peroxisomes found this way was then divided by the number of manually counted peroxisomes for each cell and the mean value and standard deviation was calculated (Figure 8). Results of a peroxisome count of images of cells incubated with PMP70 antibodies had a low standard deviation even when only three cells were used for training. The more samples were used for the training, the closer the ratio moved to the ideal value of one (dotted line). Classifiers trained with images of cells transfected with EGFP-PTS1 expression plasmids had larger standard deviations with few samples trained, but further training decreased the standard deviation and additionally brought the results of the automatic and the manual peroxisome count closer together. Training of classifiers on images of cells transfected with expression plasmids for PTS2-EGFP, assigned to import category A, showed an increase in performance when increasing the number of images trained but did not show the reduction in standard deviation as with the other training samples.

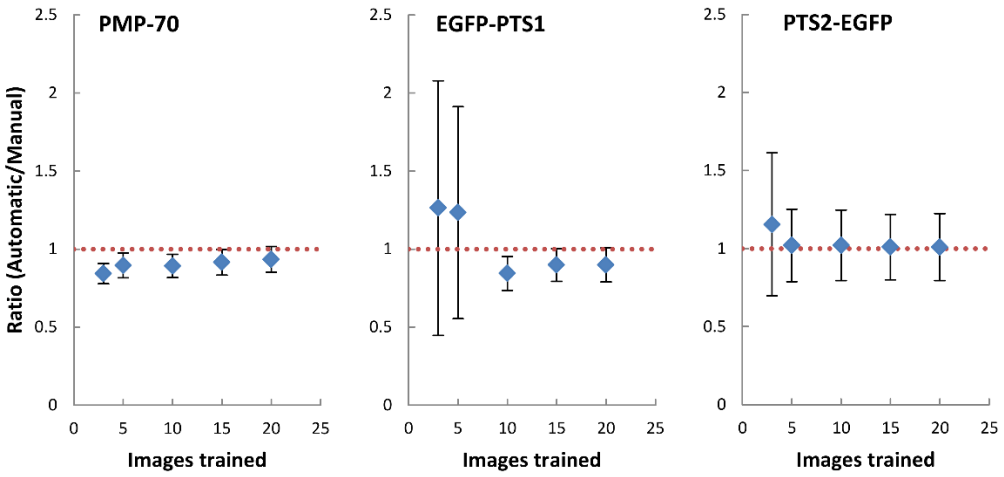


Figure 8. Training of Trainable Weka Segmentation. Per image, the TWS was trained by manually highlighting one peroxisomal and one non-peroxisomal area each. The y-axis shows a ratio of the number of peroxisomes found by the TWS and a manual peroxisome count. The x-axis shows the number of images used for training. The classifier that was trained was saved and then applied on images of the same class of fluorescence label to determine peroxisome numbers: PMP-70 n=30 (left), EGFP-PTS1 n=40 (middle) and PTS2-EGFP category A n=20 (right). Data shown as mean ± standard deviation.

5.2.4 Trainable Weka Segmentation classifiers trained on images of one type can be used on images of other types of fluorescence staining

Next, we wanted to test the performance of the TWS classifiers trained on one set, or on a combination of sets of images of fluorescence staining (PMP70, EGFP-PTS1, PTS2-EGFP) when applied to images of a different type of fluorescence staining. As a measure of quality, we used the correlation of manually and automatically counted number of peroxisomes within a cell. To train a classifier, different images of different fluorescence staining and varying intensities can be combined. The training was conducted as described in the materials and methods section.

In total, we trained three classifiers: (i) a PMP70 classifier trained on 10 images each of cells incubated with a 1:2000, a 1:4000 or a 1:10000 primary antibody dilution and Cy3 secondary antibodies, (ii) a combined classifier trained on 10 images of cells transfected with EGFP-PTS1 expression plasmid and 10 images of each peroxisomal matrix protein import category of cells transfected with PTS2-EGFP expression plasmid added to the information of the PMP70 classifier and (iii) a PTS2-EGFP classifier which was trained on 90 images of cells attributed to different peroxisomal matrix protein import categories.

The three classifiers were then applied to three sets of images: 30 images of cells incubated with a 1:2000 dilution of PMP70 primary antibodies and Cy3 secondary antibodies (Figure 9, top row), 40 images of cells transfected with EGFP-PTS1 expression plasmids (Figure 9 middle row) and 20 images of cells transfected with PTS2-EGFP expression plasmids (Figure 9, bottom row). Table 1 shows the slope and the coefficient of determination (R^2) of the resulting trendlines in Figure 9.

The PMP70 classifier works very well when applied on images of cells incubated with PMP70 as well as images of cells transfected with EGFP-PTS1 expression plasmids. Due to the large differences in intensity saved within the classifier and the test images, peroxisomes of PTS2-EGFP samples are mostly not recognized (Figure 9, PMP70 classifier row). The combined classifier has intermediate results when finding peroxisomes (Figure 9, combined classifier row). Last, the PTS2-EGFP classifier, specialized on low signal images, still works acceptably for PMP70 and EGFP-PTS1 samples, but excels at finding peroxisomes in PTS2-EGFP samples (Figure 9, PTS2-EGFP classifier row).

In conclusion, classifiers work best when applied to images of the same type but can also be applied to other images as well. A combined classifier is a compromise between sensitivity, specificity and field of use.

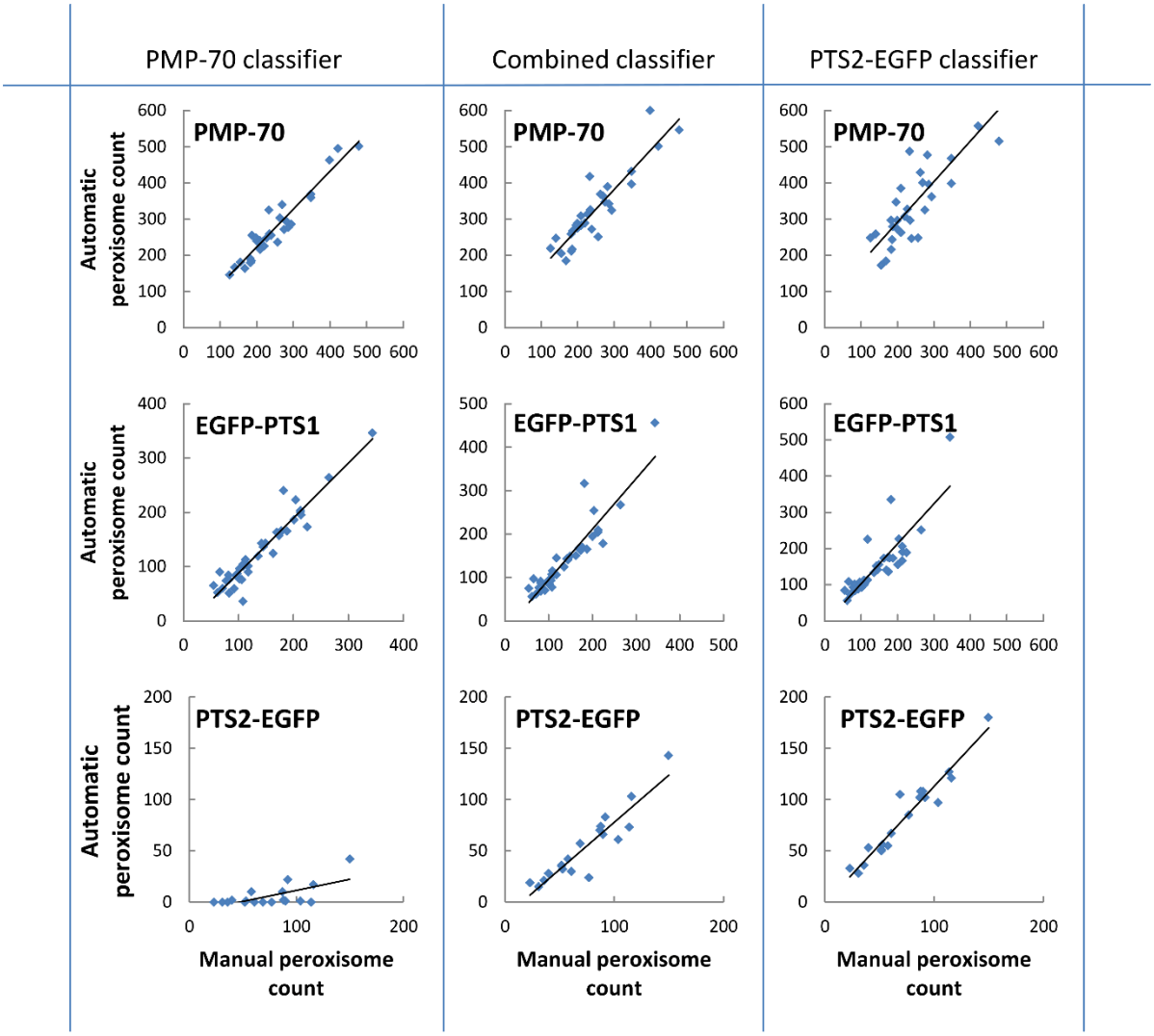


Figure 9. Quality control of TWS classifiers. The PMP70 classifier was trained on a total of 30 images of cells incubated with PMP70 primary and Cy3 secondary antibodies. The Combined classifier was trained on 30 PMP70 images, 10 images of cells transfected with EGFP-PTS1 expression plasmids and 20 images of cells transfected with PTS2-EGFP expression plasmids with import category A. The PTS2-EGFP classifier was trained on 90 cells transfected with PTS2-EGFP expression plasmids with varying import categories. Each of the three classifiers was applied to 30 images of cells labelled with PMP70 antibodies, 40 images of cells transfected with EGFP-PTS1 expression plasmids and 20 images of cells transfected with PTS2-EGFP expression plasmids to determine the peroxisome number. One of the PTS2-EGFP samples was excluded due to a statistically higher number of peroxisomes compared to the other cells.

Table 1. Slope and coefficient of determination (R2) of figure 4.

	PMP-70 classifier		Combined classifier		PTS2-EGFP classifier	
	Slope	R^2	Slope	R^2	Slope	R^2
PMP-70	1.049	0.908	1.087	0.824	1.121	0.676
EGFP-PTS1	1.014	0.904	1.162	0.844	1.106	0.721
PTS2-EGFP Category A	0.218	0.442	0.918	0.852	1.036	0.931

5.2.5 The Trainable Weka Segmentation shows better performance in detecting peroxisomes compared to the other two segmentation programs Squassh and Cell Profiler

After testing the effectivity of the TWS, we wanted to compare it to other image segmentation programs. As a measure of quality, we used the number of peroxisomes obtained by different software programs from the same images and compared the results to a manual peroxisome count.

The first program to compare to was Cell Profiler, a free, open-source software for quantitative analysis of biological images used in over 7000 publications. The software relies on buildable blocks of code, which can be easily interchanged with only few restrictions. The program was originally designed to count the number of cells in an image by using nuclear and cytoplasmic staining but has been expanded by many functions since.

The second program was a tool called Squassh (segmentation and quantification of subcellular shapes), which is a free tool implemented in ImageJ for 2-dimensional and 3-dimensional segmentation and quantification of subcellular shapes in fluorescence microscopy images. The tool is specialized on segmentation by region competition. Region Competition is a 2-dimensional and 3-dimensional multi-region image segmentation tool. It can segment arbitrarily (and not a priori known) numbers of objects in fluorescence microscopy images. The objects can have either constant or varying internal intensity. In this study, we only used the segmentation part of this software.

Images of cells incubated with PMP70 antibodies were the easiest to segment, because of the high signal to noise ratio and high intensity differences between peroxisomal and non-peroxisomal areas. Therefore, the use of the three programs

tested led to a similarly good correlation between the automatic and the manual peroxisome count (Figure 10A). For images of cells transfected with EGFP-PTS1 expression plasmids, the TWS shows the best correlation. CellProfiler and Squassh still have a good correlation regarding cells with lower amounts of peroxisomes but tend to underestimate the number of peroxisomes in cells with a larger number of smaller peroxisomes which are clustering.

When counting peroxisomes of images of cells transfected with PTS2-EGFP expression plasmids, the TWS clearly outperforms both Cell Profiler and Squassh in a comparison of the manual and an automatic peroxisome count. The reason for this is the low signal to noise ratio and low differences of peroxisomal and non-peroxisomal intensities of the PTS2-EGFP images.

This demonstrates the advantage of the TWS over the compared segmentation programs.

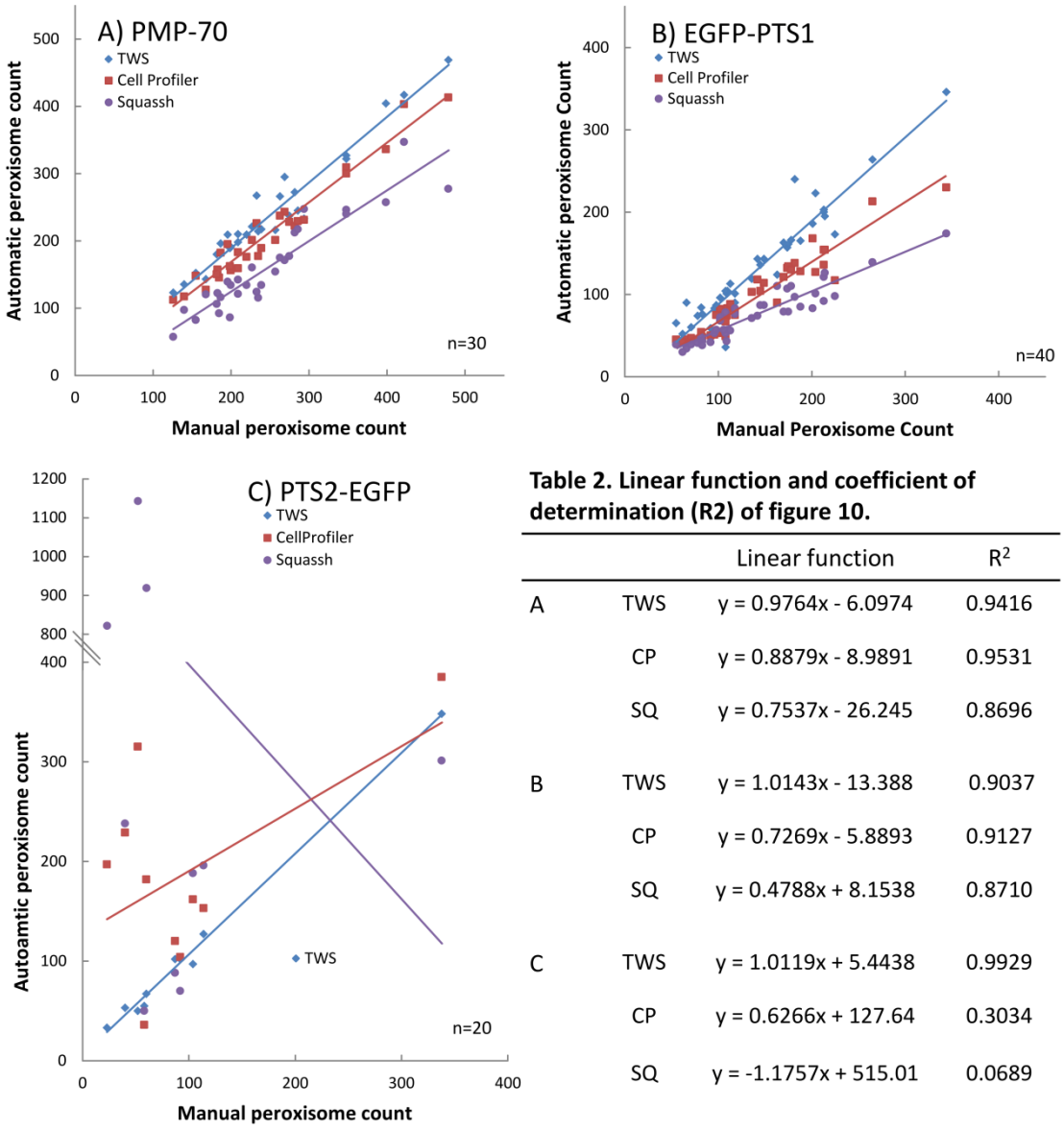


Table 2. Linear function and coefficient of determination (R2) of figure 10.

		Linear function	R^2
A	TWS	y = 0.9764x - 6.0974	0.9416
	CP	y = 0.8879x - 8.9891	0.9531
	SQ	y = 0.7537x - 26.245	0.8696
B	TWS	y = 1.0143x - 13.388	0.9037
	CP	y = 0.7269x - 5.8893	0.9127
	SQ	y = 0.4788x + 8.1538	0.8710
C	TWS	y = 1.0119x + 5.4438	0.9929
	CP	y = 0.6266x + 127.64	0.3034
	SQ	y = -1.1757x + 515.01	0.0689

Figure 10 Comparison of the novel program with other free-ware segmentation programs. The number of peroxisomes found by an automatic count after segmentation (y-axis) is compared to a manual peroxisome count (x-axis). The three segmentation programs that were compared are the TWS (a, blue diamonds), Cell Profiler (b, red squares) and Squassh (c, violet dots). The segmentation programs were tested on images with cells labelled with PMP70 antibodies primary and Cy3 labeled secondary antibodies (I), transfected with EGFP-PTS1 expression plasmids (II) or transfected with PTS2-EGFP expression plasmids (III). Of the PTS2-EGFP expression plasmid transfected cells, only category A was taken.

5.2.6 Comparing maximum and mean intensities of peroxisomes intensities of non-peroxisomal areas

After the positive quality control tests of the TWS, we wanted to focus on the analysis of images of cells transfected with plasmids expressing PTS2-EGFP. The aim of this analysis was to distinguish between different peroxisomal import categories. The images were segmented into peroxisomal and non-peroxisomal areas and the intensity information of those areas was extracted. The mean intensity of the peroxisomal maxima (Figure 11A) and the mean intensity of the peroxisomal areas (Figure 11B) was then compared to the mean intensity value of the non-peroxisomal areas within each cell, respectively.

When comparing the mean intensities of peroxisomal maxima and mean intensities of non-peroxisomal areas, significant differences in intensity can be found within cells of import categories A, B1 and B2 (Figure 11A). These are the categories where per definition peroxisomes are found. Images of category C cells show non-significant intensity difference, the TWS most likely only identified random intensity spikes within the category C images as “peroxisomes”.

When comparing the mean intensities of peroxisomal and mean intensities of non-peroxisomal areas within each cell, the results are similar to the comparison above, but the differences are not as large (Figure 11B). Here, only import category A and B1 show significant differences between the mean intensity of the peroxisomal and non-peroxisomal areas.

Next, we calculated the ratios between the mean of maxima (Figure 11A; Table 3, II) and mean intensities of peroxisomal areas (Figure 11B; Table 3, I) and the mean intensities of non-peroxisomal areas, respectively. In both cases (Figure 11C and 11D), peroxisomal matrix protein import category A depicts the highest value followed by the other import categories in descending order. The reason for the high standard deviation of the ratio of category A intensities stems from the division of high numbers (peroxisomal intensities) with low numbers (non-peroxisomal intensities). In category C cells, the intensity variability between cells can be very high, but the intensity difference within the cells is low. Therefore, the ratio of category C cell intensities has a low standard deviation.

Using the mean intensity of peroxisomal maxima leads to a greater difference in the intensity values than when using the mean value of peroxisomal intensities.

While a greater difference is desired, the use of the peroxisomal intensity maxima could lead to a biased outcome, especially when analysing import category C cells.

The summary is that there are significant differences between the intensities of peroxisomal and non-peroxisomal areas within cells showing import category A, B1 and B2 for the peroxisomal mean of maxima intensities and within cells showing import category A and B1 for peroxisomal mean intensities.

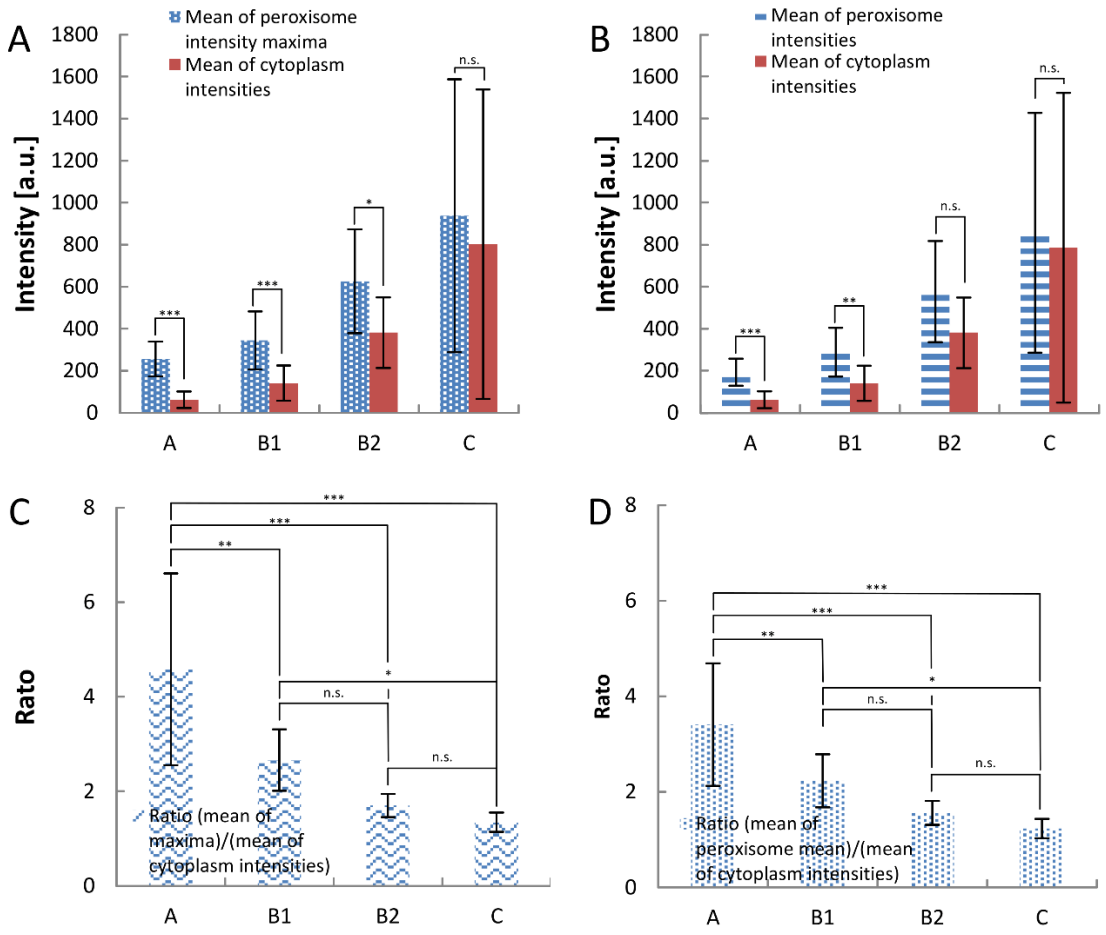


Figure 11. Peroxisomal and non-peroxisomal intensities measured after segmentation. The program used for segmentation is the TWS. The blue dotted bars (A) represent the mean of the maximum peroxisome intensities and the filled red bars represent the mean of non-peroxisomal areas within cells. The blue striped bars (B) represent the mean intensity of peroxisomal areas. The red bars in B are the same as in A. The wavy light blue bars (C) are calculated from the ratio of the mean of maxima intensity of peroxisomal areas and the mean intensities of non-peroxisomal areas within cells. The light blue bars (D) are formed by the ratio of the mean intensity of peroxisomal areas and the mean intensity of the cytoplasmic areas. The error bars are the standard deviation. The n=9 images of cells for peroxisomal matrix import category A and n=10 for categories B1, B2 and C. Statistical tests included t-tests for each category in figures A and B and ANOVA with subsequent post-hoc Tukey tests for figures C and D.

Table 3. Ratio of peroxisome and cytoplasm intensities

	I		II	
Category	Mean	StdDev	Mean	StdDev
PMP-70*	6.765	1.233	11.385	2.231
EGFP-PTS1*	3.178	0.853	4.439	1.272
A**	3.407	1.284	4.580	2.030
B1*	2.230	0.552	2.663	0.650
B2*	1.555	0.251	1.698	0.244
C*	1.231	0.204	1.343	0.205

I) Ratio of peroxisome mean intensities and mean cytoplasm mean intensities

II) Ratio of peroxisome mean maximum intensities and non-peroxisomal mean intensities

*N=10; **N=9

5.2.7 Mean and median intensities of peroxisomal intensities much lower than maximal intensities of peroxisomes

There are large differences between the mean and the mean of maximum intensities of peroxisomal areas within images. This is due to the sharp intensity peaks of peroxisomal areas within images.

A set of images of a cell in which peroxisomes were stained with PMP70 antibodies show a more detailed view of the distribution of intensity values of a peroxisome (Figure 12). From an image of a cell (Figure 12A), a peroxisomal cluster was taken (Figure 12B) and from there, a single peroxisomal area was cut (Figure 12C). We then added the intensity on the z-axis to create a 3-dimensional plot of the peroxisomal area. The top area with higher intensities (Figure 12D, yellow-orange) is small compared to the bottom area with lower intensities (Figure 12D, violet-blue). This leads to a skewed distribution of intensities and the numbers of central tendencies are lower than the maximum intensity value (Figure 12E). The intensity information of the images can be used to separate objects of interest from one another. This can either be done manually or with specialized segmentation programs.

The larger the peroxisomal area identified by the TWS, the larger the difference between the intensity maximum and the mean or median intensity of this area. To reduce the difference between mean and maximum intensity, the TWS could be trained to detect only sharp peaks and to not include the base of those

peaks. The downside of this procedure is the false positive recognition of random intensity spikes. A compromise has to be made between decreasing the risk of detecting random intensity spikes and reducing the peroxisomal area.

The conclusion is that using the mean over the median intensity of a peroxisomal area could be used to better represent the intensity peak. Using the intensity peak itself for image analysis could lead to biased results.

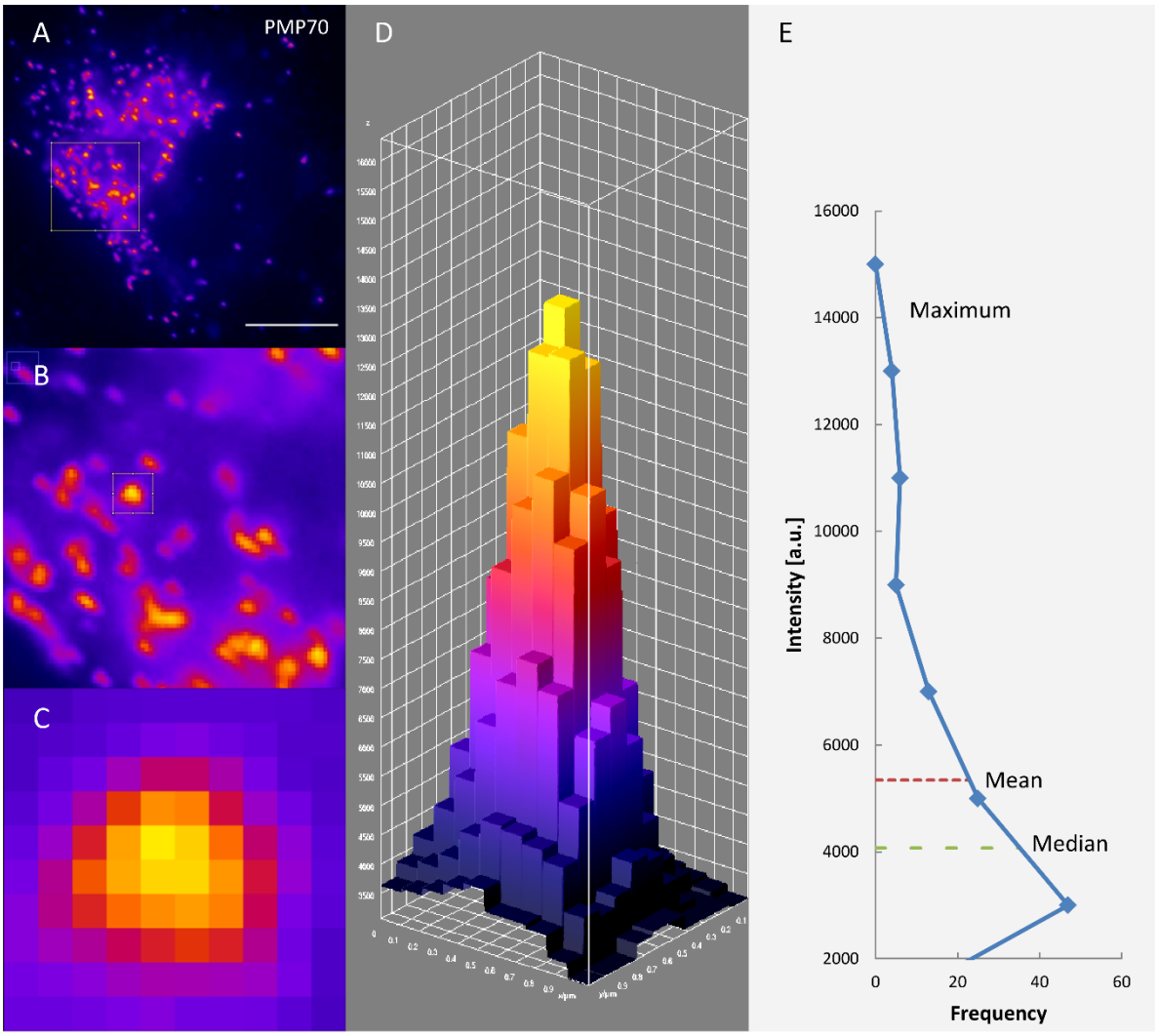


Figure 12. Peroxisome intensity peak as 2-dimensional image, 3-dimensional image and plot. A 2-dimensional image of a cell incubated with PMP70 antibodies (A) was used to closer investigate the intensities of a peroxisome area (B) and a single peroxisome (C). The intensities of the peroxisome are drawn as a 3-dimensional peak (D) and a plot (E).

5.3 Part two: Cell intensity analysis without image segmentation

5.3.1 Cell intensity analysis

A different method than calculating central tendencies from peroxisomal and cytoplasmic areas is to use the information encoded in the intensity distribution of each cell. With the second method presented in this thesis, there is no need for the computational expensive and complicated use of image segmentation. Additionally, there might not be a distortion of the intensity value due to differences in intensity peak and the base of the peroxisomal areas. The downside of this approach is that, without prior segmentation, a new system of sorting the intensity of each pixel is required.

5.3.2 Distribution of intensities and the localization of reporter proteins

This new system included ranking the intensity of each pixel from highest to lowest (Figure 13A, left section). Since in most cases peroxisomal intensities are higher than cytoplasmic ones, we hypothesised that the highest intensities primarily originate from peroxisomal areas. The intensity values of the cytosolic and nuclear areas of the cells were expected to present with varying intensities and placed in the middle section of the plot (Figure 13A, middle section). Finally, at the border of the cell the intensities are expected to be lower, because COS7 cells are hypothesised to be thinner on the cell border than in the centre (Figure 13A, right section).

Thus, we tried to conceptualize a curve that mimicks the expected distribution of pixel intensities. To calculate the baseline of the background noise level of the images, the mean of the intensities of 10-12 randomly chosen areas outside of cells was calculated and three times its standard deviation was added (Figure 13red dotted line). Every intensity value above this baseline was considered a singal, every intensity value below this baseline was considered noise.

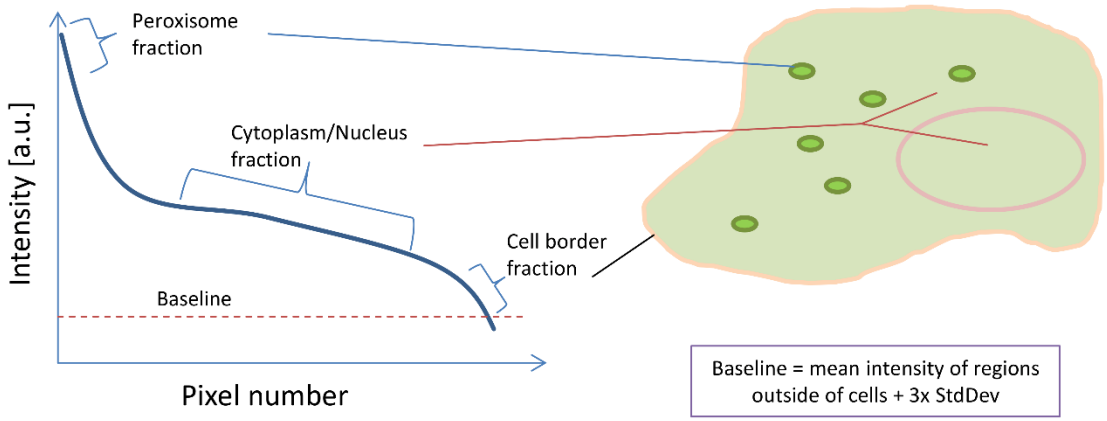


Figure 13. Hypothetical location of cell compartments within a pixel intensity curve. A cartoon of a cell transfected with peroxisomal matrix proteins shows different regions within a cell. These intensities can be plotted by ranking the intensity of each pixel in a descending manner. The baseline equals the mean value of the noise outside of cells plus three times its standard deviation.

5.3.3 Confirmation of the intensity distribution in cells expressing different reporter proteins and normalization of intensity plots

Since the maximal intensity, the form and the number of pixels can vary from cell to cell, we tried to find a way to quantify and normalize the plots.

Our approach for intensity normalization was to take the intensity value of each pixel and divide by the median of the first 100 pixels with the highest intensity value. To normalize the total number of pixels for each cell, we ranked each pixel within a cell, starting with the pixel with the highest intensity and assigned each pixel a number, starting with one. This means that the pixel with the highest intensity value was assigned the number one, which was then divided by the total number of pixels within the cell. The pixel with the second highest intensity value was assigned the number two and so on. Then, we used the pixel with highest intensity and the following pixels in 1% steps and calculated the median, the minimum, the maximum, the first quantile and the third quantile of the 101 values to create a median curve (Figure 14, B-f, C-f, D-f, E-f, F-f).

Images of cells incubated with antibodies against PMP70 generally have high peroxisomal and low cytosolic intensity (Figure 14A-a). The signal to noise ratio is the highest of the samples tested. The intensity signals start high in the intensity plot but decay quickly (Figure 14, A-b). There are little differences in intensity distribution when comparing different cells (Figure 14, A-d).

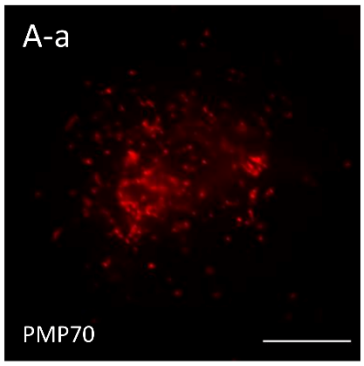
Next, we examined the intensity distribution in cells expressing the soluble reporter proteins EGFP-PTS1 (Figure 14, B-a). The first sample cell started with high,

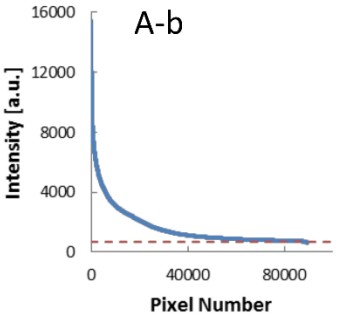
but few intensity values followed by a large number of lower intensity values (Figure 14, B-b). When testing more cells, it became clear that cell to cell intensity differences are larger than with the PMP70 samples due to differences in the number of plasmids that the cells receive during transfection (Figure 14, B-d). Therefore, we normalized the intensity and number of pixels as described above (Figure 14, B-e). The last step was to draw the median values of each of 101 pixels from the curves into one curve and add the first and third quartile and the minimum and maximum to that median line.

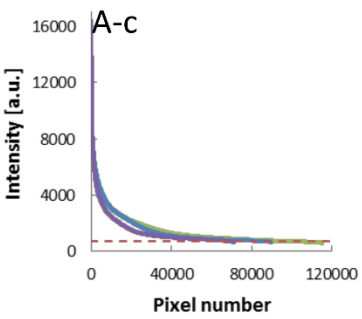
Finally, we repeated the process for cells transfected with plasmids expressing PTS2-EGFP reporter proteins (Figure 14C-F). Peroxisomal import efficiency category A cells have a low starting point of the intensities with the used exposure time of 300 ms (Figure 14C-a). The middle section of the plot is close to the noise baseline (Figure 14C-b). The maximal intensity is different for each cell observed, as expected with the use of soluble reporter proteins (Figure 14C-d). Normalization (Figure 14C-e) and calculation of a median curve (Figure 14C-f) produce distributions which have a relatively high intensity at the start and a lower intensity middle section.

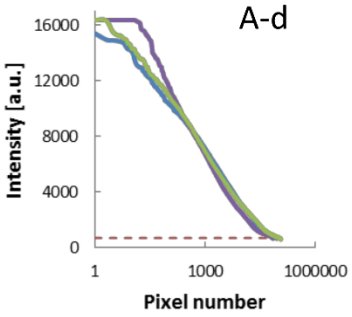
The other images of different matrix protein import distributions were treated the same. Category B1 showed a high starting point of intensities with a slightly elevated middle section (Figure 14D-f). The Category B2 plot starts with high intensities with a clearly elevated middle section (Figure 14E-f). Category C cells show a slope which is not as steep as with the other import categories (Figure 14F-f).

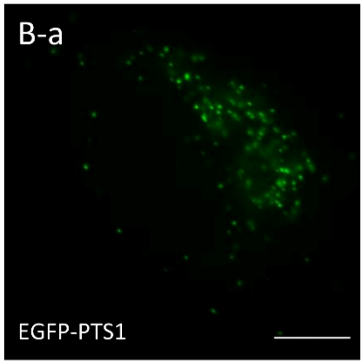
The conclusion of this section is that there are visible differences between the plots of the four categories of peroxisomal import efficiency after the quantification. In the next sections, a more refined approach for the specification of the peroxisomal and cytoplasmic areas within these plots and the calculation of numeric values to distinguish the categories will be explained.

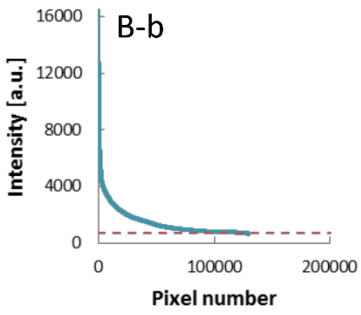

A-a
PMP70

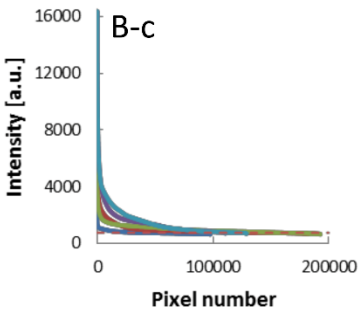

A-b
Intensity [a.u.]
16000
12000
8000
4000
0
0
40000
80000
Pixel Number

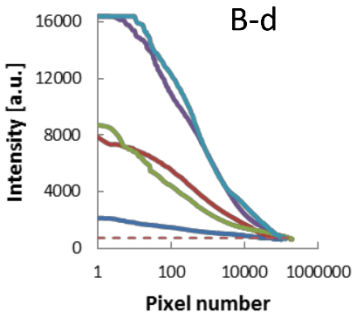

A-c
Intensity [a.u.]
16000
12000
8000
4000
0
0
40000
80000
120000
Pixel number

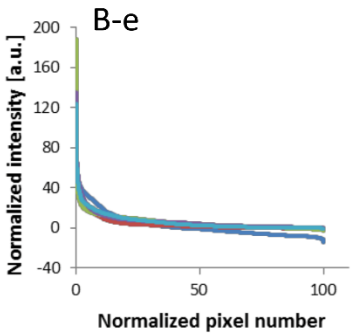

A-d
Intensity [a.u.]
16000
12000
8000
4000
0
1
1000
1000000
Pixel number

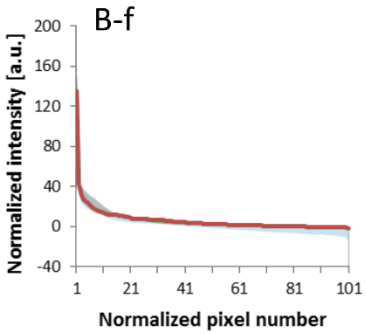

B-a
EGFP-PTS1

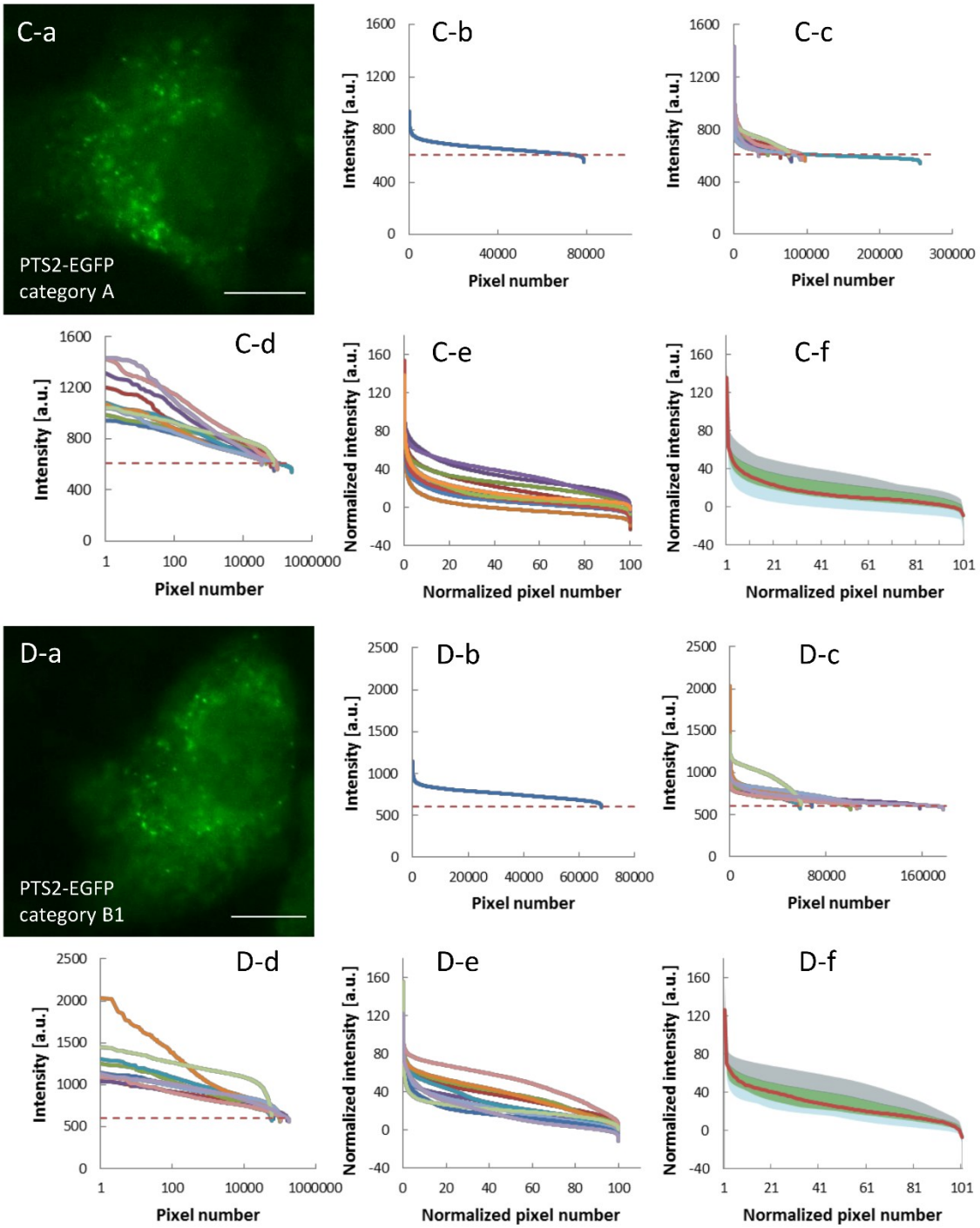

B-b
Intensity [a.u.]
16000
12000
8000
4000
0
0
100000
200000
Pixel number

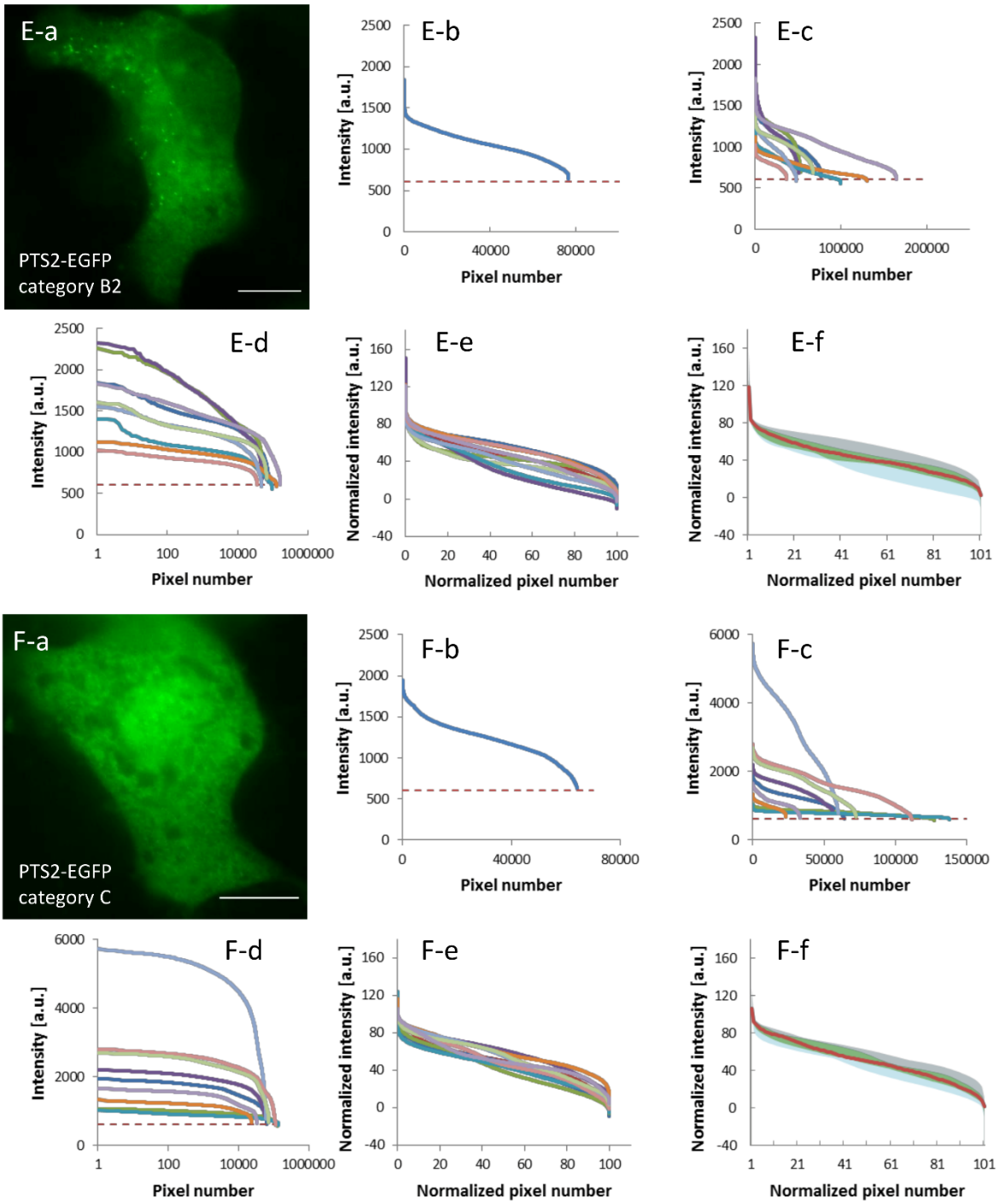

B-c
Intensity [a.u.]
16000
12000
8000
4000
0
0
100000
200000
Pixel number


B-d
Intensity [a.u.]
16000
12000
8000
4000
0
1
100
10000
1000000
Pixel number


B-e
Normalized intensity [a.u.]
200
160
120
80
40
0
-40
0
50
100
Normalized pixel number


B-f
Normalized intensity [a.u.]
200
160
120
80
40
0
-40
1
21
41
61
81
101
Normalized pixel number


C-a
PTS2-EGFP
category A
C-b
C-c
C-d
C-e
C-f
D-a
PTS2-EGFP
category B1
D-b
D-c
D-d
D-e
D-f
Intensity [a.u.]
Pixel number
Normalized intensity [a.u.]
Normalized pixel number


E-a
PTS2-EGFP
category B2
E-b
Intensity [a.u.]
Pixel number
E-c
Intensity [a.u.]
Pixel number
E-d
Intensity [a.u.]
Pixel number
E-e
Normalized intensity [a.u.]
Normalized pixel number
E-f
Normalized intensity [a.u.]
Normalized pixel number
F-a
PTS2-EGFP
category C
F-b
Intensity [a.u.]
Pixel number
F-c
Intensity [a.u.]
Pixel number
F-d
Intensity [a.u.]
Pixel number
F-e
Normalized intensity [a.u.]
Normalized pixel number
F-f
Normalized intensity [a.u.]
Normalized pixel number

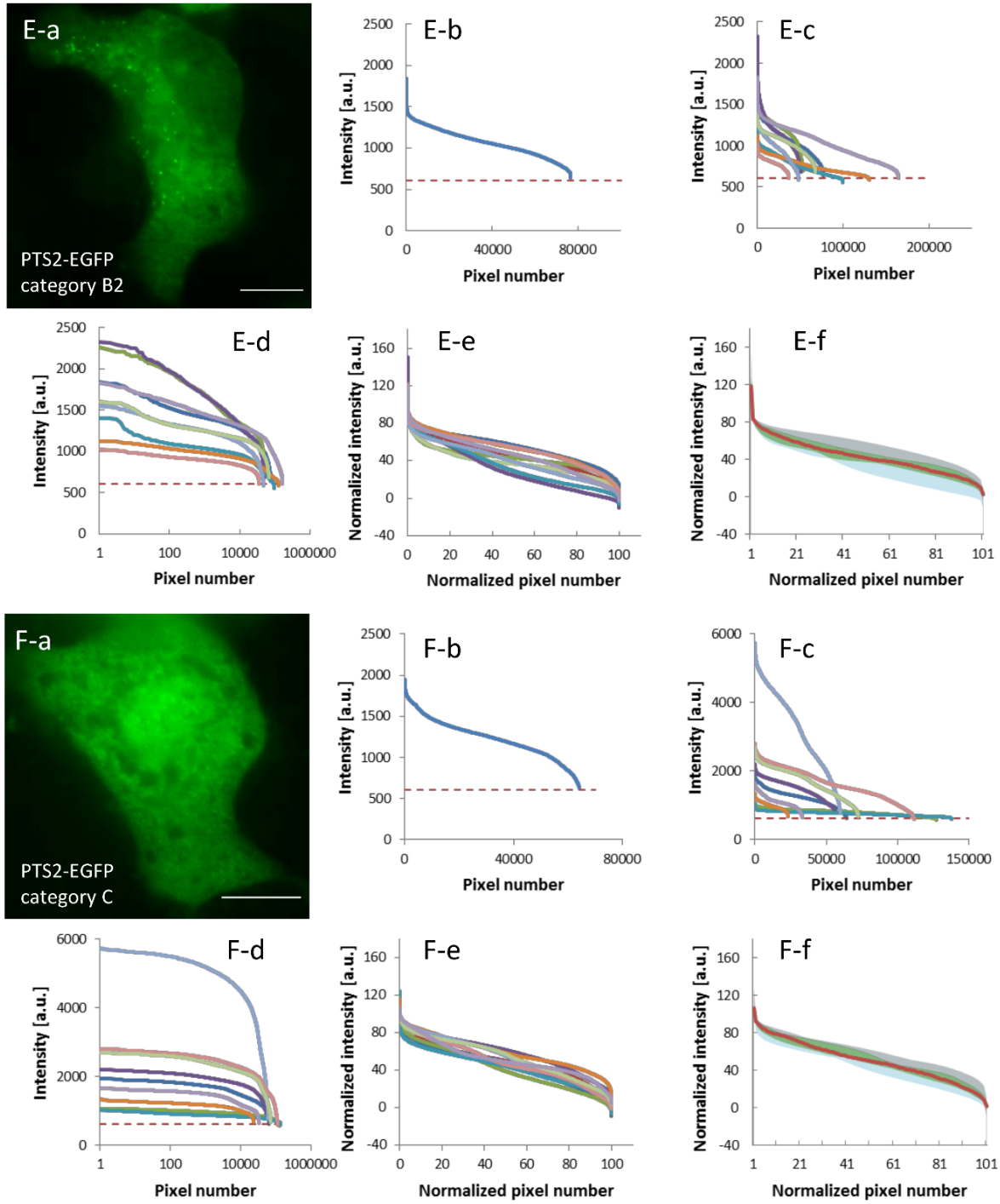


Figure 14. Intensity curves quantification and normalization. Intensity values of images of cells incubated with antibodies against PMP70 and Cy3 labelled secondary antibodies (n=3) (A), images of cells transfected with EGFP-PTS1 expression plasmids (n=5) (B) or images of cells transfected with PTS2-EGFP expression plasmids with all four matrix protein import categories (n=10 for each category) (C-F). Each section (capital letters) shows an example image (a), an intensity curve of one cell (b), intensity curves of multiple cells (c), intensity curves of multiple cells where the x-axis is $\log_{10}$ scaled (d), intensity curves of multiple cells with normalized values (e) and an intensity curve calculated from the median of the intensity curves in one section with quantile 1, 3, minimum and

maximum (f). The red dashed line represents the value of the noise in the image (intensity of outside regions mean+3xStdDev). Gamma filter was used on images of cells.

5.3.4 Hypothetical location and calculation of key values within intensity plots

5.3.4.1 Hypothetical locations of peroxisomal, cytoplasmic and border regions of the cell within the intensity plots

Following the quantification and normalization of the intensity plots, we defined the possible regions of interest within the plots (Figure 15). As stated before, the beginning of the curve is hypothesised to contain the peroxisomal pixels (either median of first 100 values, 2% median value or 10% median value), the middle part should contain the cytoplasmic pixels (value at 50% of the curve) and the last part of the curve should include the intensities of pixels close to or at the border region of the cell (value at 90 %).

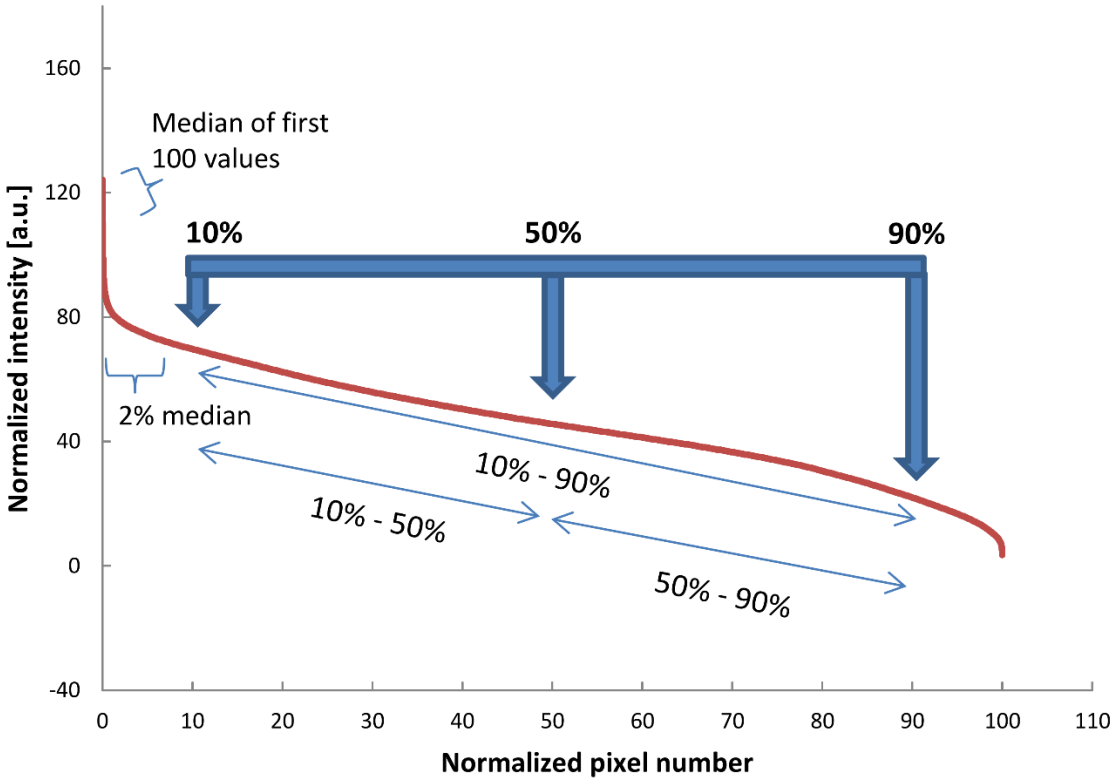


Figure 15. Hypothetical key values of an intensity curve. The intensity values of peroxisomal areas are hypothesized to be near the starting point of the plot (median of first 100 intensity values of the curve, 2% median value and 10% median value). Additionally, in the middle of the plot (intensity value at 50%) we suspect the intensities of the cytoplasmic areas and at the 90% mark we hypothesise the location of the intensities of the border regions of cells.

5.3.4.2 *Correlations of intensity plot key values and total intensity of images of cells transfected with plasmids expressing PTS2-EGFP*

To find out more about the locations within the intensity plots hypothesised as the representation of intensities of cell areas, we used a test sample of 40 images of cells showing different distribution of the expressed PTS2-EGFP. The 40 images consisted of 10 images of each peroxisomal protein import category and the intensity plots are shown above (Figure 16A-F). For this correlation test, we extracted the intensity information at the locations within the plot: (i) 2% median, (ii) 50% value and (iii) 90% value. Additionally, we extracted the amount of total intensity within each cell, because we hypothesised it to be a proxy for expressed reporter proteins.

To improve the visualization of the figures, we scaled the x-axes and y-axes with the $\log_{10}$. The individual peroxisomal import categories that were determined manually prior to this test are shown in different colour and shapes (Figure 16).

The positive correlation between the 2% median values of the plots and the total intensities of the cells (Figure 16A) suggests that cells which receive a low amount of plasmids (total intensity) expressing PTS2-EGFP are more likely to be sorted into import category A. The peroxisomal protein import categories cluster along the trend-line with category A having low starting intensity, but low total intensity when compared to category C cells with high starting intensity and high total intensity.

Plotting the 50% values against the total intensities within cells leads to a similar result (Figure 16B). The peroxisomal protein import categories cluster, but the separation between the import categories is not perfect (Figure 16A and 16B).

There is also positive correlation between the 2% median and the value at 50% of the plots (Figure 16C). The 2% median of category A intensity plots is much higher than the 50% values. This effect is reduced for category C cells.

The result is similar for the positive correlation between the intensity values at 50% and the values at 90% of the plots (Figure 16D). The slope in this correlation is not as steep suggesting that the difference between the 50% value and the 90% value of the intensity plots is not as extensive as in Figure 16C.

The conclusion of this section is that the results depicted in figure16 suggest that a low total intensity, a proxy for the expression level of reporter protein, leads to a good peroxisomal import (category A). Additionally, the difference between the 2% value (hypothesised peroxisomal intensity level) and the value at 50% (hypothesised

cytoplasmic intensity level) is larger in cells sorted into import category A than cells sorted into import category C.

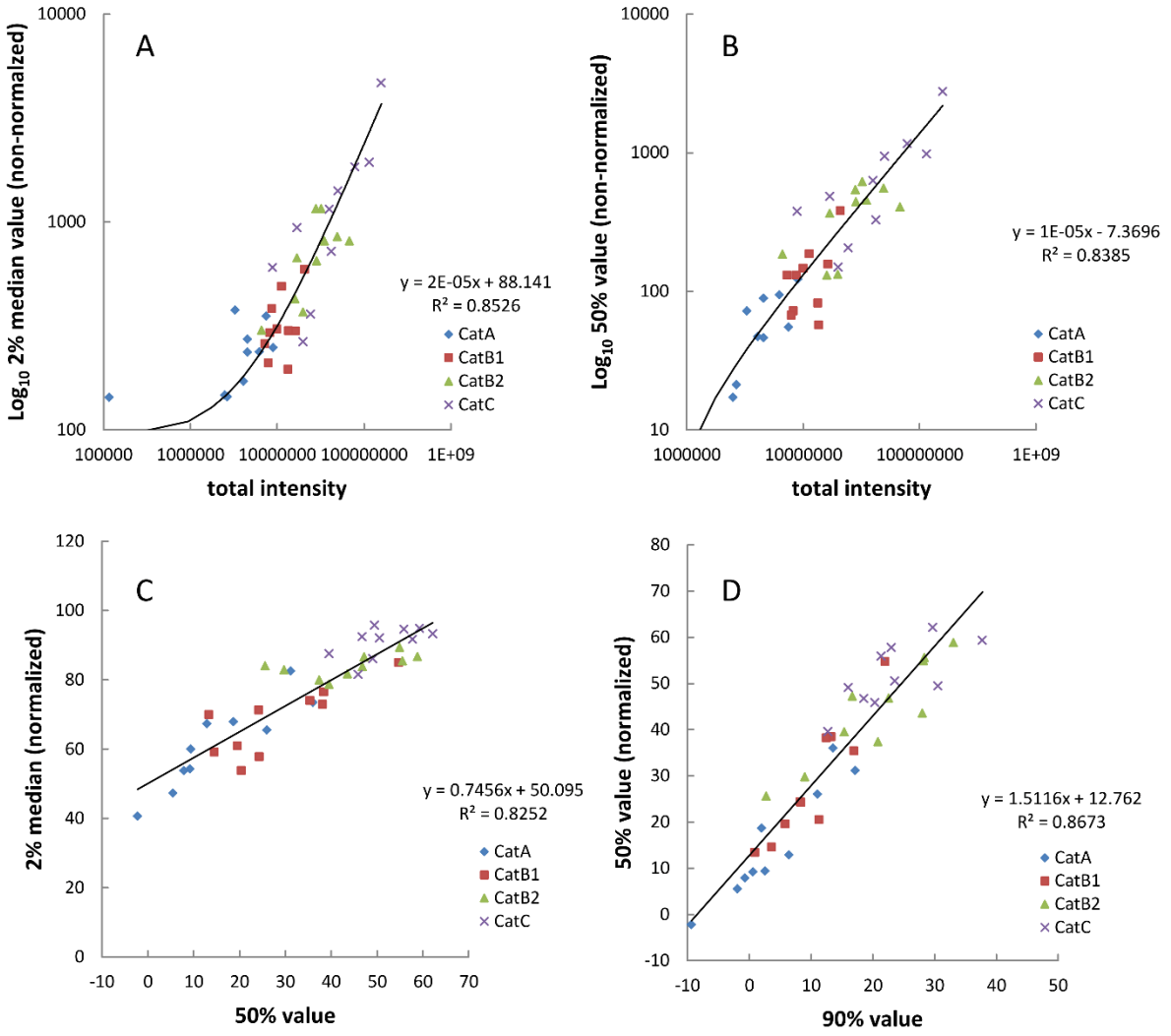


Figure 16. Normalized and non-normalized key values of the intensity curves of the PTS2 import categories. Using the microscope, cells transfected with PTS2-EGFP expression plasmids were categorized according to the import efficiency of the fluorescent peroxisomal matrix proteins. Then, 10 images of each cell with the respective import category were acquired and the intensity information was extracted using ImageJ. The $\log_{10}$ transformed key values 2% median (A) and 50% value (B) were plotted against the log transformed total intensity of a cell (A, B). Then, the normalized value of the 2% median was blotted against the normalized 50% value (C) and the normalized 50% value was blotted against the normalized 90% value (D).

5.3.4.3 Calculation of key values

Next, we wanted to separate the peroxisomal import categories of images of cells transfected with PTS2-EGPF expressing plasmids by using the value of a calculation which includes the peroxisomal, cytoplasmic and border region values of the intensity plots (Figure 17, Table 4). Additionally, we tested which of the hypothesised values for the peroxisomal fraction, either (i) the median of the 100 highest values, (ii) the 2% median or (iii) the 10% median, leads to the most separating power of the categories when implemented within a calculation (Figure 17, I, II and III).

The formula of the calculation consists of the subtraction of the peroxisomal intensity value by the cytoplasmic intensity value which is then divided by the subtraction of the peroxisomal intensity value by the intensity value of the border region of a cell.

$$Ratio = \frac{(peroxisomal\ intensity) - (cytoplasmic\ intensity)}{(peroxisomal\ intensity) - (border\ region\ intensity)}$$

We also used this calculation on intensity plots of images of cells incubated with antibodies against PMP70 or transfected with EGFP-PTS1 expressing plasmids as controls with a substantial difference between peroxisomal and cytoplasmic intensity, but low difference between cytoplasmic intensity and the intensity of the border regions.

The closer the result of the calculation to the value of one (100%), the higher the difference between the hypothesised peroxisomal intensities in comparison to the cytoplasmic intensities of the cell. This is shown by the value for the PMP70 samples being close to one with a very small standard deviation. EGFP-PTS1 samples also show a high mean value for peroxisomal intensities in all cases tested, but have higher standard deviation than the PMP70 samples. As shown before, samples that include soluble reporter proteins have a greater variability.

The bar graphs of the resulting ratio of the calculation form a “staircase” where category A has the highest value and category C the lowest. The separation between the categories is not sufficient to be significant in most cases (Figure 17, Table4).
These results confirm that the analysis of the pixel intensity distribution of cells is able to reflect differences in the apparent import efficiency suggested by visual classification.

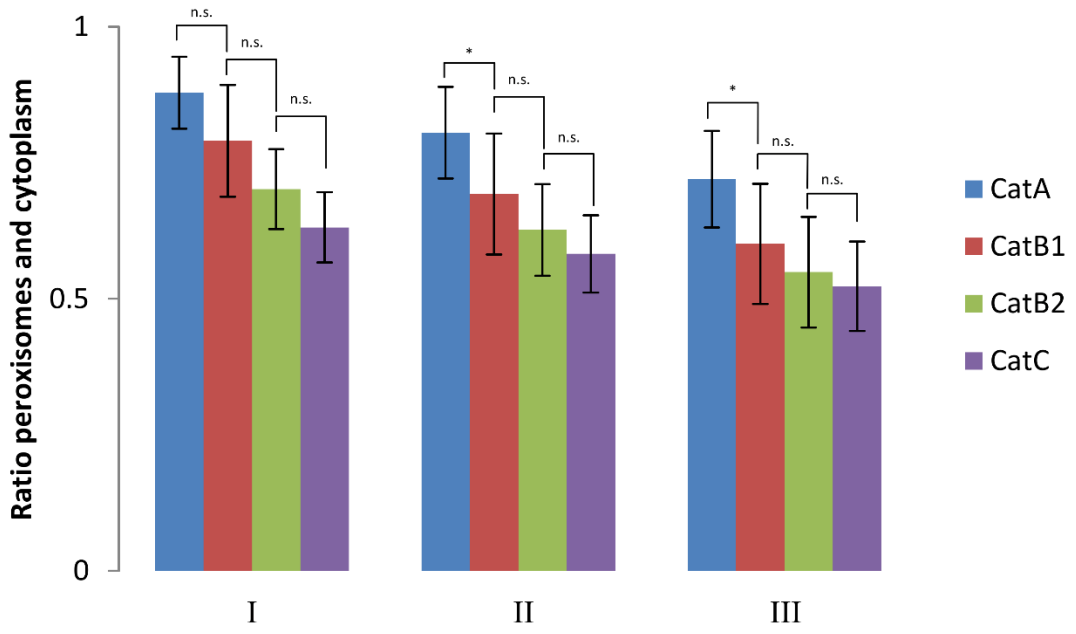


Figure 17. Bar graphs of peroxisomal intensities compared with cytoplasmic intensities for each of the four peroxisomal matrix protein import categories. The first set of bar graphs shows the comparison between the median of the first 100 values of the curve, the cytoplasmic value of 50% and the value for the border regions of the cell at 90% (I). The other two sets of bar graph depict the comparison of the intensity median value of the first 2% of the curve (II) and the median value of the first 10% of the curve (III). Shown are the mean and the standard deviation of n=10 images of cells.

Table 4. Key values for pixel intensity curves.

	I		II		III	
Category	Mean	StdDev	Mean	StdDev	Mean	StdDev
PMP-70^{+++}	0.976	0.003	0.956	0.005	0.917	0.010
EGFP-PTS1^{++}	0.964	0.016	0.910	0.036	0.847	0.052
A^{+}	0.879	0.066	0.805	0.084	0.719	0.089
B1^{+}	0.790	0.103	0.692	0.111	0.601	0.110
B2^{+}	0.701	0.073	0.626	0.084	0.549	0.102
C^{+}	0.631	0.065	0.582	0.071	0.523	0.082

I) (First 100 values - 50%)/(first 100 values - 90%)

II) (2% median-50% value)/(2% median-90% value)

III) (10% median - 50%)/(10% median - 90%)

$^{+}$N=10;$^{++}$N=5;$^{+++}$N=3

5.4 Part three: connection between classic image segmentation and intensity plot analysis

5.4.1 Combination of methods

Finally, we analysed, whether the results obtained by traditional methods are in agreement with results obtained by pixel analysis. The central question of this section is if our hypothesised regions within the intensity plots of part two of this study have similar intensities as the intensities of the areas segmented by the TWS in part one of this study.

As a first test, we marked the position of the value of the lowest peroxisomal maxima, found within an area segmented as peroxisomal by the TWS, within an intensity plot (Figure 18A and 18B). In this qualitative approach, we used an image of a cell incubated with EGFP-PTS1 expressing plasmids (Figure 18). Analysis of images of cells incubated with PMP70 or transfected with PTS2-EGFP showed similar results (data not shown).

Second, we calculated the ratio of peroxisomal pixels to cytoplasmic pixels as defined by the TWS (first part of this study) for different part of the pixel intensity curves to depict the distribution of peroxisomal pixel within this curve (Figure 19). In this test, we used images of cells incubated with PMP70 antibodies or incubated with EGFP-PTS1 or PTS2-EGFP (import category A only) expressing plasmids.

As a final control, we tested for correlations between the results of the methods presented in part one and two of this study (Figure 20). In this test, we used images of cells transfected with PTS2-EGFP expression plasmids including all four peroxisomal import categories.

5.4.2 Lowest peroxisomal intensity maximum

The first task for the last chapter was to find out more about the distribution of attributed to peroxisomal areas by the segmentation of the TWS. Therefore, we displayed the location of the lowest peroxisomal maximum. In this qualitative approach, we marked the maxima of each peroxisomal area (Figure 18A) and extracted the lowest peroxisomal intensity maximum of a cell transfected with EGFP-PTS1 (Figure 18B, red dotted line). The lowest peroxisomal intensity maximum is at the position 5850 of the curve of a total of 72318 total pixels (less than 10%) and shortly after the point of inflection.

Next, we wanted to examine the number of maxima within the first 100 pixels of the curve. Of the first 100 pixels of the curve, 14 out of 96 peroxisome maxima

were found. This means that within the first 100 pixel intensities there are only few of the highest peroxisome maxima and their surrounding intensities, which disqualifies the mean or median of the first 100 highest pixels of an intensity plot as good representation of an average of peroxisomal intensity values.

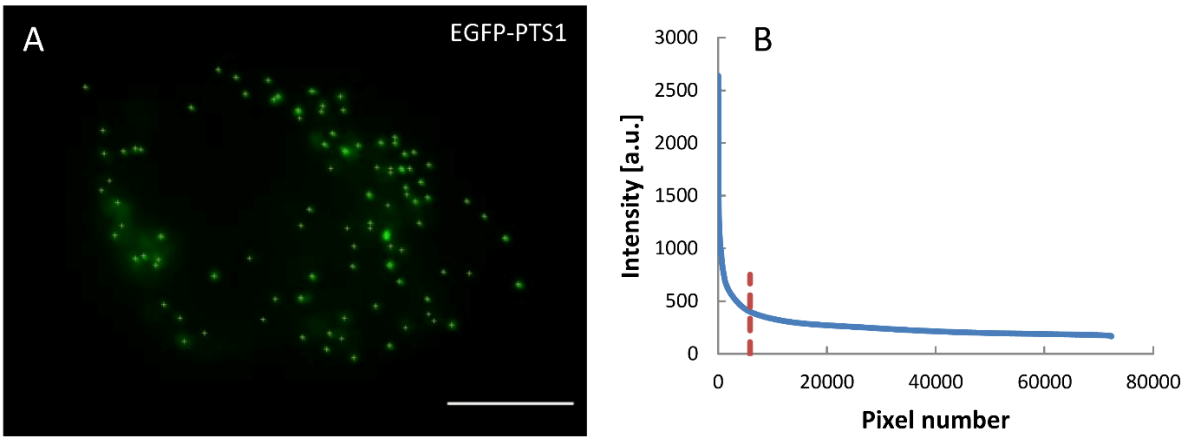


Figure 18. Position of lowest peroxisome intensity maximum within the intensity curve. The pixels belonging to a peroxisomal area of an image of a cell transfected with EGFP-PTS1 were determined using the TWS (A). Then, peroxisome maxima were determined within these peroxisome areas (A). The intensity information of each pixel of the cell area was extracted using ImageJ and plotted in a descending order and the position of the lowest peroxisome maxima was marked (B). Gamma filter was applied. Cell type is COS7. N=1.

5.4.3 Location of peroxisomal pixels within the intensity curve

Next, we used the intensity and place information of the pixels assigned to peroxisomal and non-peroxisomal areas by the TWS in the first part of this study to analyse the intensity plots of part two of this study. This was possible, because in part one and two of this study, the same images of cells were used for analysis. In a qualitative approach, we used images of cells incubated with PMP70 antibodies (Figure 19A and 19B) or images of cells transfected with either EGFP-PTS1 (Figure 19C and 19D) or PTS2-EGFP (Figure 19E and 19F) expressing plasmids.

The first approach was to rank all pixels from highest to lowest and assign each pixel either a peroxisomal or non-peroxisomal tag. Then, we formed a ratio of these tags after a certain percentage number of pixels (Figure 19, red squares).

Next, we calculated the median intensities at certain percentage amounts of pixels (Figure 19, green triangles).

Then, we included the information of how many peroxisomal pixels were counted at a certain percentage number of pixels adding up to 100% (Figure 19, blue diamonds).

The ratio of peroxisomal to cytoplasmic pixels was high at the start of the curve, but quickly fell off until around the 5% mark of total pixels and further went down from there in all three sample cells (Figure 19A, C, E, red squares). At the same time, total peroxisomal pixel number went up to 100% around the 30% mark of total pixels. From the same three images of cells, we extracted the information of the median falloff which was quite substantial from the start to the 30% mark of total pixels (Figure 19B, D, F).

The conclusion of this section is that the peroxisomal pixels found by the TWS are mixed with cytoplasmic pixels within the curve. This disqualifies the 10% median value for peroxisomes, because the ratio between peroxisomal and cytoplasmic pixels is 1:1 in the best case (cytoplasmic pixels are overrepresented) (Figure 19A, blue diamonds) or lower in the worst case (Figure 19C, blue diamonds). The compromise would be to use the 2% median to represent the peroxisomal intensities. The quick decay of the ratio is a limitation of this method, since peroxisomal and cytoplasmic pixels are mixed.

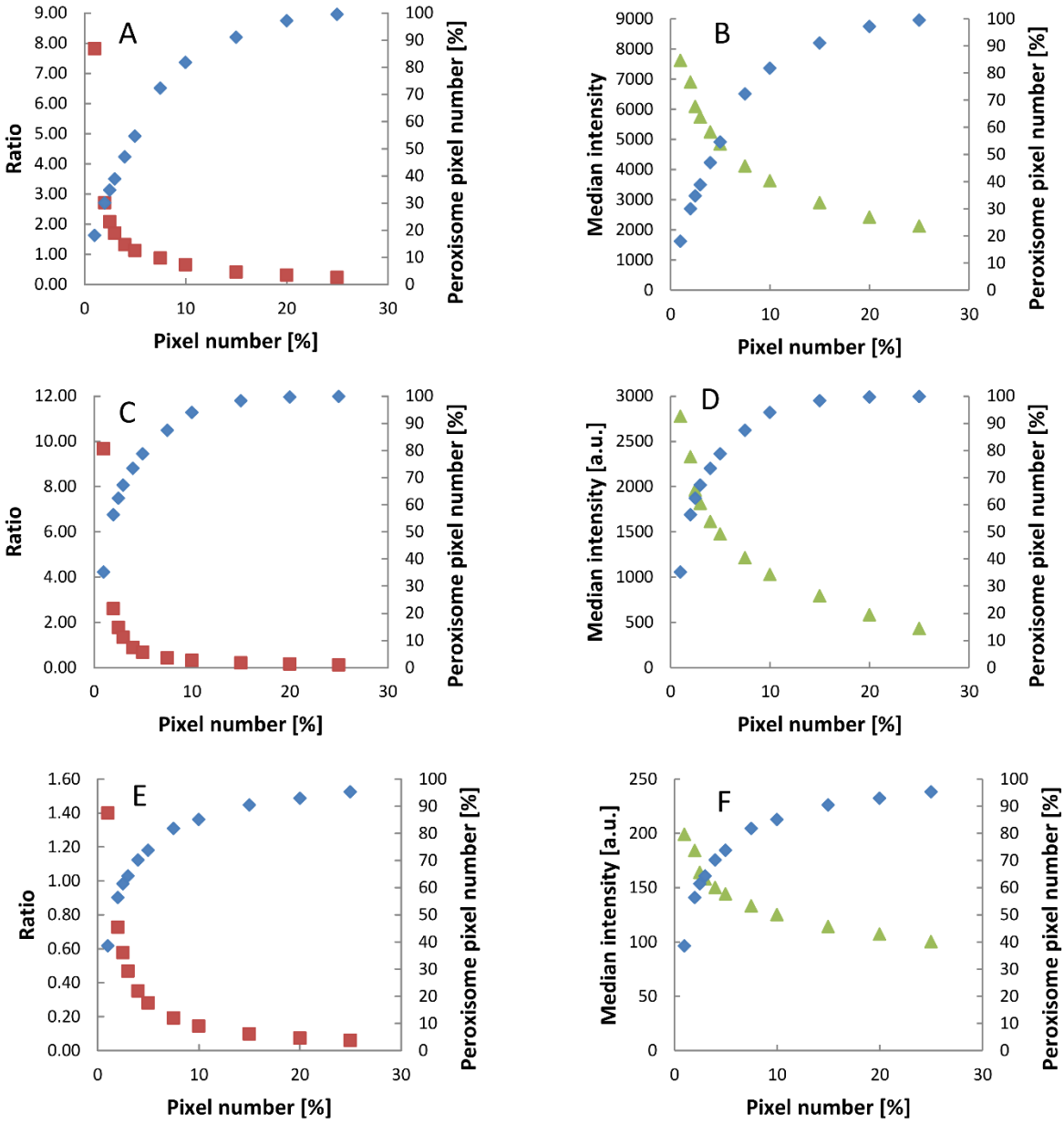


Figure 19. Analysis of positioning of peroxisome and cytoplasm pixel intensities within a descending intensity curve. The TWS was used to determine peroxisome and non-peroxisomal pixels to then extract the information of the positioning of both types within the intensity plots. The percentage of peroxisomal pixels is shown on the secondary y-axis on each blot (A, B, C, D, E, F; blue diamonds). The ratio of peroxisomal pixels and the cytoplasmic pixels was calculated at different positions within the curve (A, C, E; red squares). The median intensity was calculated at specific percentage locations within the pixel intensity curve (B, D, F; green triangles). Images of COS7 cells were used.

5.4.4 Final control: correlation between peroxisomal intensities found by the TWS and results of intensity calculation of the intensity plots

Lastly, we compared the results of the intensity measurements after the segmentation by the TWS to the results of the unbiased intensity distribution measurements (Figure 20). We used the intensities of images of cells transfected with PTS2-EGFP expressing plasmids. Only cells showing peroxisomal import category A, B1 and B2 were included, because the definition of category C is that peroxisomal areas are not clearly visible within the cell.

The average values of peroxisome maxima within each cell tested found within peroxisomal areas assigned by the TWS, show a good correlation to the 2% median values. This suggests that the average of the top 2% pixel is a reasonable measure for the peroxisomal intensity. Surprisingly, it seems as if the 2% median value correlates better with the peroxisomal intensity maxima than with the mean intensities of peroxisomal areas. Thus, the average of the top 2% pixel is a reasonable measure for the peroxisomal intensity.

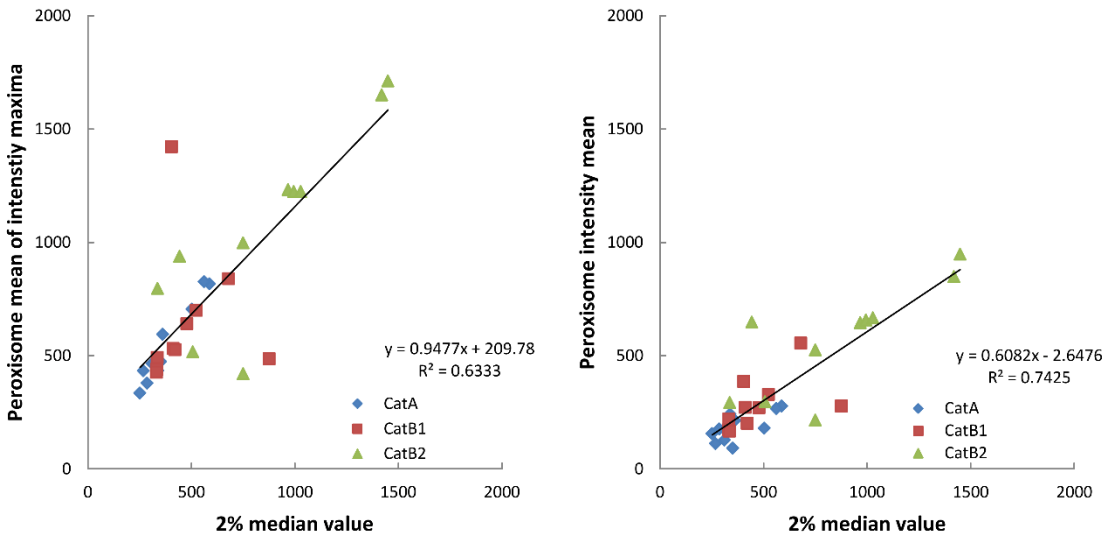


Figure 20. Correlation between peroxisome maxima or peroxisome median and 2% value of intensity plots. Using the microscope, cells transfected with plasmids expressing PTS2-EGFP were categorized concerning the import efficiency of the fluorescent peroxisomal matrix proteins. The information on the y-axis stems from the first section of this study and uses the mean calculation after segmentation by the TWS. The intensity information of the x-axis is drawn from the second section of this study, the unbiased analysis of the intensity distribution.

6 Discussion

The quality of peroxisomal import signals is hypothesised to be reflected by the import efficiency of peroxisomal matrix proteins. Therefore, we used a reporter protein expressed by a plasmid consisting of the genetic information of the PTS2 fused to EGFP to analyse its distribution within COS7 cells by using the epi-fluorescence microscope. Originally, each cell expressing such reporter protein was individually classified into one of four categories reflecting very good, good, weak and no import (A, B1, B2, C) by visual inspection. In this study, we want to present two computational methods as a first step to increase the efficiency of the classification and to reduce the bias that could be introduced by the experimenters. Additionally to the reporter protein mentioned above, we are using an EGFP-PTS1 expression plasmid and PMP70 primary and Cy3 labelled secondary antibodies as control reporters giving rise to different intensities.

The first of the two computational approaches use the traditional method of comparing the intensity of peroxisomal areas with the intensities of non-peroxisomal areas within an image of a cell. The free software Trainable Weka Segmentation (TWS), implemented in ImageJ, was our prime tool separating the peroxisomal and non-peroxisomal areas from each other, which is also called image segmentation. With the aid of the watershed and size exclusion algorithms, the TWS was able to segment the images well. With the right classifier, even PTS2-EGFP images with low signal to noise ratio were segmented reasonably well. As a quality control, we compared the number of peroxisomes found by the computational method to a manual peroxisome count. The result is a strong correlation between manual and automatic peroxisome count. The training of the TWS, essential for the use of this machine learning approach, showed that an increase in sample number trained with, increased the quality of the segmentation. Classifiers trained with images of cells with one type of fluorescence signal can be applied to images of a different type of fluorescence signal, if there are no major differences in intensities between the types of images. It is also possible to combine the training data sets of images of cells incubated with antibodies against PMP70 as well as a training data from images from cells expressing the soluble reporter protein EGFP-PTS1 When comparing the TWS to other image segmentation programs such as Squassh or CellProfiler, the other programs work well with samples that have a high signal to noise ratio like images of PMP70 and EGFP-PTS1 staining. For PTS2-EGFP images, however, both other

programs had troubles correctly finding peroxisomes due to low signal to noise ratio of the images. The final section of the first part (of this master thesis) compared the intensities of the peroxisomal areas to the intensity of non-peroxisomal areas. Next to the peroxisomal intensity mean, we also compared the mean of peroxisomal intensity maxima to the intensity mean of non-peroxisomal areas. The mean of maxima has higher intensity values, which results in larger differences in the values of peroxisomal and non-peroxisomal areas. The critique of using the mean of maxima in comparison to a cytoplasm mean is that independently which sample you choose, the maxima will always be higher than the lower cytoplasm even when observing only random variations within a cell. When analysing the results of the cells transfected with PTS2-EGFP expression plasmids and using mean of maxima, the result for the three categories A, B1 and B2 was significantly higher in comparison to the mean intensity of non-peroxisomal areas within cells. When using the mean of peroxisomal intensities, only cells of import category A and B1 were significantly higher than the mean of cytoplasmic intensities. The mean intensities and the mean of maximum intensity of the “peroxisomal” areas highlighted within category C cells were not significantly different from the mean cytoplasmic intensities. This was a good outcome, because without visible peroxisomes in category C cells, the segmentation programs only detected random intensity spikes. The conclusion is that the prediction of the peroxisomal import categories by the peroxisomal to non-peroxisomal intensity values is promising, but needs further improvement to lead to a satisfactory outcome.

The second computational method used a different approach to reach the goal of classifying images into the import categories. This approach uses a statistical analysis of cell images without the use of image segmentation algorithms. All pixels of a cell area were ranked according to their intensity. The hypothesis was that pixels of highest intensity reflect peroxisomes, whereas pixels with intermediate intensity correspond to the cytosolic and nuclear regions. Low intensity pixels were assumed to reflect the borders of the cell, which are in case of COS7 cells hypothesised to be flatter than the centre of the cell and therefore cannot contain as much reporter proteins. The baseline of the noise was calculated from regions outside of cells and subtracted from every intensity value. As a first result, a qualitative analysis showed that the course of the curve was different for each category describing peroxisomal matrix protein import efficiency (A, B1, B2, C). To quantify these results, we designed a system, in which we normalized the intensity values comprising the plots and

compensated for the size of the cells. The result was that there are clear differences between the intensity plots of images of cells of different import categories. To identify a better measure for the analysis of the plots, we advanced our hypothesis to specify regions of interest for the peroxisomal fraction. We tested the values of (i) the median of intensities of the 100 pixels with the highest intensities, which we hypothesised to represent the peroxisome maxima, (ii) the median value of the first 2% of intensities, which we thought to be the peroxisomal means and (iii) the median value of the first 10% of the intensities. The intensity value that was determined at the 50% mark of the plot was assumed to reflect the cytoplasmic intensity and the last part of the plot (value at 90%) was hypothesised to represent pixels close to or at the border region of the cell. Next, we tested the correlation strength between the values presented above to gain information about the relation to the peroxisomal import categories. Additionally, we correlated the median of the 2% highest intensity values and the intensity value at 50% of the plots with overall intensities within images of cells. The results of the correlation studies show strong positive correlation between the median of the 2% with highest pixel intensity and the total intensity on the one side and the 50% value and the total intensity of the cell on the other side. In both cases, the categories cluster along the linear trend-line, which shows that category A cells have generally lower total intensity than category C cells, which supports the hypothesis that a lower amount of reporter protein plays a critical role in producing the import categories. The positive correlation of the median of 2% of the highest pixel intensities with the value at 50% of the plots indicates that category A cells with relative low intensities within peroxisomal areas generally have lower intensities in the non-peroxisomal areas. With the slope of the linear trendline being less than 45 degrees, the differences between the two values are higher for import category A cells and much lower for import category C cells with B1 and B2 category cells being in between. This supports our hypothesis that import category A cells have a larger difference between the peroxisomal intensities and the cytoplasmic intensities than import category C cells. To find a formula for separating the peroxisomal import categories in this method, we created a calculation that includes the key values of the plots. The denominator represents the highest values hypothesised to be peroxisomal values subtracted by the value at 50% of the plot. This is divided by the numerator, which includes the subtraction of the intensities at the lower levels from

the peroxisomal intensities. This results in a value that should represent the form of the intensity plot.

$$Ratio = \frac{(peroxisomal\ intensity) - (cytoplasmic\ intensity)}{(peroxisomal\ intensity) - (border\ region\ intensity)}$$

To find out which value is the closest representation of peroxisomal intensities, we tested (i) the median of the 100 highest values, (ii) the 2% median and (iii) the 10% median. The cytoplasmic intensity was represented by the 50% value and the border region was reflected by the 90% value within the plots. The results of this calculation were that using the 2% median resulted in a significant difference when comparing the outcome of category A and category B1 cells. Additionally, category A included the highest values, because the images of cells showed high peroxisomal intensity, low cytoplasmic intensity and low border region intensity. Category C cells generally had a lower result, because the peroxisomal intensities as well as the cytoplasmic intensities were high and the border region intensities were intermediate. Category B1 and B2 were in between these two extremes. Overall, the values resulting from this calculation were a good representation of the form of the intensity plots.

We then used the combination of both methods to compare the results. In more detail, we used the place information of pixels assigned by the TWS to gain more information about the intensity plots of the second part of this study. First, we qualitatively tested the position of the intensity maxima of peroxisomal areas found in the first part of this study within an intensity plot created by the second method of this study using the same image of a cell. The conclusion was that within the 100 pixels with highest intensity, only few of the peroxisomal maxima were found together with their surrounding pixels. This was in contrast to our working hypothesis which stated that the 100 highest pixels contain most of the peroxisomal intensity maxima. Therefore, the 100 highest pixels should not be used as representative for the peroxisomal maxima. Second, we tested the positioning of the lowest peroxisomal intensity maxima within the intensity plot. The result was that the lowest maxima was found after the point of inflection which means that the maxima are also distributed in lower intensities than we expected. Next, we investigated the position of the peroxisomal pixels found in the first part of this study within the intensity plots of the second part of this study. From the start of the curve (left side, starting from the

highest pixel intensities), the ratio of peroxisomal pixels to cytoplasmic pixels decays quickly. In comparison to the total amount of pixels, the number of peroxisomal pixels is rather small. Therefore, cytoplasmic pixels quickly become dominant further down the curve. The higher the signal to noise ratio, the higher the differences between the starting ratio of peroxisomal to cytoplasmic pixels. With the sample image of a cell incubated with PMP70 antibodies, there are 10 times more peroxisomal than cytoplasmic pixels at the start of the curve. The curve of the image of a cell transfected with PTS2-EGFP shows only 1.6 times as much peroxisomal pixels than cytoplasmic pixels in the beginning. This suggests that in future studies the intensity of the signals has to be improved and noise has to be limited when using soluble reporter proteins containing a PTS2.

As a final control, we correlated the values of maxima and mean intensities of peroxisomal areas as indicated by the TWS and the 2% median values of the intensity plots. The result is that the 2% median values correlate better with the peroxisomal maxima suggesting that the findings of the two methods are in agreement.

Previously, other methods have been used to quantify import efficiency. First, a biochemical approach using an enzyme linked immunosorbent assay (ELISA)-based system. Reporter proteins are imported into a semi-permealized human cell line or fibroblasts (Terlecky *et al.*, 2001). The reporter protein is a fusion protein consisting of biotin, luciferase and PTS1. After a specific time period, avidin is added to the cell, which blocks biotin in the cytoplasm, but not within peroxisomes. The remaining free avidin is then bond by another compound and is removed. The amount fusion proteins enclosed in peroxisomes, is then measured by ELISA. The paper included different time periods for the in vitro approach with a maximum of 45 minutes. At that point it seems like a saturation of fusion proteins within peroxisomes is reached. In comparison with our method, fusion proteins located in the cytoplasm are completely removed. The in vitro approach, however, does not have to deal with uneven expression levels of plasmids, because the proteins are delivered to the cells directly. Whereas in our method, the cells are transiently transfected with plasmids expressing the reporter proteins, which leads to large differences in expression between single cells. The advantage of our method is that we include the intensities within the cytoplasm and not only the peroxisomal intensities. In this way, we can use more information about the import mechanism.

Second, I want to show a different approach to what is described above (Noguchi, Okumoto & Fujiki, 2013). The authors used a system of transfecting a fusion protein consisting of an unstable protein called FK 506 binding protein 12 variant (destabilization domain, DD), EGFP and PTS1. This protein is degraded by the proteasome in the cytoplasm unless there is a stabilizing factor Shield1. With this system, cells are grown in medium including Shield1 for some time. After the removal of Shiled1, non-imported fusion proteins are degraded in the cytosol, but remain stable in the peroxisomes. The peroxisomal EGFP fluorescence intensity is then measured by flow cytometry of the cell suspension.

Our method in comparison not only uses the intensity of the peroxisomes, but also the intensity of the cytoplasm to decide which peroxisomal import category should be used. Another advantage is that we can detect the number of peroxisomes as well as measure their intensity. In our system we already incorporate the hypothesis that each cell can receive a different amount of plasmids coding for fusion proteins during transfection, leading to highly heterogeneous phenotypes.

The conclusion of this master thesis is that the first method, with the aid of the TWS, reveals significant differences between the intensity values of peroxisomal and non-peroxisomal areas. This can be distinguished in cells with visible peroxisomes, but not in category C cells, where only random peroxisome intensity spikes were measured. To identify category C cells, the low difference between the intensity spikes and the cytoplasm could be used. If the intensity ratio within the cell is below a certain threshold, it would be categorized as category C.

The second method, based on statistics rather than segmentation, showed a difference between the different import categories when using the calculation showed above.

In future studies, the exposure time should be increased with low signal samples, because the increase of signal to noise ratio could increase the differences between the import categories. The next steps should include the testing of these methods with a larger sample size and the comparison with targeting signals that lead to lower peroxisomal protein import efficiencies.

7 Literature

Albertini M., Rehling P., Erdmann R., Girzalsky W., Kiel J.A., Veenhuis M. & Kunau W.H. (1997) Pex14p, a peroxisomal membrane protein binding both receptors of the two PTS-dependent import pathways. *Cell,* **89,** 83-92.

Arganda-Carreras I., Kaynig V., Rueden C., Eliceiri K.W., Schindelin J., Cardona A. & Sebastian Seung H. (2017) Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics,* **33,** 2424-2426.

Bankhead P. (2014) Analyzing fluorescence microscopy images with ImageJ. *ImageJ,* **1,** 195.

Braverman N., Steel G., Obie C., Moser A., Moser H., Gould S.J. & Valle D. (1997) Human PEX7 encodes the peroxisomal PTS2 receptor and is responsible for rhizomelic chondrodysplasia punctata. *Nat Genet,* **15,** 369-376.

Brocard C. & Hartig A. (2006) Peroxisome targeting signal 1: Is it really a simple tripeptide? *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research,* **1763,** 1565-1573.

Brocard C., Kragler F., Simon M.M., Schuster T. & Hartig A. (1994) The tetratricopeptide repeat-domain of the PAS10 protein of Saccharomyces cerevisiae is essential for binding the peroxisomal targeting signal-SKL. *Biochemical and biophysical research communications,* **204,** 1016-1022.

Cho D.-H., Kim Y.S., Jo D.S., Choe S.-K. & Jo E.-K. (2018) Pexophagy: Molecular Mechanisms and Implications for Health and Diseases. *Molecules and cells,* **41,** 55-64.

Chong C.S., Kunze M., Hochreiter B., Krenn M., Berger J. & Maurer-Stroh S. (2019) Rare Human Missense Variants can affect the Function of Disease-Relevant Proteins by Loss and Gain of Peroxisomal Targeting Motifs. *Int J Mol Sci,* **20**.

Dodt G., Braverman N., Wong C., Moser A., Moser H.W., Watkins P., Valle D. & Gould S.J. (1995) Mutations in the PTS1 receptor gene, PXR1, define complementation group 2 of the peroxisome biogenesis disorders. *Nat Genet,* **9,** 115-125.

Erdmann R. & Schliebs W. (2005) Peroxisomal matrix protein import: the transient pore model : Article : Nature Reviews Molecular Cell Biology. *Nat Rev Mol Cell Biol,* **6,** 738-742.

Fujiki Y., Okumoto K., Mukai S., Honsho M. & Tamura S. (2014) Peroxisome biogenesis in mammalian cells. In: *Frontiers in physiology*, p. 307.

Girzalsky W., Rehling P., Stein K., Kipper J., Blank L., Kunau W.H. & Erdmann R. (1999) Involvement of Pex13p in Pex14p localization and peroxisomal targeting signal 2-dependent protein import into peroxisomes. *Journal of Cell Biology,* **144,** 1151-1162.

Glover J.R., Andrews D.W., Subramani S. & Rachubinski R.A. (1994) Mutagenesis of the amino targeting signal of Saccharomyces cerevisiae 3-ketoacyl-CoA thiolase reveals conserved amino acids required for import into peroxisomes in vivo. *Journal of Biological Chemistry,* **269,** 7558-7563.

Gould S.J., Keller G.A. & Subramani S. (1987) Identification of a peroxisomal targeting signal at the carboxy terminus of firefly luciferase. *Journal of Cell Biology,* **105,** 2923-2931.

Honsho M., Yamashita S. & Fujiki Y. (2016) Peroxisome homeostasis: Mechanisms of division and selective degradation of peroxisomes in mammals. *Biochim Biophys Acta,* **1863,** 984-991.

Inestrosa N.C., Bronfman M. & Leighton F. (1979) Detection of peroxisomal fatty acyl-coenzyme A oxidase activity. *The Biochemical journal,* **182,** 779-788.

Jansen G.A. & Wanders R.J.A. (2006) Alpha-Oxidation. *Biochimica et Biophysica Acta - Molecular Cell Research,* **1763,** 1403-1412.

Kang W., Yang Q. & Liang R. The Comparative Research on Image Segmentation Algorithms. In: *2009 First International Workshop on Education Technology and Computer Science*, pp. 703-707. 7-8 March 2009 2009.

Kim P.K., Hailey D.W., Mullen R.T. & Lippincott-Schwartz J. (2008) Ubiquitin signals autophagic degradation of cytosolic proteins and peroxisomes. *Proc Natl Acad Sci U S A,* **105,** 20567-20574.

Kunze M., Malkani N., Maurer-Stroh S., Wiesinger C., Schmid J.A. & Berger J. (2015) Mechanistic insights into PTS2-mediated peroxisomal protein import: The co-receptor PEX5L drastically increases the interaction strength between the cargo protein and the receptor PEX7. *Journal of Biological Chemistry,* **290,** 4928-4940.

Kunze M., Neuberger G., Maurer-Stroh S., Ma J., Eck T., Braverman N., Schmid J.A., Eisenhaber F. & Berger J. (2011) Structural requirements for interaction of peroxisomal targeting signal 2 and its receptor PEX7. *Journal of Biological Chemistry,* **286,** 45048-45062.

Lametschwandtner G., Brocard C., Fransen M., Van Veldhoven P., Berger J. & Hartig A. (1998) The difference in recognition of terminal tripeptides as peroxisomal targeting signal 1 between yeast and human is due to different affinities of their receptor Pex5p to the cognate signal and to residues adjacent to it. *J Biol Chem,* **273,** 33635-33643.

Marzioch M., Erdmann R., Veenhuis M. & Kunau W.H. (1994) PAS7 encodes a novel yeast member of the WD-40 protein family essential for import of 3-oxoacyl-CoA thiolase, a PTS2-containing protein, into peroxisomes. *Embo j,* **13,** 4908-4918.

Miyazawa S., Osumi T., Hashimoto T., Ohno K., Miura S. & Fujiki Y. (1989) Peroxisome targeting signal of rat liver acyl-coenzyme A oxidase resides at the carboxy terminus. *Molecular and Cellular Biology,* **9,** 83-91.

Nichols J.A., Chan H.W.H. & Baker M.A. (2019) Machine learning: applications of artificial intelligence to imaging and diagnosis. *Biophysical reviews,* **11,** 111-118.

Noguchi M., Okumoto K. & Fujiki Y. (2013) System to quantify the import of peroxisomal matrix proteins by fluorescence intensity. *Genes to Cells,* **18,** 476-492.

Osumi T., Tsukamoto T. & Hata S. (1992) Signal peptide for peroxisomal targeting: replacement of an essential histidine residue by certain amino acids converts the amino-terminal presequence of peroxisomal 3-ketoacyl-CoA thiolase to a mitochondrial signal peptide. *Biochemical and biophysical research communications,* **186,** 811-818.

Otsu N. (1979) A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics,* **9,** 62-66.

Petriv O.I., Tang L., Titorenko V.I. & Rachubinski R.A. (2004) A new definition for the consensus sequence of the peroxisome targeting signal type 2. *Journal of molecular biology,* **341,** 119-134.

Poirier Y., Antonenkov V.D., Glumoff T. & Hiltunen J.K. (2006) Peroxisomal β-oxidation-A metabolic pathway with multiple functions. *Biochimica et Biophysica Acta - Molecular Cell Research,* **1763,** 1413-1426.

Rhodin J.a.G. (1954) *Correlation of ultrastructural organization : and function in normal and experimentally changed proximal convoluted tubule cells of the mouse kidney: an electron microscopic study*, Dept. of Anatomy, Karolinska Institutet, Stockholm.

Sagi Y., Basser P. & Assaf Y. (2009) Estimation of Cell Size Using the Composite Hindered and Restricted Model of Diffusion. *Magn Reson Med,* **17,** 1390.

Singh I., Lazo O., Kalipada P. & Singh A.K. (1992) Phytanic acid α-oxidation in human cultured skin fibroblasts. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease,* **1180,** 221-224.

Smith J.J. & Aitchison J.D. (2013) Peroxisomes take shape. *Nat Rev Mol Cell Biol,* **14,** 803-817.

Swinkels B.W., Gould S.J., Bodnar A.G., Rachubinski R.A. & Subramani S. (1991) A novel, cleavable peroxisomal targeting signal at the amino-terminus of the rat 3-ketoacyl-CoA thiolase. *Embo j,* **10,** 3255-3262.

Terlecky S.R., Legakis J.E., Hueni S.E. & Subramani S. (2001) Quantitative analysis of peroxisomal protein import in vitro. *Exp Cell Res,* **263,** 98-106.

Thomas J. Fellers K.M.V., Michael W. Davidson. (no date) CCD Signal-To-Noise Ratio.

Van Der Leij I., Fransen M., Elgersma Y., Distel B. & Tabak H.F. (1993) PAS10 is a tetratricopeptide-repeat protein that is essential for the import of most matrix proteins into peroxisomes of Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A,* **90,** 11782-11786.

Waterham H.R. & Ebberink M.S. (2012) Genetics and molecular basis of human peroxisome biogenesis disorders. *Biochimica et Biophysica Acta - Molecular Basis of Disease,* **1822,** 1430-1441.

Waterham H.R., Ferdinandusse S. & Wanders R.J.A. (2016) Human disorders of peroxisome metabolism and biogenesis. *Biochimica et Biophysica Acta - Molecular Cell Research,* **1863,** 922-933.

Watkins P.A., Howard A.E. & Mihalik S.J. (1994) Phytanic acid must be activated to phytanoyl-CoA prior to its α-oxidation in rat liver peroxisomes. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism,* **1214,** 288-294.

Webb D. & Brown C. (2013) Epi-Fluorescence Microscopy. *Methods in molecular biology (Clifton, N.J.),* **931,** 29-59.

8 Acknowledgements

I want to thank Univ.Prof. Dr. Johannes Berger for allowing me to be a part of his team and I want to thank Dr. Markus Kunze for his patience and guidance throughout my Master's Thesis. I also want to thank Dr. Fabian Dorninger and Ing. Martina Rothe for their support and I want to thank all the lab members for being helpful and friendly. I also want to thank my family and friends who were always there for me.

9 Appendix – ImageJ Macro language code used for automatic use of TWS

```
// @ File(label = "Input directory", style = "directory") dir1
// @ File(label = "Output directory", style = "directory") dir2
// @ File(label = "Specify TWS Classifier File (.model)", style = "file") classifier
// @ File(label = "Specify ROI File (.zip)", style = "file") cellRoi
// @short(label = "Particle Analyzer: minimum number of pixels", style = "spinner") MinPixel
// @short(label = "Particle Analyzer: maximum number of pixels", style = "spinner") MaxPixel
// These are examples script parameters, only work when they are on top - they are not active now

// This is important, because the script parameters are missing the "\" at the last position
// In imagej it is "/", because the "\", which would normally be in folder names seems to be
// taken by some other function
dir1 = dir1 + "/";
dir2 = dir2 + "/";
// Just to see how much time the script took afterwards
start = getTime();
// deletes all rois (regions of interest) currently in the manager
roiManager("reset");
// open the rois of the cell saved in the .zip file specified in the script parameters in the top,
// and load into the roi manager
roiManager("open", cellRoi);
// load the list of the names in the directory saved in "dir1" into the variable "list"
list = getFileList(dir1);
// count total number of rois and save the number in the variable called "NumberOfRois"
NumberOfRois = roiManager("count");
// Different output directories are created for the output images
// First, a variable is "loaded" with the path ("dir2") + the name of the folder
ClassifiedImageDir = dir2 + "/TWSClassifiedImages/";
BinaryImageDir = dir2 + "/BinaryImage/";
BareOutlines = dir2 + "/BareOutlines/";
RoiPath = dir2 + "/ParticleAnalyzerRois/";
// Here, the folders are created
File.makeDirectory(ClassifiedImageDir);
File.makeDirectory(BinaryImageDir);
File.makeDirectory(BareOutlines);
File.makeDirectory(RoiPath);
// The batchmode is used so processes run faster and also images that are opened during the process
// are closed or not shown on screen
setBatchMode(true);
// open one image after the other from the output path of the intensity adjusted images
// then open for each image the TWS, load the classifier and get the result
for (i=0; i<list.length; i++) { // with this, all images in the folder are processed
//for (i=0; i<5; i++) { // here, only x image is processed for testing the macro
   // showProgress is not so important - it just shows a progress bar for imageJ
   showProgress(i+1, list.length);
   // filename2 holds the path of the intensity adjusted output images from before
   filename = dir1 + list[i];
   // this introduces a condition: if the filename in the path specified with "filename2" ends with tif, open it
   if (endsWith(filename, "tif")) {
      // TWS does not work with batchmode activated, so batchMode false
      setBatchMode(false);
      // open the image that is saved in the varaible "filename"
      open(filename);
      // the file name without the extention is saved in the variable "NoExtensionName"
      NoExtensionName = File.nameWithoutExtension;
      // Open the Trainable Weka Segmentation
      run("Trainable Weka Segmentation");
      // Wait 3000 milliseconds for the TWS to load
      wait(3000);
      // load the classifier from the path specified in the variable "classifier"
      call("trainableSegmentation.Weka_Segmentation.loadClassifier", classifier);
      // tells the TWS to get the result (in form of a 8-bit colour image
      call("trainableSegmentation.Weka_Segmentation.getResult");
      setBatchMode(true); // you can switch this off for testing
      // save the Image in the path specified in "ClassifiedImageDir" with the name speicified in the
      // NoExtensionName variable + two zeros and the number saved in the variable i
      saveAs("tiff", ClassifiedImageDir + NoExtensionName + "TWSClassified" + "0" + "0" + i);
      // change the output of the TWS from 8-bit colour to 8-bit
      run("8-bit");
      // before running the make binary command, make sure that the background is set to black, this automatically sets the foreground to white
      setOption("BlackBackground", true);
      run("Make Binary");
      // watershedding is an important task that separates objects that are touching
      run("Watershed");
      // save the binary, watershedded image
      saveAs("tiff", BinaryImageDir + NoExtensionName + "Binary+Watershed" + "0" + "0" + i);
      // save the title of the Watershedded binarized image into the varaibel "BinaryWatershed"
      BinaryWatershed = getTitle();
      // reset the roi manager so we have a clean start
      roiManager("reset");
```

```
// load the original cell rois that we used in the beginning to define the cell outlines
roiManager("open", cellRoi);
// select the i'th entry in the roi manager
roiManager("Select", i);
// the Create Mask command leads to the creation of an image where the cell Roi is 255 (white) and the background is 0 (black)
run("Create Mask");
// save the title of the mask (usually the name is just "Mask"), so be careful to always close the mask if you dont need it
TitleMaskCell = getTitle();
// now select the window with the name saved in the varaible BinaryWatershed
selectWindow(BinaryWatershed);
// select the i'th entry in the roi manager - which is the cell roi of the specific cell
roiManager("Select", i);
// count the particles with the specified size in pixels, output bare outlines, exclude objects where at least one pixel
// is touching the border of the rois specified in cell roi, print summary and add all the dots to the manager
run("Analyze Particles...", "size=" + MinPixel + "-" + MaxPixel + " pixel show=[Masks] exclude summarize add");
// save the image name of the particle analyzer output into the variable "TitleMaskPeroxisomes"
TitleMaskPeroxisomes = getTitle();
// the next loop is necessary to be able to save the paroxisome rois without the cell rois
// so this loop selects each individual roi from the list starting with the first and ending with the number specified
// in the variable "NumberOfRois", specified in the beginning
for (j = 0; j < NumberOfRois; j++) {
   roiManager("select", 0);
   roiManager("delete");
}
// Now that we only have the rois of the individual peroxisomes without the cell outlines, we can count the number
RoiCount = roiManager("count");
// this is just for debugging, to see which number is in the variable "RoiCount"
print("The Roi Count is: "+RoiCount);
// I noticed crashes when the program detects zero or one peroxisome rois so i needed three possible outcomes
// 1) roi count is zero:
if (RoiCount == 0) {
   // print into the log: there were no peroxisomes found + iamge name
   print("There were no peroxisomes found in the image: " + NoExtensionName + "0" + "0" + i);
   // save the path of the image into the variable "filename"
   // the i stands for the i'th image of the list
   filename = dir1 + list[i];
   // open the image in the path filename
   // in this case: original input images
   open(filename);
   // save the title of the image into the variable "ImageTitle"
   ImageTitle = getTitle();
   // select image window with the name saved in the variable "TileMaskCell"
   selectWindow(TitleMaskCell);
   //run("Invert"); // this is not necessary, since there are 0 rois found
   // creating a selection means taking the mask of the cell and creating an overlay
   run("Create Selection");
   // the selection can then be added to the roi manager
   roiManager("Add");
   // select the raw image saved in the varaible "ImageTitle"
   selectWindow(ImageTitle);
   // select the first entry in in the roi manager
   roiManager("select", 0);
   // measure whatever is set in the option "set measurements" in imageJ
   run("Measure");
   // close all windows, i dont think this is really necessary
   run("Close All");
// 2) roi count is 1
} else if (RoiCount == 1) {
   // if there is a roi found with the TWS, save it
   roiManager("save", RoiPath + NoExtensionName + "0" + "0" + i + ".zip");
   // save the path of the image specified in dir1 into the variable "filename"
   filename = dir1 + list[i];
   // open the image from the path saved in the variable "filename"
   open(filename);
   // save the title of this image into the variable "ImageTitle"
   ImageTitle = getTitle();
   // show all command means that all rois in the roi manager are shown on the image
   roiManager("Show All");
   // measure whatever is set in the option "set measurements" in imageJ
   run("Measure");
   // this means delte all rois from the roi manager
   roiManager("reset");
   // this calculation subtracte the mask of the peroxisomes form the mask of the cell
   // the result is an image where the peroxisomes are "cut out" of the cell = cytoplasm + nucleus
   imageCalculator("Subtract create", ""+TitleMaskCell,""+TitleMaskPeroxisomes);
   // the resulting image needs to be inverted, otherwise it is an image with everything
   // highlighted except the cytoplasm + nucelus
   run("Invert");
   // now, a selection can be created form this image
   run("Create Selection");
   // selections can be saved into the roi manager
   roiManager("Add");
```

```
        // select the window with the name specified in "ImageTitle" (raw image)
        selectWindow(ImageTitle);
        // select the roi that was just added
        roiManager("select", 0);
        // measure whatever is set in the option "set measurements" in imageJ
        run("Measure");
        // close all, maybe not necessary when running in batch mode; this does not close result windows, only images
        run("Close All");
        // end of batch mode
        //setBatchMode(false);
      // 3) Roi count is greater than one
      } else if (RoiCount > 1) {
        // the deselect command for the roi manager means select all
        roiManager("deselect");
        // save the peroxisome rois of each cell into its own file (they are not very space-consuming
        roiManager("save", RoiPath + NoExtensionName + "0" + "0" + i + ".zip");
        // save the path of the image specified in dir1 into the variable "filename"
        filename = dir1 + list[i];
        // open the image from the path saved in the variable "filename"
        open(filename);
        // save the title of this image into the variable "ImageTitle"
        ImageTitle = getTitle();
        // this will show all rois on the iamge
        roiManager("Show All");
        // combines all rois into one big roi
        roiManager("combine");
        // since, the measure command can only measure 1 roi, we used combine
        run("Measure");
        // resets the roi manager/deletes all rois
        roiManager("reset");
        // this calculation subtracte the mask of the peroxisomes form the mask of the cell
        // the result is an image where the peroxisomes are "cut out" of the cell = cytoplasm + nucleus
        imageCalculator("Subtract create", ""+TitleMaskCell,""+TitleMaskPeroxisomes);
        // the resulting image needs to be inverted, otherwise it is an image with everything
        // highlighted except the cytoplasm + nucelus
        run("Invert");
        // now, a selection can be created form this image
        run("Create Selection");
        // selections can be saved into the roi manager
        roiManager("Add");
        // select the window with the name specified in "ImageTitle" (raw image)
        selectWindow(ImageTitle);
        // select the roi that was just added
        roiManager("select", 0);
        // measure whatever is set in the option "set measurements" in imageJ
        run("Measure");
        // close all, maybe not necessary when running in batch mode; this does not close result windows, only images
        run("Close All");
        // end of batch mode
        //setBatchMode(false);
      }
    }
}
// does what it says, selects the window named "Summary"
selectWindow("Summary");
// save the Summary
saveAs("results", dir2 + NoExtensionName + "Summary.csv");
// Summary of the Peroxisomal Intensities
selectWindow("Results");
saveAs("results", dir2 + "PeroxisomeIntensities.csv");
// print the overall time in the Log window
print("Toal amount of time: " + (getTime()-start)/1000);
```

10 Deutsche Zusammenfassung des Abstracts

Peroxisomen sind Organellen, besitzen nur eine Membran und sind verantwortlich für verschiedene Stoffwechselwege. Der Import von peroxisomalen Matrix Proteinen beginnt bei Proteinen welche entweder ein peroxisomal targeting signal 1 (PTS1, C-terminal) oder ein peroxisomal targeting signal 2 (PTS2, N-terminal) aufweisen. Das PTS1 bindet am Rezeptor PEX5; das PTS2 bindet sowohl am Rezeptor PEX7 als auch am co-Rezeptor PEX5. Diese Verbindungen können dann an der peroxisomalen Membran binden und den Import initiieren.

Die Effizienz des peroxisomalen Matrix Imports ist abhängig von der Qualität des peroxisomal targeting signals (PTS). Um die Qualität von Import Signalen zu bestimmen, wurde zunächst eine visuelle Inspektion der Import Effizienz in Zellen durchgeführt. Hierzu wurden fluoreszierende Reporter Proteine verwendet. In dieser Studie möchten wir diese manuelle Methode mithilfe von Fluoreszenzmikroskop und Bildanalyse automatisieren. Als erster Schritt wurden die Peroxisomen von einem von einem Segmentierungs-Algorithmus namens Trainable Weka Segmentation segmentiert und der Mittelwert der Intensitäts-Maxima der peroxisomalen Gebiete einer Zelle wurden mit dem Mittelwert der Intensitäten der nicht-peroxisomalen Gebiete verglichen. Die zweite Methode beschäftigt sich mit der Intensitätsverteilung innerhalb der Zellen ohne vorherige Segmentierung.

Die Ergebnisse dieser Segmentierung zeigen signifikante unterscheide zwischen den Mittelwerten der Intensitäten von peroxisomalen und nicht-peroxisomalen Gebieten innerhalb der Zelle. Die Ergebnisse der Pixel basierten Methode deuten darauf hin, dass die Form der Kurven die visuelle Evaluierung der Import Effizienz reflektiert. Zusätzlich beeinflusst die Anzahl an Reporter Proteinen die Evaluierung der Import Effizienz.

Beide Methoden, welche in dieser Studie vorgestellt wurden, haben Potential und sind gute Kandidaten für weitere Verbesserungen.