

UNIVERSITÄTSKLINIKUM HAMBURG-EPPENDORF

Zentrum für Experimentelle Medizin
Institut für Anatomie und Experimentelle Morphologie

Direktor Prof. Dr. med. Udo Schumacher

Structural characterisation of peroxisomal import receptor complexes

Dissertation

zur Erlangung des Grades eines Doctor of Philosophy
an der Medizinischen Fakultät der Universität Hamburg.

vorgelegt von:

Evdokia- Anastasia Giannopoulou
aus Patras

Hamburg 2017

Angenommen von der Medizinischen Fakultät der Universität Hamburg am: 04.12.2017

Veröffentlicht mit Genehmigung der Medizinischen Fakultät der Universität Hamburg.

Prüfungsausschuss, der/die Vorsitzende: Prof. Dr. Georg H. Lüers

Prüfungsausschuss, zweite/r Gutachter/in: Dr. Christian Löw

This thesis describes work performed under the joint supervision of Prof. Dr. Matthias Wilmanns at the European Molecular Biology Laboratory (EMBL) in Hamburg, Germany, and Prof. Dr. Georg H. Lüers at the Universitätsklinikum Hamburg-Eppendorf (UKE). This work was supported by the Marie Curie Initial Training Network (ITN) - PERFUME Early Stage Researcher (ESR) Scholarship, funded by European Union Seventh Framework Programme, under Grant Agreement No. 316723 and the Deutsche Forschungsgemeinschaft programme PerTrans FOR 1905, for the period 2016-2017.

Στην υπέροχη οικογένειά μου

I. Contents

1. Introduction	1
1.1. Peroxisomes: dynamic organelles.....	1
1.1.1. Peroxisome metabolism.....	2
1.1.2. Peroxisomal disorders.....	4
1.1.3. Peroxisome formation.....	6
1.1.4. Peroxisomal membrane biogenesis.....	7
1.1.4.1. Peroxisomal biogenesis factor 19 (Pex19).....	10
1.1.4.2. Peroxisomal biogenesis docking factor 3 (Pex3).....	14
1.1.5. Peroxisome matrix protein import.....	17
1.1.5.1. Peroxisomal targeting signal receptor 5 (Pex5).....	20
1.1.5.2. Peroxisomal membrane protein 14 (Pex14).....	21
1.1.5.3. Peroxisomal cargo proteins.....	22
1.2. Aims of this thesis.....	24
2. Materials and methods.....	25
2.1. Buffers and reagents	25
2.2. Molecular biology techniques.....	26
2.2.1. Cloning.....	26
2.2.1.1. Restriction cloning, DNA isolation and agarose gel electrophoresis	28
2.2.1.2. Bacterial transformation and clone verification.....	29
2.2.2. Protein expression and purification	30
2.2.2.1. Protein expression and cell lysis	30
2.2.2.2. Affinity chromatography protein purification	32
2.2.2.3. Size-exclusion chromatography purification (SEC).....	33
2.2.2.1. Protein analysis and quantification	35
2.2.2.2. Blue native PAGE.....	36
2.3. Biophysical characterisation of proteins.....	37
2.3.1. Circular dichroism spectroscopy (CD)	37
2.3.2. Thermofluor differential scanning fluorimetry.....	37
2.3.3. Size exclusion chromatography-light scattering techniques.....	38

2.4.	Protein-protein interaction analysis techniques	39
2.4.1.	Fluorescence anisotropy polarisation.....	39
2.4.2.	Chemical crosslinking-mass spectrometry (XL-MS)	39
2.4.3.	Native mass spectrometry	40
2.5.	Structural characterisation of proteins	42
2.5.1.	X-ray crystallography	42
2.5.1.1.	Principles of protein crystallisation	42
2.5.1.2.	Crystallographic symmetry	45
2.5.1.3.	Miller index.....	47
2.5.1.4.	Basics of X-ray diffraction.....	47
2.5.1.5.	Reciprocal lattice	48
2.5.1.6.	The Ewald construction	49
2.5.1.7.	Atomic form factor and structure factor	50
2.5.1.8.	Electron density and the phase problem	51
2.5.1.9.	Synchrotron radiation.....	53
2.5.2.	Experimental procedures in X-ray crystallography	53
2.5.2.1.	Crystallisation	53
2.5.2.2.	Data collection	54
2.5.2.3.	Data processing.....	54
2.5.2.4.	Model building and refinement.....	57
2.5.2.5.	Structure analysis	58
2.5.3.	Small angle X-ray scattering (SAXS).....	58
2.5.3.1.	Theoretical background	58
2.5.3.2.	SAXS data acquisition	60
2.5.3.3.	SAXS data analysis.....	60
2.5.3.4.	Ab initio modelling.....	61
2.5.4.	Negative stain electron microscopy (EM)	63
3.	Results	64
3.1.	Structural studies of human PEX19 complexes.....	64
3.1.1.	Structural and biophysical characterisation of the human PEX3-PEX19 complex.....	64

3.1.1.1.	Expression, purification and biophysical characterisation of the human PEX3 _{C235S(41-373)} -PEX19 ₍₁₋₂₉₉₎ complex.....	65
3.1.1.2.	Identification of PEX3 _{C235S(41-373)} - PEX19 ₍₁₋₂₉₉₎ binding sites via XL-MS	67
3.1.1.3.	SAXS analysis for PEX3 _{C235S(41-373)} -PEX19 ₍₁₋₂₉₉₎ indicates a flexible multi-domain system	71
3.1.1.4.	Crystallisation of PEX3-PEX19 constructs	75
3.1.2.	Structural and biophysical characterisation of the human PEX19-PEX14 complex.	77
3.1.2.1.	Expression, purification and biophysical characterisation of human PEX19-PEX14	77
3.1.2.2.	Identification of secondary PEX19-PEX14 binding sites via XL-MS	81
3.1.2.3.	Structural characterisation of PEX19-PEX14 via negative stain EM.....	83
3.1.2.4.	Low resolution structural characterisation of PEX19-PEX14 via SAXS suggests flexible elongated structure.....	84
3.1.2.5.	Crystallisation of PEX19-PEX14 constructs	86
3.1.3.	Interaction of human PEX19 with PMPs.....	87
3.1.3.1.	The interaction of PEX19 with ACBD5 depends on the charge of its C-terminus.....	87
3.1.3.2.	Crystallisation of PEX19 with PMP peptides.....	88
3.2.	Structural and biophysical characterisation of PTS1 import proteins	89
3.2.1.	Structural characterisation of Pcs60.....	89
3.2.1.1.	Expression, purification and characterisation of Pcs60	89
3.2.1.2.	Crystallisation and structure determination of Psc60 via X-ray crystallography.....	90
3.2.2.	Structural characterisation of the Pex5p-Pcs60 complex.....	96
3.2.2.1.	Expression and purification of Pex5p/Pcs60 complexes	96
3.2.2.2.	Low resolution SAXS analysis of Pex5p/Pcs60	97
3.2.2.3.	Crystallisation of Pex5p/Pcs60 complexes	100
3.2.2.4.	Structural analysis of Pex5pFL/Pcs60 via negative stain EM	101
3.2.3.	Structural characterisation of AtMIF1	103
3.2.3.1.	Expression, purification and biophysical characterisation of AtMIF1	103
3.2.3.2.	Structural characterisation of AtMIF1 via X-ray crystallography	103
4.	Discussion.....	107

4.1.	Structural insight into peroxisomal membrane complexes.....	107
4.1.1.	Structural characterisation of the PEX3 _{C235S(41-373)} -PEX19 complex	107
4.1.2.	The mPTS charge can affect PEX19 binding	109
4.1.3.	Novel overlapping binding site on PEX19 can accommodate both PEX3 and PEX14.....	110
4.1.4.	PEX19 facilitates PEX14 solubility.....	112
4.1.5.	Structure analysis of the full-length PEX19-PEX14 complex using a hybrid approach.....	113
4.2.	Structural insight into peroxisomal matrix protein import	116
4.2.1.	Structure of the novel peroxisomal enzyme Pcs60	116
4.2.2.	Structural analysis of Pex5p-Pcs60 complexes.....	117
4.2.3.	Characterisation of peroxisomal cargo protein MIF1 from plants.....	119
5.	Perspectives	120
6.	Summary.....	121
7.	Zusammenfassung	122
8.	Abbreviations	123
9.	Bibliography	125
10.	Appendix	138
10.1.	DNA and protein sequences.....	138
11.	<i>Curriculum vitae</i>	143
12.	Acknowledgements	144
13.	Eidesstattliche Versicherung	146

II. Nomenclature

In this doctoral thesis, the following nomenclature has been followed for the distinction of genes and proteins from various organisms, as is commonly used by scientists in the peroxisomal community (N denotes number of peroxin):

- Genes from *Saccharomyces cerevisiae*: *pexN*
- Proteins from *Saccharomyces cerevisiae*: PexNp
- Genes from *Homo sapiens*: *PEX_N*
- Proteins from *Homo sapiens*: PEXN

1. Introduction

1.1. Peroxisomes: dynamic organelles

Eukaryotic cells are divided into elaborate subcellular compartments, which are specialised to carry out a variety of different processes, essential to the survival of the cell. These enclosed organellar compartments are separated from the cytoplasm by intracellular membrane systems that are characterised by the presence of membrane transport proteins, responsible for the import and export of specific components into their respective organelles.

Peroxisomes are such single membrane-bound eukaryotic compartments that contain a multitude of essential enzymes. To date, more than 50 peroxisomal enzymes with various peroxisome-specific, anabolic and catabolic functions have been identified to be involved in fundamental metabolic processes,

such as detoxification of reactive oxygen species (ROS) and the biosynthesis of several types of lipids (Wanders & Waterham, 2006). Other major enzymatic components that they contain are hydrogen peroxide (H_2O_2)-producing oxidases and the hydrogen peroxide-decomposing catalase, the presence of which contributed to their name (De Duve & Baudhuin, 1966). In addition to H_2O_2 , a range of other substrates, including uric acid, amino and fatty acids are being employed by oxidative reactions in peroxisomes (Nordgren &

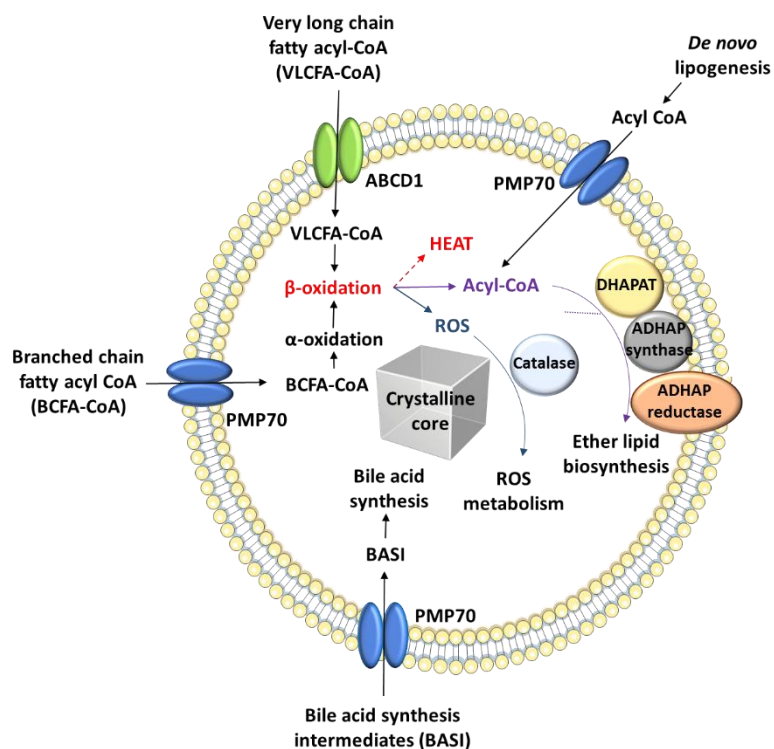


Figure 1: Function of peroxisomes. The main metabolic functions of peroxisomes cells include β -oxidation of very long chain fatty acids, α -oxidation of branched chain fatty acids, synthesis of bile acids and ether-linked phospholipids, and metabolism of reactive oxygen species. Peroxisomes in many, but not all, cell types contain a dense crystalline core of oxidative enzymes (Adapted from (Lodhi & Semenkovich, 2014).

Fransen, 2014). It has been reported that due to the high concentration of enzymes in the peroxisomal lumen, customarily urate oxidase, electron-dense crystalline cores have been detected in several species (Hruban and Swift 1964; Tsukada, Mochizuki, and Fujiwara 1966; van Dijken et al. 1975; Tanaka et al. 1976; Völkl, Baumgart, and Fahimi 1988) (**Figure 1**).

Besides their unique metabolic capabilities, peroxisomes are able to adapt their content, numbers and morphology to the requirements of the cell upon changing environmental cues, in a dynamic manner (van der Klei *et al.*, 2006; Till *et al.*, 2012a; Schrader *et al.*, 2013). However, in contrast to other dynamic organelles such as mitochondria and chloroplasts, peroxisomes do not contain their own genetic material and are thus believed to import all their proteins post-translationally from the cytoplasm (Lazarow & Fujiki, 1985).

Regardless of their organism of origin, a distinct set of proteins encoded by *PEX* genes, which are collectively called **peroxins**, are indispensable for peroxisome maintenance and are involved in several processes, including peroxisome biogenesis (Purdue & Lazarow, 2001), proliferation (Motley & Hetteema, 2007; Schrader *et al.*, 2012) and turnover by autophagy (Till *et al.*, 2012b; Fagarasanu *et al.*, 2010; Reggiori, 2005). In yeast, more than 30 peroxins have been identified so far (Dimitrov *et al.*, 2013) and many of them are conserved in mammals, where, with the exception of minor variations, the principle mechanisms are conserved amongst all higher eukaryotes (Brocard & Hartig, 2014).

1.1.1. Peroxisome metabolism

Fatty acid β -oxidation

Peroxisomes are rich in lipid metabolising enzymes (Kannenberg *et al.*, 1999; Lazarow & De Duve, 1976; van den Bosch *et al.*, 1992) and one of their most important functions is complementation of mitochondrial β -oxidation by contributing to fatty acid oxidation (Eaton *et al.*, 1996; Felber & Golay, 1995; Elferink *et al.*, 2000; Kunau *et al.*, 1995). Although the process of β -oxidation is similar between mitochondria and peroxisomes, there are some significant differences; mitochondrial β -oxidation mostly involves short-, medium- and long-chain fatty acids, whereas its peroxisomal counterpart is primarily responsible for the degradation of very-long chain fatty acids or VLCFAs ($> C_{20}$) (Eaton *et al.*, 1996; Hashimoto, 1999; Reddy & Mannaerts, 1994), dicarboxylic

fatty acids, prostanoids, and the bile acid intermediates di-/tri-hydroxycholestanoic acid (DHCA, THCA) (Wanders, 2004; Mannaerts & Van Veldhoven, 1993). Deficiency in the metabolism of VLCFAs has been shown to cause the peroxisomal disorder X-adrenoleukodystrophy (Singh *et al*, 1984; Kemp *et al*, 2004).

Ether-Phospholipid and plasmalogen biosynthesis

Ether phospholipids represent a special class of phospholipids that are characterised by an alkyl- or alkenyl- bond at the sn-1 position of the glycerol backbone. Plasmalogens are the most abundant subclass of alkenyl-glycophospholipids, which contain an α,β -unsaturated ether bond and their synthesis requires functional peroxisomes (da Silva *et al*, 2012). Plasmalogens are important antioxidant factors as well as mediators of membrane structure, dynamics and signal transduction (Wanders & Brites, 2010).

Fatty acid α -oxidation

The process of α -oxidation is fundamental for the metabolism of 3-methyl-branched fatty acids (such as phytanic acid), which cannot be metabolised *via* β -oxidation, as they contain a methyl group on their 3-position. Having undergone α -oxidation, this methyl group is removed and the product is activated on the outer peroxisomal membrane by a long chain acyl-CoA synthetase (LACS), whereupon it enters the peroxisome to undergo further metabolism and β -oxidation. Enzymatic deficiency in α -oxidation can result in Refsum's disease, a peroxisomal disorder characterised by high levels of phytanic acid, whose derivatives and can lead to neurological damage (Wanders *et al*, 2001).

Biosynthesis of cholesterol and other isoprenoids

Peroxisomes are crucial for cholesterol biosynthesis. They contain a number of enzymes, such as mevalonate kinase, phosphomevalonate kinase and mevalonate pyrophosphate decarboxylase, as well as other rate-limiting enzymes for the synthesis of cholesterol, which were previously thought to reside in the cytosol or the endoplasmic reticulum (ER) (Kovacs *et al*, 2002; Wanders & Brites, 2010).

Detoxification of Glyoxylate

Peroxisomes contain the enzyme glyoxylate aminotransferase which is involved in the metabolism of glyoxylate into glycine. In absence of this enzyme, glyoxylate is converted to oxalate by the enzyme lactate dehydrogenase, which is a toxic metabolite that can cause severe consequences for the cell (Wanders & Brites, 2010).

1.1.2. Peroxisomal disorders

The function of peroxisomes in cell physiology has become more apparent in recent years. As one of the central roles of peroxisomes is lipid metabolism, defects in one or more peroxisomal functions can cause a range of serious clinical phenotypes. Peroxisomal disorders can be classified in two major groups: (1) disorders of peroxisome biogenesis (PBDs) which are caused by defects in any of at least 14 different *PEX* genes (Waterham & Ebberink, 2012) and (2) single peroxisome enzyme deficiencies that affect single metabolic pathways (Wanders, 2004).

PBD patient cells are characterised by deficiency of functional peroxisomes. There are two distinct clinical subtypes: the Zellweger syndrome spectrum (ZSS) disorders and Rhizomelic Chondrodysplasia punctata (RCDP) type 1. Cells from severe ZSS disorders display entire absence of peroxisomal organelles as a result of a variety of defective *PEX* genes, although peroxisomal membrane remnants are in some cases still present. In milder cases, small numbers of functional peroxisomes can be detected. Conversely, RCDP type 1 is caused by mutations in the *PEX7* gene, which is encoding an essential protein involved in

Table 1: Most common *PEX* gene deficiencies in PBD patients as reported by Waterham and Ebberink, 2012.

<i>PEX</i> gene	Frequency (%) among PBDs	Frequency (%) among ZSS disorders
<i>PEX1</i>	48.5	58.9
<i>PEX2</i>	3.2	3.9
<i>PEX3</i>	0.5	0.5
<i>PEX5</i>	1.4	1.7
<i>PEX6</i>	13.1	15.9
<i>PEX7</i>	17.7	–
<i>PEX10</i>	3.4	4.2
<i>PEX11B</i>	< 0.1	< 0.1
<i>PEX12</i>	5.9	7.1
<i>PEX13</i>	1.4	1.7
<i>PEX14</i>	0.2	0.3
<i>PEX16</i>	1.0	1.2
<i>PEX19</i>	0.3	0.4
<i>PEX26</i>	3.4	4.2
TOTAL	100	100

peroxisomal matrix protein import. This deficiency results in the impaired import of a subset of peroxisomal enzymes, affecting the biochemical, cellular and clinical phenotype of the patients (Waterham & Ebberink, 2012). Therefore, absence of functional peroxisomes in PBD patients promotes degradation of peroxisomal enzymes in the cytoplasm and consequently affects important metabolic pathways. This becomes apparent from the accumulation of substrates commonly metabolised in peroxisomes that are found in PBD patients (VLCFAs, pristanic acid, phytanic acid, DHCA and THCA) and a shortage of peroxisome metabolism products (plasmalogens) (Waterham & Ebberink, 2012). **Table 1** summarises typical *PEX* gene deficiencies recorded to cause PBDs. The second group of peroxisomal disorders encompasses all single peroxisomal enzyme/ transporter deficiencies, which can be further distributed into subgroups depending on the metabolic pathways affected (**Table 2**).

Table 2: List of the single peroxisomal enzyme deficiencies (Wanders, 2004)

Peroxisomal pathway affected	Peroxisomal disease	Enzyme defect	Gene involved
Ether phospholipid synthesis	Rhizomelic chondrodysplasia punctata Type 2 (DHAPAT deficiency)	DHAPAT	<i>GNPAT</i>
Peroxisomal β-oxidation	Rhizomelic chondrodysplasia punctata Type 3 (alkyl-DHAP synthase)	ADHAPS	<i>AGPS</i>
	X-linked adrenoleukodystrophy	ALDP	<i>ABCD1</i>
	Acyl-CoA oxidase deficiency	ACOX1	<i>ACOX1</i>
	D-bifunctional protein deficiency	DBP	<i>HSD17B4</i>
	2-MethylacylCoA racemase deficiency	AMACR	<i>AMACR</i>
	Sterol carrier protein X deficiency	SCPx	<i>SCP2</i>
Peroxisomal α-oxidation	Refsum disease (phytanoyl-CoA hydroxylase deficiency)	PHYH/PAHX	<i>PHYH/PAHX</i>
Glyoxylate detoxification	Hyperoxaluria Type 1	AGT	<i>AGXT</i>
H₂O₂-metabolism	Acatasaemia	CAT	<i>CAT</i>

1.1.3. Peroxisome formation

The origin of peroxisomes is a widely investigated topic, but remains controversial to this day (Dimitrov *et al*, 2013; Agrawal & Subramani, 2016; Schrader *et al*, 2016). The most popular theories suggest that peroxisomes can arise either by growth and fission of pre-existing organelles (Motley & Hettema, 2007; Purdue & Lazarow, 2001; Schrader *et al*, 2012; Lazarow & Fujiki, 1985) or, alternatively, through a *de novo* budding process from the ER (Hoepfner *et al*, 2005; Kragt *et al*, 2005b; Tam *et al*, 2005; Kim *et al*, 2006). There is evidence to support both of these hypotheses, indicating that the two pathways may exist in parallel, but their regulation may depend upon discrete growth conditions and metabolic requirements (Agrawal & Subramani, 2013; Nuttall *et al*, 2011) (**Figure 2**).

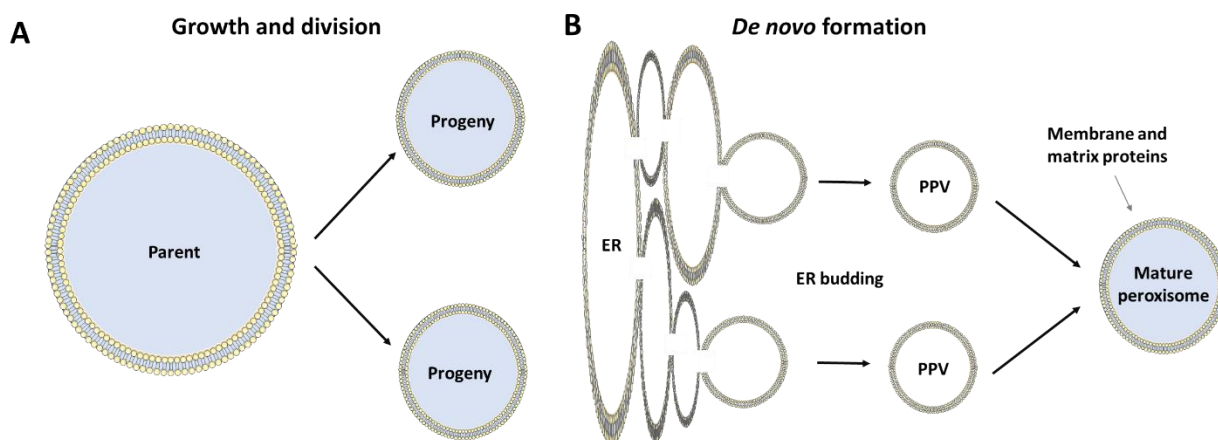


Figure 2: Peroxisome biogenesis models. (A) Peroxisomes are generated autonomously through growth and division of pre-existing organelles, (B) Peroxisomes can arise via a *de novo* process involving budding from the ER followed by import of membrane and matrix proteins from the cytosol.

According to the growth and fission model (**Figure 2A**), new peroxisomes are formed by duplication of pre-existing ones, making use of an intricate fission and proliferation machinery. This model can be described as a process of several sequential steps, including (1) peroxisome polarisation, (2) membrane elongation, (3) import of peroxisomal membrane and matrix proteins, and (4) membrane fission (Koch *et al*, 2010; Delille *et al*, 2010; Fagarasanu *et al*, 2007; Koch & Brocard, 2012). Although the mechanistic details may vary between different organisms, proteins of the Pex11 family have been implicated as essential in the regulation of peroxisome abundance (Erdmann & Blobel, 1995; Marshall *et al*, 1995; Fagarasanu *et al*, 2007). It has been shown that

deletion of Pex11 in yeast can interrupt the fission process and promote formation of giant peroxisomes in yeast cells (Erdmann & Blobel, 1995). Coupled with the various Pex11 proteins, mitochondrial fission factors, such as mitochondrial fission-1 protein Fis-1 and the dynamin-related protein Drp-1 are also employed in the peroxisome division process (Gandre-Babbe & van der Blik, 2008; Koch *et al.*, 2003, 2005).

The alternative model suggests that peroxisomes are semi-autonomous and peroxisomal membranes arise *de novo* from the ER (**Figure 2B**), where a number of peroxisomal membrane proteins are already localised, whilst their matrix proteins are imported from the cytoplasm (Tabak *et al.*, 2013). In this premise, newly synthesised peroxin 3 (Pex3) is originally sorted into the ER membrane, into foci that will subsequently bud into the cytoplasm, forming new peroxisomes. Other hypotheses propose that distinct classes of pre-peroxisomal vesicles (PPVs) stem from the ER to heterotypically fuse and form mature peroxisomes (Titorenko *et al.*, 2000; van der Zand *et al.*, 2012).

Despite the differences in these two models, it is generally accepted that in both cases two peroxins, Pex3 and Pex19, are essential for the formation of peroxisomal membranes and mature peroxisomes. Pex3 and Pex19 are classified as peroxisomal biogenesis factors and are important for the correct topogenesis of peroxisomal membrane proteins (PMPs) (Fujiki *et al.*, 2006). Both $\Delta pex3$ and $\Delta pex19$ cells exhibit lack of peroxisomal structures (Höhfeld *et al.*, 1991; Subramani, 1998); in the fission model, PMPs are unstable and become mislocalised to the cytoplasm (Hetteema *et al.*, 2000), whereas in the *de novo* model PMPs accumulate to the ER membrane (van der Zand *et al.*, 2010, 2012). Interestingly, it has been discovered that PPVs can exist in $\Delta pex3$ and $\Delta pex19$ cells, which contain the PMPs Pex13 and Pex14, but lack other PMPs. This could indicate that these two PMPs are targeted to peroxisomes in a Pex3/Pex19-independent manner (Knoops *et al.*, 2014).

1.1.4. Peroxisomal membrane biogenesis

Essential requirement for the biogenesis and formation of mature peroxisomes is the assembly of PMPs on the peroxisomal membrane, which will allow the import of peroxisomal matrix proteins and other metabolites to the peroxisomal lumen *via* their corresponding import machinery.

From known peroxisome biogenesis factors, Pex3, Pex19, and Pex16 appear to be crucial for peroxisomal membrane biogenesis (Kim *et al*, 2006; Fujiki *et al*, 2006), as in their absence, peroxisome structures are absent and PMPs are mislocalised or degraded (Kim & Hettema, 2015).

In principle, PMPs are synthesised on free polyribosomes and are inserted post-translationally into the peroxisomal membrane (Purdue & Lazarow, 2001). PMPs are generally classified as **Class I** and **Class II** PMPs that correspond to whether they are translocated to the membrane in a Pex19-dependent or Pex19-independent fashion respectively. Unlike Class II PMPs, Class I PMPs contain a target sequence referred to as **mPTS** (membrane peroxisome targeting signal) that is recognised by Pex19, targeting directly at the peroxisomal membrane. On the other hand, Class II PMPs are hypothesised to first insert into the ER membrane, thus sorting to peroxisomes indirectly (Fujiki *et al*, 2006).

The target sequence of Class I PMPs contains one or more non-overlapping signal motifs that consist of a Pex19-targeting element and a membrane-docking sequence (Jones *et al*, 2001; Heiland & Erdmann, 2005), which can be functionally or physically separated (Jones *et al*, 2001; Fransen *et al*, 2001; Girzalsky *et al*, 2006). Although no unambiguous mPTS sequence consensus has been deduced, it is generally accepted that most mPTS sequences exhibit some common characteristics, such as presence of one or more short α -helical segments that are made up of positively charged and hydrophobic residues, as well as at least one transmembrane domain (Dyer *et al*, 1996; Honsho & Fujiki, 2000; Jones *et al*, 2001; Fransen *et al*, 2001; Wang *et al*, 2001; Rottensteiner *et al*, 2004). A special category of PMPs includes tail-anchored (TA) protein which are characterised by the presence of a C-terminal mPTS signal, consisting of a transmembrane domain and a positively charged luminal tail region (Halbach *et al*, 2006).

The general model used to describe PMP transport into the peroxisomal membrane utilises peroxisomal biogenesis factors Pex3 and Pex19 (Fujiki *et al*, 2006). Pex19 is a cytosolic protein which functions as a cycling receptor for newly synthesised PMPs, and displays chaperone-like activity by potentially shielding the hydrophobic parts of the PMP from the aqueous cytosolic environment (Sacksteder, 2000; Fransen *et al*, 2001, 2004; Jones *et al*, 2001; Kashiwayama *et al*, 2005; Shibata *et al*, 2004; Jones *et al*, 2004). Pex3, on the other hand, is a peroxisomal membrane protein that serves as the docking factor for Pex19-PMP complexes and facilitates PMP insertion

into the peroxisomal membrane through an unknown mechanism (Lazarow & Fujiki, 1985; Muntau *et al*, 2003; Jones *et al*, 2004; Fang *et al*, 2004) (**Figure 3**).

Class II PMPs, as mentioned, are employing a Pex19-independent pathway for sorting to peroxisomes. This class of proteins does not contain an identifiable mPTS motif and the mechanism of peroxisome targeting has not been elucidated to date. It has been suggested that PMPs of this pathway (originally Pex3) are sorted indirectly to peroxisomes *via* the ER, through an unknown procedure that utilises Sec61 or the Get pathway and includes a vesicular step (Dimitrov *et al*, 2013; Tabak *et al*, 2013; Aranovich *et al*, 2014). Other proteins that are directly associated to the ER include Pex16 in plants and mammals, as well as Pex15, Pex22, Pex13, and Pex14 in yeast (Jones *et al*, 2004; Fransen *et al*, 2004; Kim *et al*, 2006; Halbach *et al*, 2009; Lam *et al*, 2011).

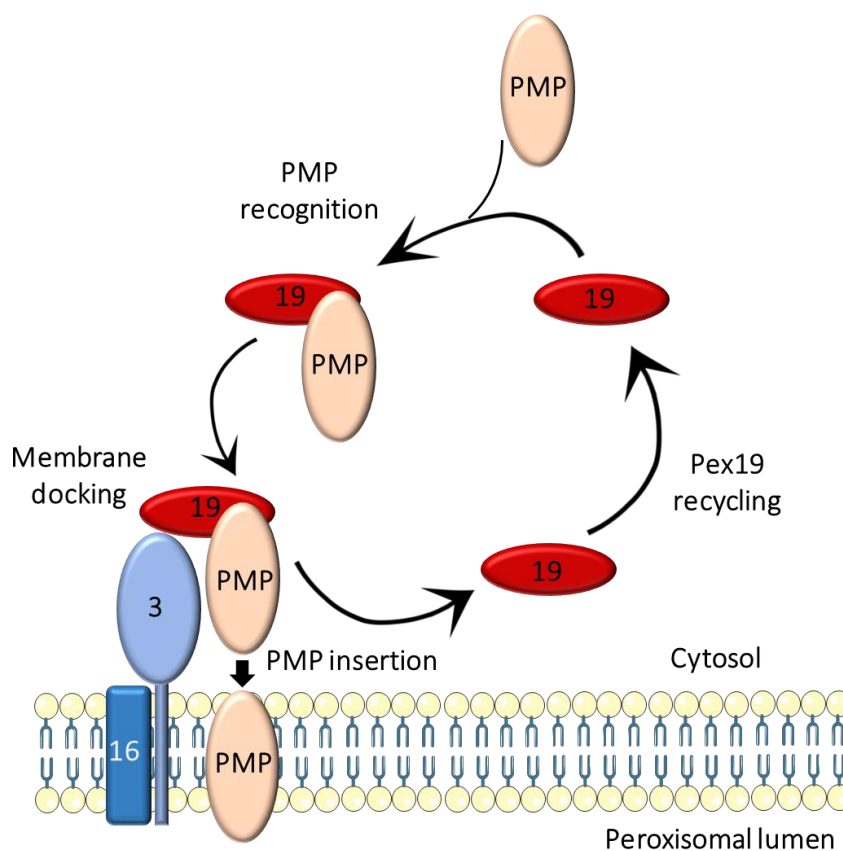


Figure 3: Pex19-mediated peroxisomal membrane protein import. Class I PMPs are imported post-translationally to the peroxisomal membrane. Pex19 recognises PMPs, binds them in the cytosol and transports them to the peroxisomal membrane, where it docks on a complex containing Pex3 (and Pex16 in mammals). Following insertion of the PMP, Pex19 is recycled back to the cytosol.

1.1.4.1. *Peroxisomal biogenesis factor 19 (Pex19)*

Pex19 is a multifunctional peroxisomal protein that is essential for peroxisome biogenesis in all peroxisome-containing organisms. It has been characterised as both a shuttling PMP receptor and a chaperone, with the capacity to recycle in the cytosol after PMP release (Matsuzono & Fujiki, 2006). Although, it displays a variety of functions, its central role involves PMP import (Fujiki *et al*, 2006; Agrawal & Subramani, 2013; Theodoulou *et al*, 2013; Hettema *et al*, 2014). The main role of Pex19 is to bind and stabilise Class I PMPs in the cytoplasm through mPTS motif recognition. Through its chaperone-like function, it can act as a quality control element for proper folding of PMPs before insertion into the peroxisomal membrane (Halbach *et al*, 2006; Jones *et al*, 2004).

Particularly in the *de novo* peroxisome biogenesis model, Pex19 has been associated with the ATP-dependent budding of PPVs from the ER, which will eventually mature into functional peroxisomes (Agrawal & Subramani, 2013; Dimitrov *et al*, 2013). Furthermore, Pex19 has been speculated to be involved in peroxisomal inheritance by binding to motor protein Myosin 2 (Myo2) in a farnesylation-dependent way, although the precise mechanism has not been elucidated (Otzen *et al*, 2012). Moreover, there is evidence that Pex19 can form multiple, dynamic complexes *in vivo* and *in vitro*, and has been implicated as an important factor for the import of peroxisomal matrix proteins through the regulation of the Pex14 docking complex, although the exact mechanism is not well understood at this point (Fransen *et al*, 2004; Bharti *et al*, 2011; Veenhuis & van der Klei, 2014; Knoops *et al*, 2014). In this respect, Pex19 has been attributed an additional function as an assembly or disassembly factor of peroxisomal membrane protein complexes (Snyder *et al*, 2000), which is not strictly involved with PMP topogenesis.

Pex19 displays a highly conserved domain architecture amongst different vertebrate species, although plant, fungi and amoebozoa exhibit differences in length and sequence of Pex19, implying functional diversity. Pex19 is a generally monomeric and hydrophilic protein that contains a farnesylation motif at its C-terminus (CTD) (Götte *et al*, 1998; Kammerer *et al*, 1997). An exception is found in *Arabidopsis thaliana*, the Pex19 homologue of which has been reported to form a dimer (Hadden *et al*, 2006). Pex19 is predominantly located in the cytosol, but due to its function as a cycling receptor (Jones *et al*, 2004; Schliebs & Kunau, 2004; Fujiki *et al*, 2006),

small amounts can also be found on the peroxisomal membrane (Sacksteder, 2000; Götte *et al*, 1998; Matsuzono *et al*, 1999).

Human PEX19 is a 299-residue, partly disordered protein, characterised by a double-domain architecture that reflects its multiple functions as a PMP chaperone and import receptor (Jones *et al*, 2004), as well as its docking on peroxisomes *via* its interaction with Pex3 (Fang *et al*, 2004; Schmidt *et al*, 2010; Sato *et al*, 2010). It is composed of an unstructured N-terminal and a compact C-terminal domain (Shibata *et al*, 2004; Hattula *et al*, 2014). The C-terminal domain of Pex19 generally displays a higher degree of conservation compared with the N-terminus, with the exception of its N-terminally located Pex3 binding site (**Figure 4A**).

As mentioned, the N-terminal domain of Pex19 contains the Pex3 and Pex14- interacting regions, as well as predicted amphipathic segments that could be involved in stabilising the binding and release of PMPs (Fransen *et al*, 2002; Muntau *et al*, 2003; Jones *et al*, 2004; Fang *et al*, 2004; Hoepfner *et al*, 2005; Neufeld *et al*, 2009; Chen *et al*, 2014). The C-terminus, on the other hand, is characterised by the presence of the PMP-binding site and a so-called CAAX farnesylation motif which has been conserved across several species (**Figure 4B**), with trypanosomal Pex19 being an exception (Banerjee *et al*, 2005). This PMP binding region interacts with the majority of PMPs that have been investigated so far, including peroxisomal tail-anchored proteins, which are PMPs with a single C-terminal transmembrane domain (Halbach *et al*, 2006; Chen *et al*, 2014; Girzalsky *et al*, 2006; Jones *et al*, 2001; Brosius *et al*, 2002; Fransen *et al*, 2001; Halbach *et al*, 2005; Yernaux *et al*, 2009; Gloeckner *et al*, 2000).

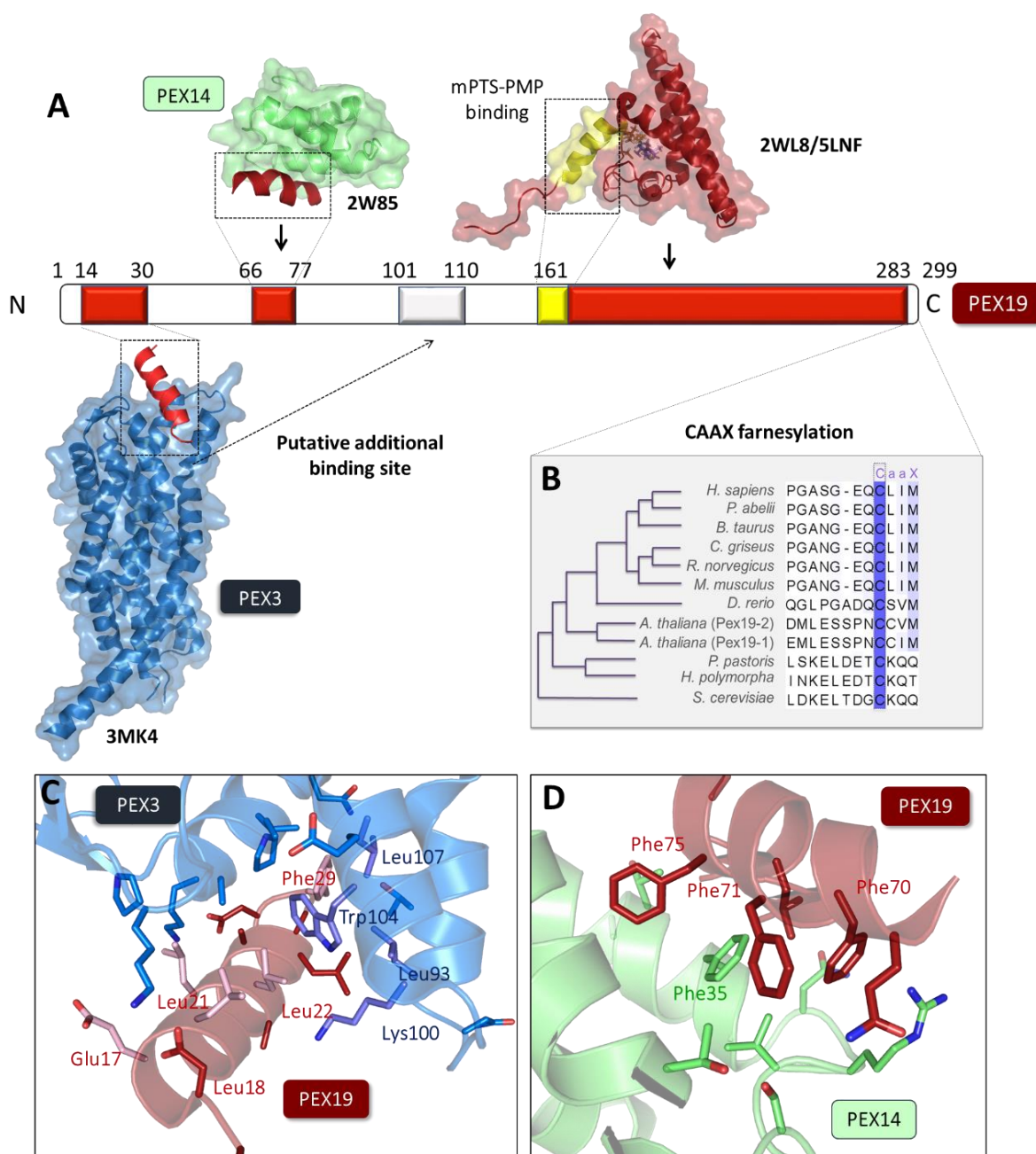


Figure 4: Structure of human PEX19. (A) Schematic representation of PEX19 architecture. It consists of an unstructured N-terminus that contains PEX3 (3MK4) and PEX14 (2W85) binding sites (red), as well as a putative secondary PEX3 binding site (grey). The C-terminus is a globular α -helical bundle (2WL8/ 5LNF) and contains the mPTS recognising helix (yellow). (B) The utmost C-terminal region of PEX19 contains a CAAX farnesylation motif which is conserved in several species. The structure alignment was produced using Clustal-W (Thompson et al, 1994) and manually coloured. (C) Residues involved in PEX3 (blue)–PEX19 (red) interaction. (D) PEX14 (green)–PEX19 (red) interacting residues. Available high-resolution structures are shown with semitransparent surfaces and overlaid ribbons, made using the programme Pymol (Schrodinger LLC, 2010).

Although the three-dimensional structure of full-length Pex19 has not been determined, segmented structural information is available. Most notably, the structure of the folded C-terminus of human PEX19 (residues 161-283), lacking the CAAX farnesylation motif has been resolved *via* X-ray crystallography (PDB entry: 2WL8) and can be described as an α -helical bundle (Schueller *et al.*, 2010). With regards to the partly disordered N-terminal region of Pex19, two short α -helical structures comprising residues 15-40 (PDB entry: 3AJB) and 14-30 (PDB entry: 3MK4) have been determined in complex with Pex3 (**Figure 4C**) (Sato *et al.*, 2010; Schmidt *et al.*, 2010). Additionally, it has been shown *via* NMR spectroscopy that residues 66-77 of human Pex19 also adopt an α -helical conformation when bound to Pex14 (**Figure 4D**) (PDB code 2W85) (Neufeld *et al.*, 2009). Lastly, the N-terminal region of Pex19 has been predicted to contain a number of amphipathic α -helices rich in hydrophobic residues, highly conserved amongst different species (Shibata *et al.*, 2004; Fransen *et al.*, 2005; Chen *et al.*, 2014).

From the above, it can be concluded that Pex19 can interact with multiple proteins, the best studied of which is its interaction with Pex3, showing a high binding affinity of ≤ 10 nM (Sato *et al.*, 2010; Schmidt *et al.*, 2010). Comparatively, the binding affinity of the N-terminal binding site of Pex19 that has been structurally determined is about 5-10 times lower than that of the full length Pex19, but sufficient to establish a stable interaction. A second low affinity binding site within the N-terminus of Pex19 could potentially explain the difference in the recorded binding affinities (Fransen *et al.*, 2005; Matsuzono *et al.*, 2006; Schmidt *et al.*, 2010, 2012).

Moreover, the N-terminal part of Pex19 contains a Pex14 binding region. Pex14 is a peroxisomal membrane protein that is an essential component of the matrix protein import machinery. Unlike the majority of Class I PMPs, it does not contain a typical mPTS motif, but its Pex19-interacting site is located at its N-terminus (residues 16-80) and contains an F/YFxxxF sequence motif (Neufeld *et al.*, 2009). In the same study it was shown that the same Pex14 N-terminal segment can bind to a WxxxF/Y sequence motif found on the peroxisomal matrix protein import receptor Pex5, but with much higher affinity (0.5 versus 9 μ M) and in an antiparallel orientation (Hattula *et al.*, 2014).

As has been demonstrated in several works, the structure of the C-terminal part of Pex19 defines an autonomous mPTS-binding domain and its structural stability is crucial for the

interaction with a variety of PMPs (Gloeckner *et al*, 2000; Brosius *et al*, 2002; Shibata *et al*, 2004; Rottensteiner *et al*, 2004; Fransen *et al*, 2004; Kashiwayama *et al*, 2005; Halbach *et al*, 2005; Matsuzono & Fujiki, 2006; Girzalsky *et al*, 2006; Halbach *et al*, 2009; Schueller *et al*, 2010; Yagita *et al*, 2013). Despite being the subject of numerous studies, little is known about the mechanism of mPTS-mediated binding of PMPs by Pex19. Truncated Pex19 constructs have been used to pinpoint the exact mPTS recognition site, which seems to be formed by an amphipathic α -helix (residues 171-182) placed adjacent to the helically folded C-terminal domain (Schueller *et al*, 2010). Binding affinity measurements indicate a moderate dissociation constant of about 10 μ M between PMP peptides and the Pex19 C-terminus. Substitution of a number of exposed hydrophobic residues in this region results in loss of PMP binding, indicating that non-specific hydrophobic interactions are essential for this interaction (Schueller *et al*, 2010).

The structure of the Pex19 C-terminus contains a large cavity at its core, which has been suggested to accommodate the C-terminal post-translationally farnesylated C-terminus (Schueller *et al*, 2010), and later demonstrated by NMR spectrometry (Emmanouilidis *et al*, 2017). Furthermore, it has been shown that farnesylated Pex19 displays increased PMP binding affinity (Rucktäschel *et al*, 2009), whereas the non-farnesylated CAAX motif considerably reduces the binding affinity between Pex19 and PMPs (Schueller *et al*, 2010). The role of the farnesylation remains controversial to this point, but its existence may be important for efficient PMP recognition and targeting to peroxisomes (Götte *et al*, 1998; Matsuzono *et al*, 1999; Sacksteder, 2000; Rucktäschel *et al*, 2009; Fransen *et al*, 2004). Theories suggest that farnesylation may occur prior to Pex3 binding and may increase its affinity, although it is not required for this interaction (Fransen *et al*, 2001, 2002; Jones *et al*, 2004; Fransen *et al*, 2005; Rucktäschel *et al*, 2009).

1.1.4.2. Peroxisomal biogenesis docking factor 3 (Pex3)

Pex3 is an integral membrane protein that is central to peroxisome biogenesis in addition to a number of other cellular processes (Yamashita *et al*, 2014; Ghaedi *et al*, 2000; Fujiki *et al*, 2006; Schliebs & Kunau, 2004). It is a fundamental part of the PMP import machinery, as it functions as the docking factor for Pex19-PMP complexes (Fang *et al*, 2004; Fransen *et al*, 2005; Matsuzono & Fujiki, 2006). Although Pex3 contains a transmembrane N-terminal domain, it does not interact

with the PMP binding site of Pex19, but instead with a distinct site on the Pex19 N-terminus. Moreover, its membrane insertion does not depend on Pex19, classifying Pex3 as a class II PMP (Jones *et al*, 2004).

The N-terminal transmembrane domain of Pex3 is preceded by a short and basic luminal segment, termed as n-region and h-region, due to its similarity with ER signal anchor-like sequences (Thoms *et al*, 2012). This segment has also been referred to as mPTS2 (Kim & Hettema, 2015) and is essential for anchoring Pex3 into the peroxisomal membrane or peroxisomal precursor structures such as the ER membrane and PPVs (Thoms *et al*, 2012; Fakieh *et al*, 2013). Recognition of the n/h-regions likely relies on the Sec61 endomembrane channel system associated with the ER (Thoms *et al*, 2012), providing mechanistic insight into Pex3 transport *via* the ER in yeast (Hoepfner *et al*, 2005; Kragt *et al*, 2005b; Tam *et al*, 2005). Alternatively, in vertebrates, the ER receptor function is provided by the elusive protein Pex16 (Kim & Mullen, 2013), which also serves as a Pex3/Pex19 membrane docking component in mature peroxisomes (Matsuzaki & Fujiki, 2008).

Similar to Pex19, vertebrate Pex3 sequences are also highly conserved and span about 370 amino acid residues in length. On the contrary, fungi and plant Pex3 only share <35% sequence identity with human PEX3 and are divergent in length, number of predicted transmembrane domains, and topology (Ghaedi *et al*, 2000; Soukupova *et al*, 1999; Gloeckner *et al*, 2000; Hunt & Trelease, 2004; Haan *et al*, 2002; Giannopoulou *et al*, 2016).

Human PEX3 is a 373 residue-long protein that is anchored on the peroxisomal membrane *via* its N-terminus, while the rest of the protein forms a globular cytosolic domain (Kammerer *et al*, 1998; Soukupova *et al*, 1999; Ghaedi *et al*, 2000) (**Figure 5**). Despite the fact that not much is known about the arrangement of the full length PEX3 on the peroxisomal membrane, the two structures that have been determined in complex with N-terminal PEX19 fragments (PDB entries 3AJB and 3MK4) have contributed considerable insight into the interaction of PEX3 and PEX19 (Sato *et al*, 2010; Schmidt *et al*, 2010). The overall structure of the PEX3 cytosolic domain forms an α -helical bundle in which the interfaces between tightly packed helices are mostly hydrophobic, whereas the exterior is mostly hydrophilic. There are three structurally conserved regions within PEX3: (1) a hydrophobic groove near the base of the protein, which could be important for PMP

insertion (Schmidt *et al*, 2010; Chen *et al*, 2014), (2) a cluster of acidic residues located on another face of the cytosolic domain of PEX3 whose potential role has not yet been elucidated (Schmidt *et al*, 2012), and most importantly (3) the hydrophobic and compact PEX19-binding site, located distally from the membrane anchoring site (Sato *et al*, 2010) (**Figure 4A, C**).

The PEX19-binding site covers about 600 Å² in size, with three PEX3 loop regions of segments 90-107, 196-197 and 321-330 contributing to the interaction. Several residues involved in this interaction are conserved amongst various species, indicating that the interaction is also conserved. Even though a number of hydrogen bonds are required for the interaction, mainly hydrophobic residues involved in binding are conserved in both PEX3 and PEX19. Furthermore, it has been reported that PEX19 has a limited stabilising effect on PEX3 *in vitro* and in cellular assays, but does not display a chaperone-like activity for it, unlike for other PMPs (Matsuzaki & Fujiki, 2008; Schmidt *et al*, 2012).

Upon Pex19-PMP binding, Pex3 possibly undergoes conformational changes that would allow PMP insertion into the peroxisomal membrane. Ternary Pex3-Pex19-PMP complexes can be isolated *in vitro*, indicating the formation of a transient structural and functional unit (Shibata *et al*, 2004; Matsuzono & Fujiki, 2006) .

The surface of Pex3 exhibits highly hydrophobic interhelical grooves, suggesting additional Pex3 binding partners. Indeed, there is evidence that Pex3 can make contacts with several other proteins to serve processes other than peroxisome biogenesis. Various data showed a connection of Pex3 to pexophagy (Yamashita *et al*, 2014; Williams & van der Klei, 2013; Deosaran *et al*, 2013; Kim *et al*, 2008), which serves to balance peroxisome biogenesis and maintenance by a regulated autophagy procedure. Interaction of Pex3 with proteins identified as important for pexophagy, such as Atg36 and Atg30 (Motley *et al*, 2012; Burnett *et al*, 2015) has been reported, but the exact molecular basis of the interaction has not yet been elucidated. Pex3 has further been implicated in the procedure of peroxisome inheritance in yeast, *via* interaction with peroxisomal inheritance factor Inp1 (Knoblach *et al*, 2013; Munck *et al*, 2009) that reportedly regulates peroxisome motility from the mother to daughter yeast cell.

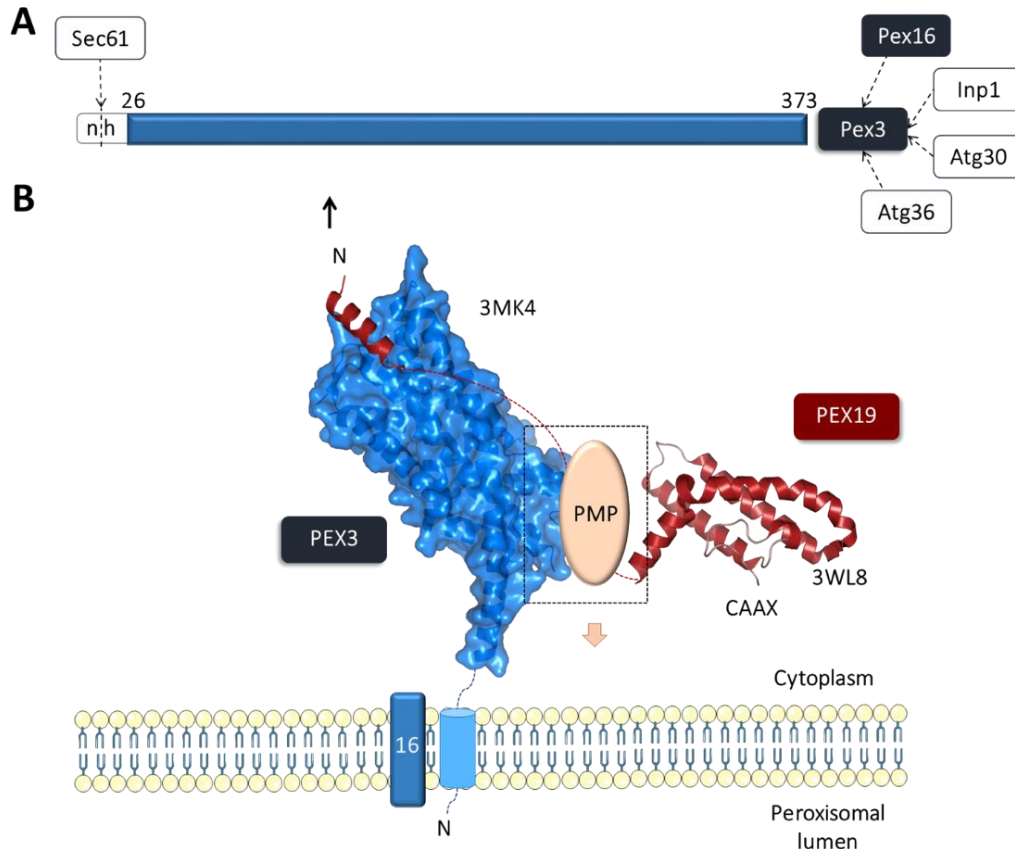


Figure 5: Structure of human PEX3. (A) Schematic representation of PEX3, which contains a transmembrane/peroxisomal domain (white) and a globular cytosolic domain (blue). Reported protein interactions, which have not yet been mapped (including the yeast Sec61, Inp1, Atg30, Atg36 and the human PEX16 are indicated. (B) Possible model of the role of PEX3 in PMP import. Available high-resolution structures are shown with surfaces and ribbons.

1.1.5. Peroxisome matrix protein import

Peroxisomes, unlike other organelles such as mitochondria, do not contain their own DNA. For that reason, an important step for their functional maturation is the import of their required enzymes from the cytoplasm into their matrix (Lazarow & Fujiki, 1985). A noteworthy aspect of peroxisomal matrix protein import is the ability of peroxisomes to import fully folded and oligomeric proteins, unlike other organelles (Wickner & Schekman, 2005). This ability of peroxisomes relies on their highly dynamic pore, which can accommodate a wide range of cargoes (Brul *et al*, 1988b, 1988a; Glover *et al*, 1994; Keller *et al*, 1987; Titorenko *et al*, 2002; Walton *et al*, 1992, 1995; Zhang *et al*, 2003) and their oligomeric states (Freitas *et al*, 2015), as well as allowing proteins to be imported as import partners (“piggyback”) (Titorenko *et al*, 2002;

Effelsberg *et al*, 2015; Islinger *et al*, 2009), with its diameter changing in accordance to the size of the cargo to be imported (Meinecke *et al*, 2010).

Newly synthesised peroxisomal matrix proteins carry specific motifs that can be recognised by cycling receptors in the cytoplasm, which target them to a docking site on the peroxisomal membrane. Subsequently, the receptor-bound cargo protein is translocated to the peroxisomal lumen and the complex disassembles, causing release of the cargo and recycling of the receptor, following an ubiquitination step (Erdmann & Schliebs, 2005a; Rehling *et al*, 2000; Nair *et al*, 2004; Kragt *et al*, 2005a; Liu *et al*, 2012; Platta *et al*, 2007).

The recognition of peroxisomal target proteins by their corresponding peroxisomal receptors occurs *via* a Peroxisomal Targeting Signal motif (PTS) that can be located either at their C-terminus (PTS1) or their N-terminus (PTS2) (Gould *et al*, 1987; Swinkels *et al*, 1991). PTS1 and PTS2 containing proteins employ distinct pathways for their peroxisomal import, although the basic molecular mechanism is similar. The import receptor for PTS1-containing proteins is Pex5 (Dodt & Gould, 1996) which recognises the tripeptide motif (S/A/C)(K/R/H)(L/M/I) (Gould *et al*, 1989; Brocard & Hartig, 2006), whereas PTS2-containing peroxisomal proteins are translocated by the receptor Pex7, which recognises the N-terminal nonapeptide consensus sequence (R/K)/(L/V/I)X5(H(Q))(L/A) (Braverman *et al*, 1997; Swinkels *et al*, 1991).

Pex5 functions both as a receptor for PTS1 cargo (Rucktäschel *et al*, 2011; Terlecky *et al*, 1995), but also as a chaperone, ensuring that they remain folded and active (Freitas *et al*, 2011). It interacts with the PTS1 motif on the cargo *via* a series of tetratricopeptide repeats (TPRs) within its C-terminus (Van der Leij *et al*, 1992; Brocard *et al*, 1994; Szilard *et al*, 1995; Gatto *et al*, 2000). The number of TPR repeats can differ among different species. Once at the peroxisomal membrane, the cargo-loaded Pex5 interacts with the membrane associated proteins Pex14 and Pex13, which form the minimum docking complex required for cargo import (Albertini *et al*, 1997; Schliebs *et al*, 1999; Bottger *et al*, 2000; Saidowsky *et al*, 2001; Niederhoff *et al*, 2005) (**Figure 6A**).

The interaction of Pex5 with Pex14 is mediated by short WxxxF sequence motifs located at the N-terminus of Pex5 (Saidowsky *et al*, 2001; Otera *et al*, 2002; Choe *et al*, 2003; Williams *et al*, 2005). Crystal structures of cargo-loaded or unloaded Pex5 show that cargo binding induces major conformational changes within the receptor, which could be required for docking of the

loaded Pex5 on the peroxisomal pore docking complex (Stanley & Wilmanns, 2006; Shiozawa *et al*, 2009; Fodor *et al*, 2012). It has been suggested that Pex14 favours the binding of cargo-loaded Pex5, while Pex13 exhibits higher affinity for the unloaded Pex5 (Otera *et al*, 2002; Mukai & Fujiki, 2006; Lanyon-Hogg *et al*, 2010), suggesting that Pex13 may be involved in receptor recycling (Erdmann & Schliebs, 2005b). In yeast, another peroxisomal protein, Pex17, is associated with the Pex13/Pex14 docking complex, but its exact role is not well understood (Huhse *et al*, 1998; Mast *et al*, 2010).

After docking to the so-called importomer complex, cargo is released and imported into peroxisomes through an unknown mechanism. Pex8, a membrane-associated protein in the peroxisomal lumen has been implicated in this process in yeast (Stanley & Wilmanns, 2006). Once cargo has been released, Pex5 undergoes ubiquitination and exits the membrane in a process facilitated by the exportomer (Pex1/Pex6/Pex15 in yeast) in an ATP-dependent manner (Platta *et al*, 2014). Depending on the number of ubiquitinations events, Pex5 can either be recycled to the cytosol (mono-ubiquitination) or degraded by the proteasome (poly-ubiquitination) (Platta *et al*, 2014; Smith & Aitchison, 2013; Kiel *et al*, 2005).

Conversely, the PTS2 pathway utilises Pex7 as the main receptor for cargo proteins (Rehling *et al*, 1996), but presents higher variability with regards to the co-receptors required. Although Pex5 alone is enough for the recognition and transport of PTS1 proteins to the docking complex, in the case of PTS2 pathway, Pex7 cooperates with Pex18 and Pex21 in yeast or the orthologous Pex20 in other fungal species (Schliebs & Kunau, 2006). These co-receptors are required for targeting and stabilisation of the receptor-cargo complexes. Although Pex7 is functionally similar to Pex5, it lacks the WxxxF motifs that would enable its interaction with the docking complex. For that reason, in the PTS2 pathway, the Pex7 binding co-receptors contain the WxxxF motifs instead, which are required for Pex14 binding and protein import in an otherwise similar manner to the PTS1 pathway (Stein *et al*, 2002; Purdue *et al*, 1998; Otera *et al*, 2000).

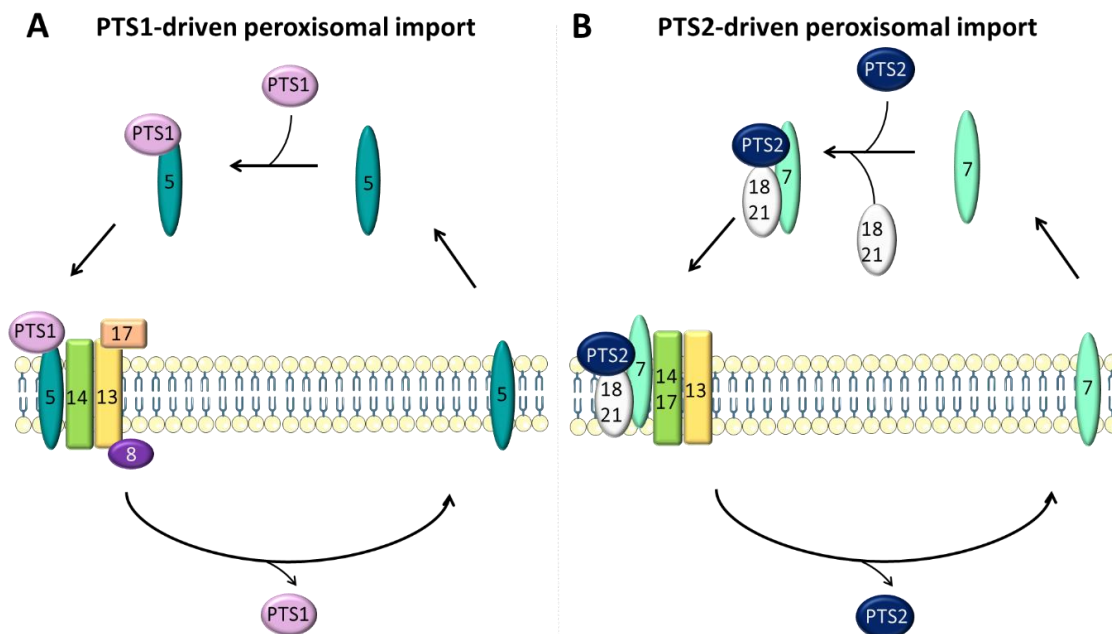


Figure 6: Peroxisomal matrix protein import. (A) PTS1-driven import. Pex5 recognises PTS1 cargo in the cytosol, and transports them to the docking complex (Pex13p, Pex14p, and Pex17p) at the peroxisomal membrane. Assembly of the cargo-loaded Pex5p with the docking complex results in the formation of a transient pore and the cargo is translocated into the peroxisomal lumen in an unknown manner, possibly remaining associated with its receptor. The receptor-cargo complex dissociates and the cargo is released, through a process that possibly involves Pex8. (B) PTS2-driven import. The receptor Pex7 recognises PTS2-cargoes and cooperates with Pex18 or Pex21 for import to peroxisomes, where a complex similar to the PTS1 importomer is located.

1.1.5.1. Peroxisomal targeting signal receptor 5 (Pex5)

Pex5 is fundamental factor for peroxisome biogenesis, which can localise both in the cytosol and the peroxisomal membrane (Dodt *et al*, 1995). It is characterised by an intrinsically disordered N-terminal domain (Carvalho *et al*, 2006; Neuhaus *et al*, 2014; Shiozawa *et al*, 2009; Su *et al*, 2009) and a globular C-terminal domain that contains a series of TPR subdomains used to recognise and bind PTS1 cargoes (Neufeld *et al*, 2009; Neuhaus *et al*, 2014) (**Figure 7**).

The N-terminal domain of Pex5 contains a number of diaromatic, partly helical WxxxF/Y motifs (Neufeld *et al*, 2009), with the *Saccharomyces cerevisiae* Pex5p containing three and the human PEX5 containing seven as well as a novel LVAEF motif. All of these bind to the N-terminal region of Pex14, an interaction crucial to the docking of cargo-loaded Pex5 onto the peroxisomal membrane (Saidowsky *et al*, 2001; Neufeld *et al*, 2009; Neuhaus *et al*, 2014; Platta *et al*, 2013; Kim & Hettema, 2015). Conversely, the Pex5 C-terminus, is structured and contains a series of TPR domains followed by a helical bundle (Stanley *et al*, 2006). In human, the C-terminus is

folded into two TPR clusters separated by TPR-4 (Dodt *et al*, 1995). This ring-like conformation forms the PTS1 binding groove, where conserved amino acids are responsible for its recognition and binding.

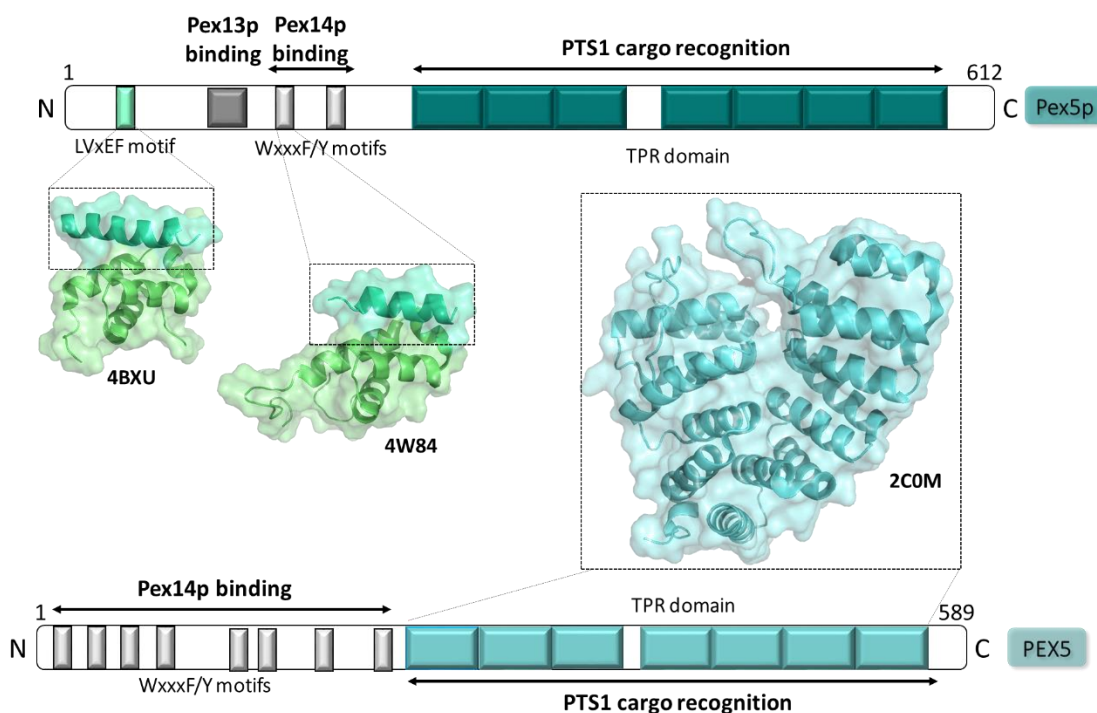


Figure 7: Structure of yeast *Pex5p* and human *PEX5*. The TPR domains are coloured in teal, while the *WxxxF/Y* motifs in grey and the *LVxEF* motif in cyan. Available NMR and crystal structures are indicated.

1.1.5.2. Peroxisomal membrane protein 14 (*Pex14*)

Depending on the organism, *Pex14* is characterised as an intrinsic membrane- or membrane-associated peroxisomal protein that constitutes one of the basic components of the peroxisomal importomer machinery. *Pex14* contains three major structural and functional domains: an N-terminal helical domain, a hydrophobic -putative transmembrane- domain, and a coiled-coil region (**Figure 8**) (Albertini *et al*, 1997; Shimizu *et al*, 1999; Will *et al*, 1999; Shimozawa *et al*, 2004). Mammalian *Pex14* is an integral membrane protein with its C-terminus exposed in the cytosol, where the coiled-coil domain offers potential for homo-oligomerisation (Will *et al*, 1999; Otera *et al*, 2002; Itoh & Fujiki, 2006).

The conserved N-terminus of Pex14 comprises of 60 residues and has been shown to mediate binding to various other peroxisomal proteins, including Pex5 and Pex19, as well as facilitating non-peroxisomal interactions such as binding to tubulin (Neufeld *et al*, 2009; Neuhaus *et al*, 2014; Bharti *et al*, 2011a). It forms a helical bundle which has been shown to bind competitively either Pex5 and Pex19 in an antiparallel orientation (Neufeld *et al*, 2009) (**Figure 8**). The binding between Pex14 and these two proteins depends upon a WxxxF and a LVxAF motif found in Pex5, or an F/FxxxF motif located at the N-terminus of Pex19 (Neufeld *et al*, 2009; Neuhaus *et al*, 2014). Additional to the aforementioned motifs, Pex14 also contains the Pex13 binding motif PxxxP (Fransen *et al*, 2004; Neufeld *et al*, 2009), which is followed by a patch of hydrophobic amino acids important for its association with the peroxisomal membrane.

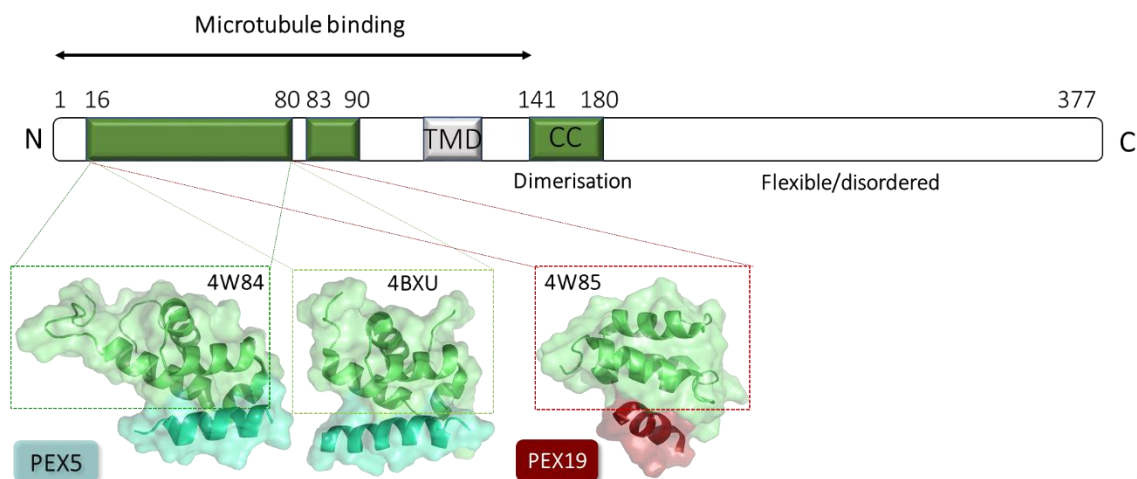


Figure 8: Schematic overview of human PEX14. The conserved N-terminus (residues 16-80), the PEX13 binding site (83-90) and the coiled-coil (CC) region are indicated in green. The putative transmembrane domain (TMD) is coloured grey and the available NMR structure of PEX14-PEX5 and PEX14-PEX19 complexes are shown.

1.1.5.3. Peroxisomal cargo proteins

Peroxisomal coenzyme-A synthetase (Pcs60) from yeast

Pcs60 is a 60 kDa peroxisomal protein from *Saccharomyces cerevisiae* which belongs to the family of AMP-binding proteins. It contains a C-terminal PTS1 –SKL tripeptide motif that is required for its targeting to peroxisomes *via* Pex5p. Expression of Pcs60 is induced under oleic acid growth conditions, upon which, peroxisome localisation is observed (Blobel & Erdmann,

1996). Pcs60 is involved in the first step of oxalic acid metabolism, converting oxalate to oxalyl-CoA (Foster & Nakata, 2014):



Oxalic acid is a simple dicarboxylic acid which is found in the majority of organisms and can be beneficial for nutrient uptake, tolerance to metals, and pathogenicity amongst other roles (Foster & Nakata, 2014). However, despite these advantages, high concentrations of oxalic acid can cause several problems for the cell, including disruption of cellular membranes, mitochondrial metabolism, and free radical formation. For that reason, Pcs60 is an essential enzyme for the regulation of oxalic acid levels and protecting the cell from its harmful effects.

Pcs60 has been shown to interact with Pex5p in a PTS1-dependent manner. Deletion of the SKL motif inhibits Psc60 import to peroxisomes, but Δ SKL Psc60 still exhibits low binding affinity for Pex5p, indicating that additional sites may be required for this interaction (Hagen *et al*, 2015), also referred to as “extended PTS1” (Fodor *et al*, 2012).

Microphage migration inhibitory factor (MIF1) from Arabidopsis thaliana

MIF1 is a pro-inflammatory protein released by immune cells in response to stress in human, displaying dopachrome tautomerase/isomerase activity (Calandra & Roger, 2003) In *Arabidopsis thaliana*, three homologues of MIF1 have been identified, with MIF1 carrying an SKL PTS1 motif that enables its localisation to peroxisomes.

MIF1 belongs to the family of isomerases and is involved in the metabolism of the phytotoxic compound L-dopachrome to 5,6-dihydroxyindole-2-carboxylic acid (DHICA). It exhibits stress-inducible transcript accumulation in accordance with findings in *Arabidopsis*, which indicate that L-dopachrome alters the expression of biotic and abiotic stress response genes (Golisz *et al*, 2011). Based on studies on human MIF1, the catalytic activity of the enzyme relies on an N-terminal Pro-1 residue (after cleavage of Met-1), which is highly conserved amongst many species and is required for its catalytic activity (Merk *et al*, 2012).

1.2. Aims of this thesis

This doctoral work aims to provide structural insight into peroxisomal receptor-cargo complexes, involved in peroxisomal membrane protein import (mediated by the receptor Pex19), and peroxisomal matrix protein import (regulated by Pex5). To obtain information about the biological function and structure of these complexes, as well as to elucidate the mechanistic aspects of complex formation, a variety of structural biology methods will be employed, including X-ray crystallography, small angle X-ray scattering (SAXS), electron microscopy (EM), and crosslinking coupled with mass spectrometry (XL-MS).

The following questions will be addressed with regards to the interaction of human PEX19 with membrane proteins PEX3 and PEX14:

- Are there auxiliary binding sites between human PEX3 and PEX19 and what is their biological relevance?
- How does full-length PEX19 behave in solution as a monomer, and as a complex with PEX3?
- What is the role of the interaction of PEX19 and PEX14 and what are the structural characteristics of the full-length PEX19-PEX14 complex?

In the second part of this thesis, which concerns peroxisomal matrix protein import, we will apply a hybrid structural approach to answer the following questions:

- What is the structure of peroxisomal matrix proteins, such as Pcs60 and MIF1?
- How do they interact with the peroxisomal receptor Pex5p and what are the structural characteristics of these complexes?
- How do multimeric cargoes become imported through the peroxisomal membrane?

Answering these questions will aid in producing more detailed models for peroxisomal membrane and matrix protein import, processes that are essential for peroxisome biogenesis and maturation, and consequently survival of the cell.

2. Materials and methods

2.1. Buffers and reagents

All chemicals used in this study were purchased from ROTH and Sigma-Aldrich. Reagents for molecular biology were acquired from Fermentas (Thermo Fisher Scientific/ Fermentas) and New England BioLabs (NEB). The following table contains a summary of all the chemicals and reagents used (**Table 3**).

Table 3: List of buffers and reagents.

Bacterial strains	
DH5 α	F- ϕ 80lacZ Δ M15 Δ (lacZYA-argF) U169 recA1 endA1 hsdR17(rk-, mk+) phoA supE44 thi-1 gyrA96 relA λ -
BL21-CodonPlus (DE3) RIL	<i>E. coli</i> B F- <i>ompT hsdS</i> (r - m -) <i>dcm</i> + Tetr gal λ (DE3) <i>endA Hte</i> [<i>argU ileY BB leuW Camr</i>]
Cloning reagents	
6x DNA loading dye (Thermo Scientific R0611)	10mM Tris-HCl (pH 7.6), 0.03% (w/v) bromophenol blue, 0.03% xylene cyanol FF, 60% (v/v) glycerol, 60mM EDTA
TAE (Tris-Acetate-EDTA) buffer	40mM Tris (pH 7.6), 20mM acetic acid, 1mM EDTA
LB Agar	1% (w/v) peptone, 1% (w/v) NaCl, 0.5% (w/v) yeast extract, 1.5% (w/v) agar
SOC medium	2% (w/V) peptone, 0.5% (w/v) yeast extract, 10mM NaCl, 2.5mM KCl, 10mM MgCl ₂ , 10mM MgSO ₄ , 20mM glucose
Expression and purification reagents and buffers	
Auto-induction medium	Basic medium: 1% w/v tryptone/peptone, 0.5% w/v yeast extract 25mM Na ₂ HPO ₄ ·7H ₂ O, 25mM KHPO ₄ , 50mM NH ₄ Cl, 5mM Na ₂ SO ₄ , 2mM MgSO ₄ ·xH ₂ O, 0.05% w/v glucose, 0.2% w/v lactose, 0.5% w/v 87% glycerol Trace elements (1000x): 50mM FeCl ₃ , 20mM CaCl ₂ , 10mM MnCl ₂ , 10mM ZnSO ₄ , 2mM CoCl ₂ , 2mM CuCl ₂ , 2mM NiCl ₂ , 2mM Na ₂ MoO ₄ , 2mM Na ₂ SeO ₃ , 2mM H ₃ BO ₃
Lysogeny Broth (LB) medium	1 % (w/v) peptone, 1 % (w/v) NaCl, 0.5 % (w/v) yeast extract
Ni ²⁺ affinity purification buffers	Lysis buffer: 50mM HEPES pH 7.5, 150mM NaCl, 20mM Imidazole Wash buffer: 50mM HEPES pH7.5, 1M NaCl, 20mM Imidazole Elution buffer: 50mM HEPES pH7.5, 150mM NaCl, 500mM Imidazole
GST purification buffers	Lysis-GST buffer (D): 50mM HEPES pH7.5, 150mM NaCl Elution-GST buffer (E): 50mM HEPES pH7.5, 150mM NaCl, 20mM reduced glutathione
SEC buffers	SEC1: 50mM TRIS pH7.5, 150mM NaCl, 1mM TCEP SEC2: 50mM HEPES pH7.5, 150mM NaCl, 1mM TCEP

Protein analysis	
SDS PAGE 6x Loading dye	350mM TRIS pH6.8, 60% v/v glycerol, 40mM DTT, 10% w/v SDS, 0.06% w/v bromophenol blue
12% (w/v) SDS gels	350mM BisTris pH 6.8, 12% v/v acrylamide, 0.02% w/v ammonium persulfate (APS), 1% v/v tetramethylethylenediamine (TEMED) Stacking gel: 350mM BisTris pH 6.8, 5% v/v acrylamide, 0.02% w/v ammonium persulfate (APS), 1% v/v tetramethylethylenediamine (TEMED)
20x SDS running buffer	1M MES pH 7.3, 1M TRIS-base, 2% w/v SDS, 20mM EDTA
Coomassie staining solution	30 % (v/v) ethanol, 10 % (v/v) acetic acid, 0.25 % (w/v) Coomassie Brilliant Blue G-250
Coomassie destaining solution	30 % (v/v) ethanol, 10 % (v/v) acetic acid
Coomassie staining solution	Commercial InstantBlue Protein Stain (Expedeon): 1L reagent, containing Coomassie dye, ethanol, phosphoric acid and solubilizing agents in water
Crosslinking	
BS3 crosslinking buffer	50mM HEPES pH7.5, 150mM NaCl
EDC/sulfo-NHS activation buffer	0.1M MES, 0.5M NaCl, pH 6.0
EDC/sulfo-NHS coupling buffer	Phosphate-buffered saline (PBS)
Quenching buffer	50mM TRIS pH7.5, 150mM NaCl
Additional methods	
Circular dichroism (CD) spectroscopy buffer	50mM NaH ₂ PO ₄ , 100mM NaF
Fluorescence anisotropy (FA) buffer	50mM Tris-HCl, pH7.5, 250mM NaCl, 0.5mM TCEP, 0.67 mg.ml ⁻¹ BSA

2.2. Molecular biology techniques

2.2.1. Cloning

All enzymes used were purchased from NEB. DNA purification and plasmid extraction were performed using kits from Promega and QIAGEN respectively, while peptides and DNA were synthesised by GenScript.

The plasmids used are summarised in **Table 4**.

Table 4: List of constructs used in this study

Construct Name	Organism	Vector	Purification Tag	Protein Boundaries	Cloning
PEX3 _{C235S(41-373)}	Human	pET-32a (Novagen)	N-terminal Trx-His6-tag, TEV cleavage site	41-373	Courtesy of Professor G. Dodt (University of Tübingen)
PEX19 ₍₁₋₂₉₉₎	Human	pETM-11 (G. Stier, EMBL Heidelberg)	N-terminal His6-tag, TEV cleavage site	1-299	Pre-existing (Schueller et al, 2010) Forward:CACTGCGTCTCCCAT GGCCGCCGCTGAGGAAGG Reverse:CGGGATCCTTATCAC ATGATCAGACACTGTTCACC
PEX19 ₍₅₄₋₂₉₁₎	Human	pETM-11 (G. Stier, EMBL Heidelberg)	N-terminal His6-tag, TEV cleavage site	54-291	Pre-existing construct (N. Schueller, EMBL-Hamburg)
PEX19 ₍₁₆₁₋₂₈₃₎	Human	pETM-11 (G. Stier, EMBL Heidelberg)	N-terminal His6-tag, TEV cleavage site	161-283	Pre-existing (Schueller et al, 2010) Forward:CAGATCCATGGGCA TGGACGAAGGGGATGGG Reverse:CGGGATCCTTAGAG GGCATCCAGGTCAAAGTT
PEX19 _{Δ(62-148)}	Human	pETM-11 (G. Stier, EMBL Heidelberg)	N-terminal His6-tag, TEV cleavage site	1-299 substitution of 62-148 with HEWL linker (GGGGSLVPRGSGGGGS)	Synthesised by GenScript Recloned into pETM-11 using restriction enzymes
PEX14 ₍₁₋₃₇₇₎	Human	pnEA-vHis (C. Romier)	N-terminal His6-tag, TEV cleavage site	1-377	Pre-existing (D. Passon): Forward:GATATACATATGGC GTCC TCG Reverse:CGCCGCAGATCTCTA CTAGTCCCGCTC
PEX14 ₍₁₆₋₈₀₎	Human	pETM-11 (G. Stier, EMBL Heidelberg)	N-terminal His6-tag, TEV cleavage site	16-80	Pre-existing (Neufeld et al, 2009)
Pex5p ₍₁₋₆₁₂₎	<i>Saccharomyces cerevisiae</i>	pETM-30(G. Stier, EMBL Heidelberg)	N-terminal His6-tag, N-terminal GST tag, TEV cleavage site	1-612	Pre-existing (D. Passon): Forward:GCTCATATGGACGT AGGAAGTTGCTCA Reverse:GCTGGATCCAAACG AAAATTCTCC
Pex5p ₍₁₉₈₋₆₁₂₎	<i>Saccharomyces cerevisiae</i>	pETM-33(G. Stier, EMBL Heidelberg)	N-terminal His6-tag, N-terminal GST tag, TEV cleavage site	198-612	Pre-existing (D. Passon): Forward:GCTCATATGGAGCA AGAACAACAACCCT Reverse:GCTGGATCCAAACG AAAATTCTCC
Pex5p ₍₃₁₂₋₆₁₂₎	<i>Saccharomyces cerevisiae</i>	pETM-33(G. Stier, EMBL Heidelberg)	N-terminal His6-tag, N-terminal GST tag, TEV cleavage site	312-612	Pre-existing (D. Passon): Forward:GCGCCCATGGAAAA TCCTAATGCTTATAAAATT Reverse:GCGCGAATTCTCAA AACGAAAATTCTCCTTAA
Pcs60	<i>Saccharomyces cerevisiae</i>	pGEX-4T-3 (GE-Healthcare)	N-terminal GST-tag, thrombin cleavage site	1-543	Pre-existing (Hagen et al, 2015) Forward:GATCGGATCCATGA CAAGTGCCGCTACTG Reverse:GATCCTCGAGCTACA ACTTACTCTTATTTCTGCTG
AtMIF1	<i>Arabidopsis thaliana</i>	pQE-31 (QIAGEN)	N-terminal His6-tag	1-112	Cloned by D. Crappe (University of Stavanger)

Cloning was performed using polymerase chain reaction (PCR) to amplify a specific DNA sequence (Mullis *et al*, 1986) into a vector of choice. The PCR product contains distinct restriction enzyme cleavage sites on either end created by the primers used. The size of the PCR product is verified by agarose gel electrophoresis and the purified DNA is ligated with a bacterial vector containing the same restriction sites. The ligated plasmid is then used to transform *E. coli* DH5 α cells and is selected based on the antibiotic resistance gene included in the vector. To further validate the clones, the plasmids are purified and subjected to restriction by endonucleases and sequencing.

2.2.1.1. *Restriction cloning, DNA isolation and agarose gel electrophoresis*

PEX19 $\Delta_{(62-148)}$ was synthesised and codon optimised by GenScript and was initially cloned into vector pUC57 using EcoRV sites. Additional nucleotides were added, providing spacers and sites for the restriction enzymes NcoI upstream of the gene and NotI downstream. In order to obtain DNA sequences suitable for ligation, the high fidelity enzymes NcoI-HF (NEB: R3193S) and NotI-HF (NEB: R3189S) were used to digest the synthesised plasmid and the pETM-11 vector, according to the following reaction:

PEX19 $\Delta_{(62-148)}$ - pUC57 / pETM-11	1 μ g
NcoI-HF	1 μ l
NotI-HF	1 μ l
10x NEB CutSmart® buffer	5 μ l
Nuclease-free water	Up to 50 μ l final volume
Incubation time	1 hour
Incubation temperature	37°C

The digested DNA fragments were analysed by agarose gel electrophoresis. The reaction was loaded onto a 1% w/v agarose TAE gel and were submitted to electrical current for 40 minutes at 120V (Mupid One, ADVANCE). The separation of DNA molecules of different size relies on the negative charge of DNA which allows it to migrate towards the positive electrode, when subjected to an electric field. The separated bands corresponding to the PEX19 $\Delta_{(62-148)}$ insert and the linearized pETM-11 vector were cut from the agarose gel and purified using the Wizard SV Gel and PCR Clean-Up System (Promega) according to the manufacturer's instructions.

The isolated DNA fragments were then ligated using the NEB T4 DNA ligase in the following reaction:

PEX19 Δ (62-148) insert	40 ng
pETM-11 vector	50 ng
T4 DNA Ligase	2 μ l
T4 DNA Ligase Buffer (10X)	1 μ l
Nuclease -free water	Up to 10 μ l final volume
Incubation time	2 hours
Incubation temperature	Room temperature

2.2.1.2. *Bacterial transformation and clone verification*

The ligation reaction products were used to transform 50 μ l of competent *Escherichia coli* (*E.coli*) DH5 α cells under sterile conditions using the heat-shock approach. The DNA-cell mixture was initially incubated in ice for 30 minutes and submitted to heat-shock by placing into a 42°C water bath (Aqualine, Roth) for 50 seconds. After placing briefly in ice, the cells were subsequently mixed with 200 μ l of SOC medium and incubated at 37°C in a shaking incubator (New BrunswickTM Innova® 42R) to ensure optimal aeration. After 45 minutes, the cells were plated onto LB agar plates containing the appropriate antibiotic of choice. In the case of pETM-11, which contains a kanamycin resistance gene, the LB agar plates contained 50 μ g/ml of kanamycin for selection of correct clones. The plated cells were incubated at 37°C for 12 hours and the colonies that appeared were tested for correct DNA fragment insertion. From the plate, 10 colonies were selected, inoculated into 10 ml of LB medium (containing 50 μ g/ml kanamycin) and grew for 8 hours at 37°C in a shaking incubator. The cells were harvested by centrifugation at 3,100g for 10 minutes (5810R, Eppendorf) and the plasmids were isolated using the QIAprep Spin Miniprep Kit (QIAGEN), following the manufacturer's instructions. To verify the correct DNA fragment insertion into the vector, a diagnostic restriction digest was performed for the purified plasmids in the following reaction:

Plasmid DNA	100 ng
NcoI-HF	1 µl
NotI-HF	1 µl
10x NEB CutSmart® buffer	1 µl
Nuclease-free water	Up to 10 µl final volume
Incubation time	1 hour
Incubation temperature	37°C

The digested DNA was analysed by agarose gel electrophoresis as previously described (Chapter 2.2.1.1). The fragments that appeared at the expected size were further analysed by Sanger sequencing (MWG Eurofins) using T7 promoter (5'TAATACGACTCACTATAGGG 3') and T7 terminator (5'CTAGTTATTGCTCAGCGGT 3') primers.

2.2.2. Protein expression and purification

For the expression and purification of proteins, all chemicals were purchased by Roth and Sigma Aldrich, unless otherwise stated. Ni-NTA agarose (Nickel-nitrilotriacetic acid) resin came from QIAGEN, Glutathione sepharose from GE Healthcare and affinity purification was performed at room temperature. All size exclusion chromatography columns were purchased by GE Healthcare and size exclusion chromatography was performed at 10°C by an ÄKTA Pure system (GE Healthcare) using the Unicorn 7.0 software.

2.2.2.1. Protein expression and cell lysis

E. coli is a very commonly used organism of choice for recombinant protein production. It can present several advantages, such as fast growth kinetics, inexpensive growth media, possibility to achieve high cell density cultures, as well as fast and easy transformation with exogenous DNA. For this study, the *E. coli* strain BL21 (DE3) CodonPlus-RIL (Stratagene/ Agilent Technologies) was used for protein production for all constructs. This strain contains extra copies of the *argU*, *ileY*, and *leuW* tRNA genes, encoding tRNAs that recognise the rare arginine codons AGA and AGG, the isoleucine codon AUA, and the leucine codon CUA, respectively.

All plasmids used contain the gene of interest cloned downstream of a T7 promoter, a lac operator (*lacO*), and a lac repressor sequence (*lacI*), which are responsible for tuning the

expression depending on metabolites present in the cell. Usual procedures of overexpression make use of the ability of IPTG (Isopropyl β -D-1-thiogalactopyranoside) to block the *lacI* repressor, which in turn would lead to activation of transcription. In this study, instead of IPTG-induced overexpression of proteins, the alternative method of autoinduction was used (Studier, 2005). This method is based on the ability of certain media to induce protein expression when cells reach saturation and is the result of different metabolism states of the bacteria. As a principle, autoinduction is based on carbon sources present in the media that are metabolised differentially to promote high density cell growth and automatic induction of expression. In this method, a limited concentration of glucose is initially metabolised during growth, preventing lactose uptake until glucose is depleted. Eventually, lactose can be absorbed and metabolised to the inducer allolactose, which can release the lac repressor and induce expression of T7 RNA polymerase, allowing activation of transcription.

Practically, for protein expression, the same protocol was followed for all constructs. Initially 100 ng of plasmid DNA were used to transform 50 μ l of *E. coli* BL21 (DE3) CodonPlus-RIL cells by means of the heat-shock protocol (Chapter 2.2.1.2). The cells are plated onto LB-agar plates containing the appropriate antibiotics (100 μ g/ml Ampicillin, 25 μ g/ml Chloramphenicol, 50 μ g/ml Kanamycin) and are incubated for 12 hours until colonies appear. Next, a small volume of 10 ml of LB with the appropriate antibiotics is inoculated with single colonies and is incubated for 8 hours at 37°C in a shaking incubator. From this pre-culture, 2 ml were used to inoculate 500 ml of freshly-made autoinduction media which grew first at 37°C for 4 hours in a shaking incubator at 142 rpm (New Brunswick™ Innova® 44R) and then at 20°C overnight (12-14 hours).

Cell harvesting was performed by centrifugation at 9,300g for 20 minutes (J-LITE® JLA-8.1000 rotor, Avanti® JXN-26 centrifuge, Beckman-Coulter) and the cell pellet was re-suspended in lysis buffer (**Table 3**), following the addition of DNase I (#DN-25) and protein inhibitor mix (#39106, Serva) at 1 ml per 100 ml of cell suspension.

Cell lysis was carried out by means of sonication (SonoPuls HD3200, Bandelin) using a number of short pulses (5-10 sec) with pauses (10-30 sec) to ensure that the cell suspension does not overheat. A following step of centrifugation (JA-25.50 rotor, Avanti® JXN-26 centrifuge,

Beckman-Coulter) and subsequent filtration (Millex-HV Filter, 0.45 μm , Merck) ensures that cell debris are removed and the supernatant contains the cell lysate.

2.2.2.2. *Affinity chromatography protein purification*

After cell lysis, the overexpressed protein is subjected to the first purification step by affinity chromatography. In this study, the PEX19, PEX3, PEX14 and *AtMIF1* constructs contain N-terminal His6-tags, while the Pex5p and Pcs60 constructs contain N-terminal GST fusion tags, so they were purified by means of Ni-NTA agarose or Glutathione sepharose resin respectively. All steps were performed at room temperature at the bench.

Ni-NTA metal ion affinity chromatography (IMAC)

Ni-NTA purification is based on the selectivity of the Ni-NTA resin for proteins containing 6 or more histidine residues on their N- or C- termini. The cell lysate is loaded onto 5 ml of equilibrated with lysis buffer Ni-NTA agarose resin in a gravity flow set up and is incubated for 1 hour at 4°C. The following washing steps (buffers summarised in **Table 3**) are performed and the protein is eluted after application of high imidazole concentration:

Lysis buffer wash I	50 ml
Wash buffer wash	50 ml
Lysis buffer wash II	50 ml
Elution	25 ml

Subsequently, the His6-tag is removed after proteolytic cleavage using His6-TEV protease (1 mg to 40 mg of target protein). Cleavage was performed overnight in a dialysis membrane tube (Dry Spectra/Por[®] Dialysis Tubing, MWCO 3.5 kDa, SpectrumLabs) at 4°C while dialysing against 1 litre of SEC buffer. An additional step of Ni-NTA purification is required in order to remove the cleaved tag, TEV and other bound contaminant proteins.

Glutathione S- transferase affinity chromatography

Glutathione is a tripeptide (Glu-Cys-Gly) that is the specific substrate for glutathione S-transferase (GST), a 211 amino acid protein (26kDa). Similar to Ni-NTA purification, cell lysate is loaded onto 5 ml of glutathione sepharose resin, equilibrated with lysis-GST buffer (**Table 3**) and is incubated for 1 hour at 4°C. After discarding the flow-through, the resin is washed with 50 ml of lysis-GST buffer and on-column cleavage of the tag is performed by addition of 3C protease (1 mg to 100 mg of target protein) in the case of the Pex5p constructs or thrombin (#T6884, Sigma-Aldrich) in the case of Pcs60 (1 mg to 500 mg of target protein). Cleavage was carried out overnight at 4°C, resulting in the cleaved protein eluting, while the GST tag remains bound to the resin and can be removed after washing with 50ml of elution-GST buffer.

2.2.2.3. *Size-exclusion chromatography purification (SEC)*

SEC can be used to separate protein molecules according to differences in the hydrodynamic radius of the proteins, as they pass through a SEC column. The stationary phase of a SEC column is composed of a porous matrix of spherical particles, with the pore size changing depending on the material used. Proteins are eluted isocratically, with larger molecules emerging first. As proteins can absorb at 280 nm, it is possible to use this property to monitor the amount of protein purified.

In this project, the proteins obtained after affinity chromatography were concentrated (Spin-X UF, Corning) by centrifugation at 3,100g and further purified by SEC, after equilibrating the respective columns with the SEC buffer (SEC1 buffer for all constructs except for Pcs60 which was purified in SEC2 buffer) (**Table 3**).

The following SEC steps were performed for each protein:

Table 5: SEC columns used for each construct.

PEX3 _{C235S(41-373)}	Hiprep Superdex 200 16/60 GL
PEX19 ₍₁₋₂₉₉₎	Hiprep Superdex 200 16/60 GL
PEX19 ₍₁₋₂₉₉₎ - PEX3 _{C235S(41-373)}	Superdex 200 increase 10/300GL
PEX19 ₍₅₄₋₂₉₁₎	Hiprep Superdex 200 16/60 GL
PEX19 ₍₁₆₁₋₂₈₃₎	Hiprep Superdex 200 16/60 GL
PEX19 _{Δ(62-148)}	Hiprep Superdex 200 16/60 GL
PEX19 ₍₁₋₂₉₉₎ - PEX14 ₍₁₋₃₇₇₎	Hiprep Superdex 200 16/60 GL
	Superose 6 increase 10/300 GL
PEX14 ₍₁₆₋₈₀₎	Hiprep Superdex 75 16/60 GL
Pex5p ₍₁₋₆₁₂₎	Hiprep Superdex 200 16/60 GL
Pex5p ₍₁₉₈₋₆₁₂₎	Hiprep Superdex 200 16/60 GL
Pex5p ₍₃₁₂₋₆₁₂₎	Hiprep Superdex 200 16/60 GL
Pcs60 ₍₁₋₅₄₃₎	Hiprep Superdex 200 16/60 GL
	Superdex 200 increase 10/300GL
Pex5p ₍₁₋₆₁₂₎ - Pcs60 ₍₁₋₅₄₃₎	Superdex 200 increase 10/300GL
Pex5p ₍₁₉₈₋₆₁₂₎ - Pcs60 ₍₁₋₅₄₃₎	Superdex 200 increase 10/300GL
Pex5p ₍₃₁₂₋₆₁₂₎ - Pcs60 ₍₁₋₅₄₃₎	Superdex 200 increase 10/300GL
AtMIF1 ₍₁₋₁₁₂₎	Hiprep Superdex 75 16/60 GL

2.2.2.1. Protein analysis and quantification

The purified proteins are analysed qualitatively on the basis of their purity by means of SDS-PAGE (discontinuous sodium dodecyl-sulfate polyacrylamide gel electrophoresis). The proteins are denatured by incubating with SDS-containing 1x SDS-PAGE loading dye (**Table 3**) and heating at 75°C for 5 minutes. A protein chain can contain SDS proportional to its relative molecular mass. The negative charges of SDS are strongly attracted towards a positively charged electrode in an electric field and proteins can be separated as the porous polyacrylamide gels restrain larger molecules from migrating as fast as smaller molecules.

For this project, the gels used were composed of a 12% separating gel and a 5% stacking gel and were loaded with 1-5 mg/ml of sample. Electrophoresis was performed using 1x SDS running buffer in room temperature at 200 V for 40 minutes, using the Mini-Protean III Tetracel system (Bio-Rad). The gels were stained using the commercial Coomassie stain, InstantBlue (Expedeon). Quantitative analysis of the purified proteins can be performed by measuring their absorbance at 280nm using a NanoDrop ND-1000 spectrophotometer (PeqLab). Proteins that contain Trp, Tyr residues or Cys-Cys disulphide bonds generally absorb in the UV range, enabling quantification by absorbance spectroscopy. Using the ProtParam tool (Gasteiger *et al*, 2005) it is possible to calculate the molar extinction coefficient of a protein molecule based on its amino acid sequence (**Table 6**).

Table 6: Extinction coefficients of constructs used in this study, as calculated by the ProtParam tool

Construct Name	Extinction Coefficient (M ⁻¹ Cm ⁻¹)
PEX3 _{C235S(41-373)}	23295
PEX19 ₍₁₋₂₉₉₎	10220
PEX19 ₍₅₄₋₂₉₁₎	10095
PEX19 ₍₁₆₁₋₂₈₃₎	10095
PEX19 _{Δ(62-148)}	10220
PEX14 ₍₁₋₃₇₇₎	23950
PEX14 ₍₁₆₋₈₀₎	no Trp, Tyr or Cys
Pex5p ₍₁₋₆₁₂₎	71195
Pex5p ₍₁₉₈₋₆₁₂₎	55600
Pex5p ₍₃₁₂₋₆₁₂₎	35660
Pcs60 ₍₁₋₅₄₃₎	38070
AfMIF1 ₍₁₋₁₁₂₎	4470

2.2.2.2. *Blue native PAGE*

In contrast with SDS-PAGE, in the case of blue native PAGE the proteins are analysed in their native state, yielding information about their oligomeric state and overall size in non-denatured conditions. The separation principle relies on binding of Coomassie blue G250 that provides negative charges to the surface of the protein. During migration to the anode, protein complexes are separated according to molecular mass and/or size and high resolution is obtained by the decreasing pore size of a polyacrylamide gradient gel.

Blue native PAGE was employed for the characterisation of peroxisomal complexes. The protein samples were prepared with 1x NativePAGE sample buffer (Novex, Life Technologies) and loaded on 4-16% w/v Bis-Tris gels (Life Technologies). The electrophoresis is performed at 4°C with a voltage of 150V, using 1x Native PAGE Dark Blue Cathode buffer (Life technologies) and 1x Native PAGE Anode buffer (Life Technologies). The gels were initially treated with Coomassie destaining solution to remove excess Coomassie stain and re-stained using Coomassie staining solution (**Table 3**).

2.3. Biophysical characterisation of proteins

In order to obtain information and characterise in detail the overall shape, size and oligomerisation state, complementary biophysical techniques, such as Circular Dichroism Spectroscopy (CD) and Size Exclusion Chromatography- Light Scattering techniques were employed.

2.3.1. Circular dichroism spectroscopy (CD)

Circular dichroism refers to the property of chiral molecules to absorb right- and left- circularly polarised light differentially. The secondary structure elements of proteins, such as α -helices, β -strands and random coil can produce characteristic CD spectra that exhibit specific minima or maxima of ellipticity at different wavelengths in a recorded spectrum. These features can be used to estimate the secondary structure content of a protein.

For this project, proteins with 10 mM concentration in CD buffer (**Table 3**) were analysed by CD spectroscopy using a Chirascan CD Spectrometer (Applied Photophysics) and analysis of the secondary structure was performed using the accompanying software provided by Applied Photophysics.

2.3.2. Thermofluor differential scanning fluorimetry

This method is commonly used for assessment of favourable buffer conditions for protein stability. It relies on the ability of the SYPRO Orange dye (Thermo Fischer Scientific) to bind to hydrophobic patches/denatured protein/molten globules and emit fluorescence. As the temperature is increased and the protein unfolds, it is possible to monitor the increase in fluorescence and determine a melting temperature. The temperature and fluorescence monitoring are done using a qPCR machine (MyIQ RT-PCR, BioRad). The fluorescence signal is then used to determine the protein melting point (T_m).

2.3.3. Size exclusion chromatography-light scattering techniques

Size exclusion chromatography coupled with light scattering techniques is a powerful method for the absolute determination of molar mass and root mean square radius in solution. As the name of the method suggests, the first part of the procedure includes a fractional elution of a protein according to its molecular weight. Subsequently, the eluted protein is analysed by its scattering properties using either Right Angle Light Scattering (RALS), which uses one scattering angle of 90° to characterise the protein, or the more powerful Multiple Angle Light Scattering (MALS), which uses several angles. The most widespread application of MALS is as an absolute molecular weight detector in conjunction with a concentration detector (measuring refractive index, RI).

In this study, we used the analytical HPLC Wyatt SEC-MALS system (Wyatt) which utilises the ASTRA software for characterisation of the PEX19₍₁₋₂₉₉₎-PEX14₍₁₋₃₇₇₎ complex. The analytical Superose 6 increase 10/300 GL column was connected to the Wyatt system and equilibrated at 20 °C in SEC buffer. 100 µl of sample was injected on column and ran at a flow rate of 0.3 ml/min. The PEX19₍₁₋₂₉₉₎-PEX3_{C235S(41-373)} complex was analysed with a Viscotek 305 tridetector (Malvern Instruments) monitoring light scattering, refractive index, and UV absorbance. Similar to the SEC-MALS procedure, an analytical Superdex 200 10/300 GL column was connected to the setup and the process was performed as previously. Data was acquired and evaluated with the provided OmniSEC software (Malvern). For both the SEC-MALS and SEC-RALS experiments, BSA was used as an internal control. Molecular mass was estimated using a refractive index combined with light-scattering data refractive index increment with protein concentration (dn/dc) of 0.185 ml/g.

2.4. Protein-protein interaction analysis techniques

2.4.1. Fluorescence anisotropy polarisation

The aim of fluorescence anisotropy polarisation is used to non-disruptively quantify the associations between a fluorescent ligand with a larger molecule. This technique can measure the rotational diffusion of a molecule, so when the fluorescent ligand interacts with its binding partner, the product exhibits lower rotational diffusion and higher fluorescent anisotropy. Typically, the concentration of the ligands is constant and is titrated in a range of concentrations of the untagged partner molecule, making it possible to measure their affinity.

Fluorescently-labelled peptides ACBD5_{wt}, ACBD5_{mut1} and ACBD5_{mut2} were synthesised by GenScript and used in the assay at a final concentration of 6.7 nM. Assays were performed in black 96-well plates (Greiner Bio-One) with an Infinite M1000 plate reader (TECAN) regulated at 25 °C, using the excitation diode at 470 nm and detecting the emitted light at 530 nm. The experiment was performed in FA buffer (**Table 3**) complemented with 0.67 mg.ml⁻¹ BSA to prevent unspecific binding on the surface of the plastic well. The protein concentration series was obtained by successive dilution by a factor of 0.75 allowing to establish complete titration curves. Three independent measurements were performed and binding data were analysed using Prism (GraphPad Prism software, Inc.). Binding profiles were fitted using a simple model assuming the stoichiometry one to one.

2.4.2. Chemical crosslinking-mass spectrometry (XL-MS)

Chemical crosslinking coupled with mass spectrometry (XL-MS) is used as a means to detect transient intermolecular or intramolecular interactions between biomolecules and can provide indispensable insight into both the structure and the organisation of proteins in a wide variety of conditions. Common crosslinking protocols utilise homo-bifunctional NHS-esters which can form bridges between lysine residues that are in close proximity or hetero-bifunctional crosslinkers that are able to form bonds between distinct groups, such as amines and carboxyl groups.

For this work, the homo-bifunctional crosslinker bis(sulfosuccinimidyl)suberate or BS³ (#21580, Thermo Scientific Pierce) and the hetero-bifunctional crosslinkers 1-ethyl-3-(3-

dimethylaminopropyl)carbodiimide hydrochloride (EDC, PG82073, Thermo Scientific Pierce) and N-hydroxysulfosuccinimide (sulfo-NHS, PG82071, Thermo Scientific Pierce) were used for complexes PEX19₍₁₋₂₉₉₎- PEX3_{C235S(41-373)}, PEX19₍₁₋₂₉₉₎- PEX14₍₁₆₋₈₀₎.

For the BS³ crosslinker reaction, the proteins were prepared at 15µM concentration in BS³ buffer and mixed in equimolar ratio prior to the reaction. BS³ was dissolved in the same buffer and used at a final concentration of 0.25mM. The reaction was incubated at room temperature for 5 minutes and quenched with the addition of quenching buffer at a final concentration of 25mM Tris buffer (**Table 3**).

For the EDC/sulfo-NHS two-step reaction, PEX3_{C235S(41-373)} was prepared at 30µM in activation buffer and was incubated for 15 minutes at room temperature with 5mM EDC. Similarly, 30mM of PEX19₍₁₋₂₉₉₎ were prepared in coupling buffer and mixed with the activated PEX3_{C235S(41-373)}. After addition of 2mM of sulfo-NHS, the reaction was incubated at room temperature for 2 hours and was quenched with 25mM Tris buffer.

Finally, both reactions were analysed by SDS-PAGE. The bands corresponding to the crosslinked product were isolated and treated to tryptic digest. Liquid chromatography-mass spectrometry (LC-MS) was performed by collaborators at the University of Freiburg (Dr. Friedel Drepper) who used the instruments Orbitrap Elite with CID fragmentation (Thermo Fisher Scientific) and Q-Exactive+ with HCD fragmentation (Thermo Fisher Scientific) for analysis of the resulting peptides. The peptides that were detected were further processed using the programme xQuest (Rinner *et al*, 2008).

2.4.3. Native mass spectrometry

Native mass spectrometry is a useful approach for studying intact biomolecular structure of proteins and protein complexes in the near-native state, in the gas phase. This analysis was performed by Dr. Friedel Drepper and Julian Bender (University of Freiburg, Germany). The PEX19₍₁₋₂₉₉₎- PEX14₍₁₋₃₇₇₎ and PEX19₍₁₋₂₉₉₎- PEX14₍₁₆₋₈₀₎ complexes were transferred into 200 mM ammonium acetate, directly loaded onto in-house prepared gold-coated nano-electrospray ionization needles (nano-ESI) and sprayed on a nano-ESI source with no backing pressure applied. Spectra were recorded on a Micromass Q-ToF Ultima mass spectrometer (Waters) modified for

the transmission of high masses. Background of data was subtracted using Massign (Morgner & Robinson, 2012).

To determine the components of the samples, the proteins were precipitated by acetone, reduced with TCEP and alkylated with iodoacetamide before being subjected to tryptic in-solution digest. LC-MS analyses were carried out on a Q Exactive mass spectrometer (Thermo Fisher Scientific). For protein identification and quantification the programme MaxQuant (Tyanova *et al*, 2015) was used to search for PEX19₍₁₋₂₉₉₎ and PEX14₍₁₋₃₇₇₎ against the UniProt organism-specific proteome set and sequences. Identified proteins were sorted by the estimate of an intensity based absolute quantification (iBAQ).

2.5. Structural characterisation of proteins

2.5.1. X-ray crystallography

The primary objective of protein crystallography is the determination of the 3-dimensional structure of protein macromolecules using X-ray diffraction techniques.

2.5.1.1. *Principles of protein crystallisation*

A crystal is a solid-state phase in which all the molecules comprising the crystal are periodically placed in three dimensions. Protein crystals are held together by weak, noncovalent interactions, such as van der Waals forces, hydrogen bonds or hydrophobic interactions and usually contain a high percentage of solvent which can range from 40-60% of the crystal volume (Matthews, 1968).

Protein crystallisation can be characterised as a reversible equilibrium phenomenon driven by the need to minimise the free energy of a system converting from soluble to crystalline state (Pusey *et al*, 1986). Supersaturation is an important catalyst for crystallisation, as it affects the rate of nucleation and growth of crystals, and it can be accomplished by gradual dehydration of the solvent and change of protein solubility. Generally, two phenomena are crucial for crystallisation; nucleation and crystal growth. Nucleation is the primary step in which the soluble protein molecules cluster into nuclei, causing an increase in the local concentration of the solution. It is during the nucleation stage that molecules become arranged in a defined and periodic manner forming a crystal lattice.

A classic explanation for nuclei formation can be given using a phase diagram, which describes the crystal formation procedure as a function of protein concentration with regards to an adjustable parameter (such as precipitant concentration) (**Figure 9**). A phase diagram can be divided into four areas, (1) a high supersaturation zone in which the protein will form amorphous precipitate, (2) a moderate supersaturation zone which is where spontaneous nucleation will take place (nucleation zone), (3) a lower saturation zone where crystal growth occurs (metastable zone), and lastly (4) an undersaturated zone, in which protein concentration is too low for crystallisation to occur (Chayen

et al, 1996). Cessation of crystal growth can happen for several reasons, such as diminishing of available protein, resulting in an undersaturated solution.

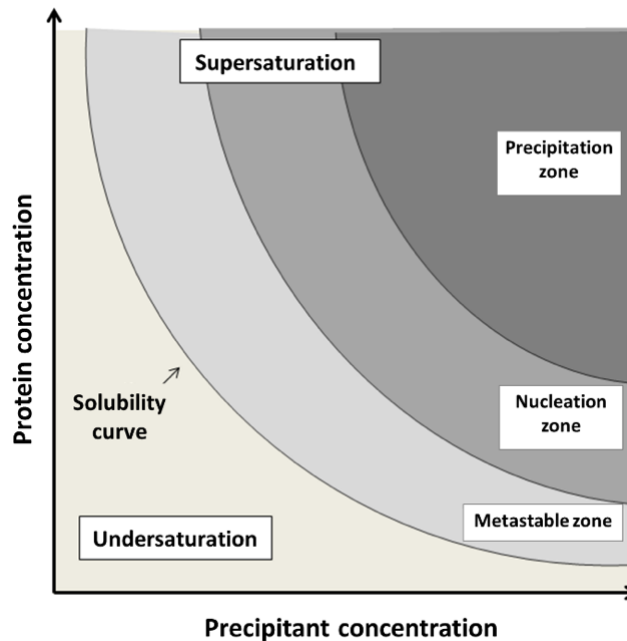


Figure 9: Indicative phase diagram

Successful crystallisation involves the systematic examination of several parameters that can affect crystal formation. The factors that can influence crystallisation are summarised below:

- Homogeneity: high level of protein purity (~97%) is usually required for protein crystallisation (Geerlof et al, 2006).
- Protein concentration: as suggested by the phase diagram, higher protein concentration enables the equilibrium to shift towards supersaturation which can advance precipitation and nuclei formation. However, if protein concentration is too high, crystallisation is prevented and amorphous precipitation is observed.
- Protein solubility: protein solubility is lowest at the isoelectric point (pI) of each protein, as its charge is zero at that point. Variation of the pH around the isoelectric point will allow the formation of electrostatic interactions between the protein molecules and promote crystal formation.
- pH: solvent pH affects the amphoteric side chain charge of the protein and consequently interferes with the intermolecular protein-protein and protein-solvent interactions.

- Temperature: temperature is an important factor for crystallisation as it can affect protein solubility. The standard temperatures used for protein crystallisation trials vary between 4-20°C, but other temperatures are also applicable.
- Crystallisation duration: depending on the complexity and flexibility of the system, crystals may need minutes or months to appear.
- Ionic strength: addition of salts in the crystallisation solution can influence its ionic strength, as salt molecules compete with protein molecules, reduce their solubility and thus facilitate crystal formation. There are two methods in which salts can be used for crystallisation; the ‘salting out’ method employs the addition of a soluble salt into a protein solution, which reduces protein solubility and aids its precipitation. Conversely, the ‘salting in’ method causes protein molecules to become more soluble, as the salt ions bind on their surface and enhance their hydrophilicity. In this case, the protein can crystallise by reducing salt concentration (Arakawa & Timasheff, 1982; Collins, 2004).
- Ligands: small molecules can alter the folding or conformation of a protein they are bound to and subsequently the interaction surfaces that are exposed to the solvent and its neighbouring molecules.
- Non-polar solvents: organic solvents can reduce protein solubility. In theory, non-polar solvent molecules bind to the polar side chains on the surface of the protein, rendering it less polar, or alternatively, reducing the effective number of polar solvent groups.
- Precipitating agents: use of polymers such as polyethylene glycol reduces the solubility of protein molecules, as these polymers serve to inhibit the interaction between solvent and protein.

Due to the difficulties posed by protein crystallisation, several methods have been devised to facilitate this procedure. The most commonly used methods take advantage of the process of vapour diffusion between solvents (‘hanging drop’ and ‘sitting drop’ crystallisation), but alternatives such as microdialysis, batch (or microbatch) crystallisation, or lipidic cubic phase for membrane proteins, are also popular for more difficult cases (Chayen & Saridakis, 2008).

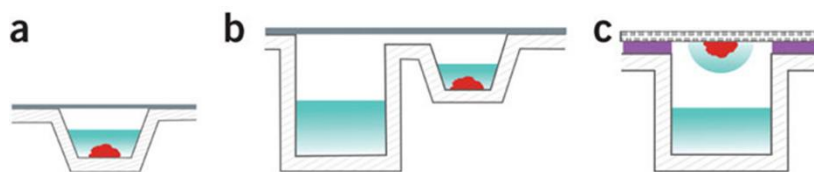


Figure 10: Selected crystallisation methods. These include (a) microbatch, (b) sitting drop, and (c) hanging drop. The precipitant solution is coloured light teal, the protein is red, the vacuum grease purple and the sealing tape gray (Caffrey & Cherezov, 2009).

Vapour diffusion involves the equilibrium between a liquid reservoir of precipitant solution with a drop that contains a ratio of protein with the same combination of reagents. As the drop contains lower reagent concentration compared to the reservoir, water evaporates gradually from the drop and consequently increases protein concentration until an equilibrium is achieved. As mentioned, there are two ways for vapour diffusion (**Figure 10**) both widely used, with the only difference being the orientation of the protein solution relative to the reservoir; in the hanging drop method, the protein solution is positioned vertically above the reservoir, whilst in the sitting drop method, the protein solution is placed on a shelf within the reservoir.

2.5.1.2. *Crystallographic symmetry*

A crystal is a three-dimensional periodic repetition of identical structural units. The periodicity is described by a mathematical lattice and the structural units are atoms that form a specific arrangement. A **unit cell** is the smallest part of a lattice that can be used to reconstruct the full crystal if translated in all three directions. Its dimensions can be defined as the lengths of its three axes a , b and c , as well as the resulting angles α , β and γ (**Figure 11A**). Furthermore, a unit cell contains multiple identical molecules arranged in such manner that can reproduce the crystal if symmetry operations are applied (described below). Nonetheless, the largest collection of molecules that cannot be reproduced by symmetry, but can produce the unit cell upon introduction of symmetry is called the **asymmetric unit** (ASU) (**Figure 11B**).

There are seven Crystal Systems that can be used to describe and classify the symmetry of a crystal. Depending on the relationship between the unit cell parameters, a crystal can belong to either of the following crystal systems: Triclinic, Monoclinic, Hexagonal, Orthorhombic, Tetragonal and Cubic. Furthermore, a lattice can be described by its centering, which refers to the

position of lattice points relative to a unit cell. Combination of the crystal system classification and the lattice centering gives rise to 14 Bravais lattices (**Figure 11C**). Each Bravais lattice corresponds to a specific type of centering:

- **Primitive (P)**: lattice points only on unit cell edges
- **Body-centered (I)**: extra lattice point in unit cell center
- **Face-centered (F)**: extra lattice points on each face of the unit cell
- **Base-centered (A, B, C)**: extra lattice points on either of the bases of the unit cell

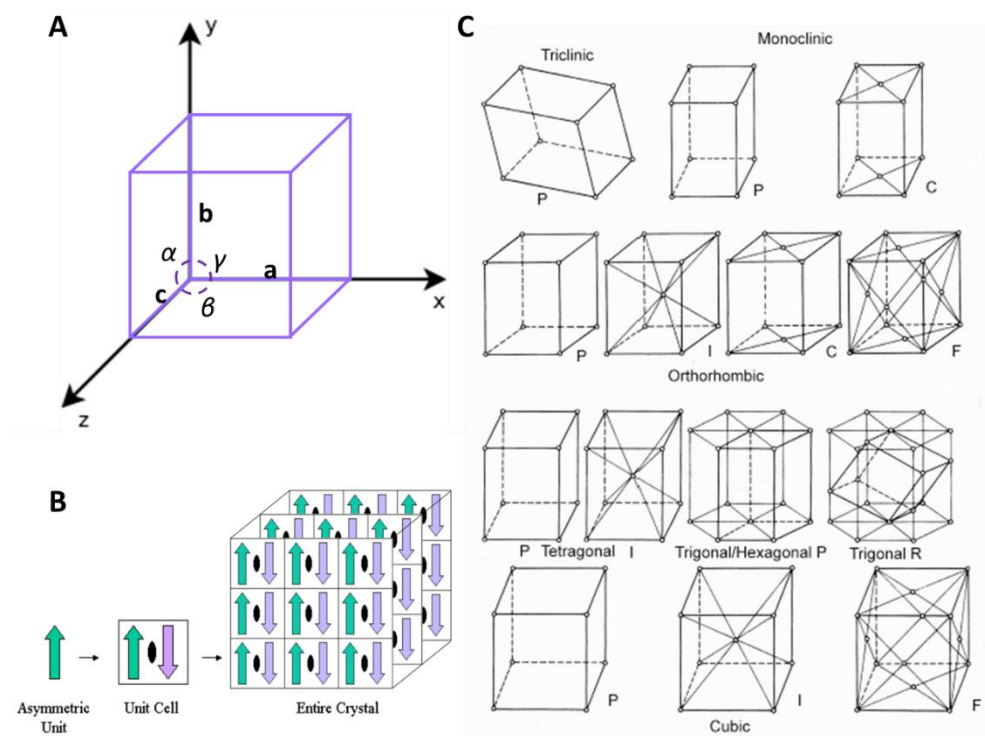


Figure 11: Crystal symmetry. (A) Diagram of a unit cell, with its corresponding axes a , b , c and angles α , β and γ . (B) Relation between the asymmetric unit, the unit cell and the crystal (source: <https://pdb101.rcsb.org/>). (C) Schematic representation of the 14 Bravais lattices (source: <http://bioweb.uwlax.edu/>).

Another important crystallographic notion is that of crystallographic **point groups**, which are mathematical descriptions of the rotational symmetry operations used to define the symmetry of a finite object within a unit cell. There are 32 possible point groups, which, in combination with the Bravais lattices result in 230 space groups that depict all possible symmetries. The simplest symmetry operations required for the description of crystal symmetry are **translation** (movement by a certain distance), **rotation** around an axis, and **reflection** (reflection of a 2-dimensional object

across a plane), as well as their combined operations, **screw axis** (rotation and translation), and **glide reflection** (reflection and translation). Nevertheless, proteins are chiral molecules (they contain only L-amino acid residues) and for this reason certain symmetry operations are impossible to apply (reflection and glide reflection operations), limiting the number of protein space groups to 65. Each space group can be represented by an ensemble of symbols which state the lattice type (Bravais lattice) as well as the symmetry elements existing in a specific crystal.

2.5.1.3. *Miller index*

The bases of the unit cell, in addition to other equally-distanced planes passing through the crystal lattice, can be considered as the source of diffraction and they are described by three parameters, otherwise called lattice constants. Specifically, a family of equivalent lattice planes can be represented by three arithmetic parameters h, k, l that are widely known as **Miller indices**. As a rule, Miller indices define the direction and orientation of the planes in a crystal lattice (Ashcroft & Mermin, 1976). These indices indicate the number of planes or respectively the number of parts in which each unit cell axes a, b or c is intercepted by a family of planes. Miller indices can be notated differently to describe a crystal plane (hkl), a set of equivalent planes $\{hkl\}$, a direction $[hkl]$, and equivalent directions $\langle hkl \rangle$.

2.5.1.4. *Basics of X-ray diffraction*

Similar to visible light, an X-ray beam can be described as an electromagnetic wave, characterised by an electric field and a perpendicular magnetic field. The difference between the two lies on the wavelength, as visible light wavelengths range from 4000-7000 Å, whereas X-rays typically have wavelengths in the 0.1-10 Å range. As the wavelength of X-rays is comparable to the interatomic distances within crystals, it is possible to detect the characteristic interactions between X-ray photons and the electrons of a crystalline material, thus making X-rays suitable for structure determination. Incidence of electromagnetic X-ray radiation on an object results in their scattering from the electrons within the object. The X-ray field forces the electrons to oscillate and

A **Fourier transform** is used to deconstruct any function into an equivalent set of waves of different periodicities. It is commonly used in crystallography to connect the notions of direct and reciprocal space.

emit secondary waves of the same frequency and wavelength as the original wave. The scattered waves are superimposed and in the case of **constructive interference**, detectable diffraction phenomena occur.

Bragg's law was formulated by Sir William Henry Bragg and Sir William Lawrence Bragg, and can be used to explain the phenomenon of diffraction. According to Bragg's law, the angle with which an outgoing diffracted beam exits a crystal can be calculated, if diffraction is considered as a reflection event occurring from a set of equivalent, parallel planes made up of the atoms the crystal consists of. Generally, Bragg's law suggests that when X-rays are scattered from a crystal lattice, peaks of scattered intensity are observed, which abide by the following conditions: (1) the angle of incidence is equal to the angle of scattering, (2) the path-length difference is equal to an integer number of wavelengths. Bragg's law can be summarised by the following equation:

$$n\lambda = 2d \sin\theta$$

with n representing an integer number of wavelengths, λ the wavelength of the X-ray wave, d the interplanar spacing and 2θ being the diffraction angle (**Figure 12**).

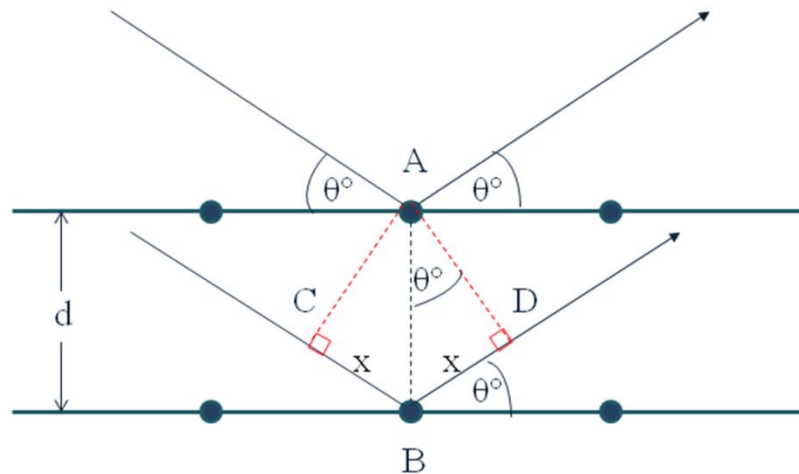


Figure 12: Schematic illustration of Bragg's law. X-rays scattering from planes of atoms exhibit interference effects leading to specific angles where reflections are observed.

2.5.1.5. Reciprocal lattice

Based on Bragg's law, the reciprocal relationship between the diffraction angle θ and the interplanar spacing, d is apparent. This inverse relationship between a crystal and its corresponding

diffraction pattern can be explained by the mathematical construction of **reciprocal lattice**, which is a direct result of a Fourier transformation of the Bravais lattice of a crystal. The reciprocal space lattice corresponds to a set of imaginary points which can be described using the inverse of direct space vectors \vec{a} , \vec{b} and \vec{c} , indicated as \vec{a}^* , \vec{b}^* and \vec{c}^* . These vectors can be calculated using the following equations:

$$\vec{a}^* = \frac{bxc}{a \cdot (bxc)}, \quad \vec{b}^* = \frac{axc}{b \cdot (axc)}, \quad \vec{c}^* = \frac{axb}{c \cdot (axb)}$$

The position of each point on the reciprocal lattice is furthermore described by coordinates h , k and l on the direction defined by the above vectors.

2.5.1.6. *The Ewald construction*

The Ewald construction, or Ewald's sphere, is a geometric construct used in crystallography to describe the relationship between the wavelength λ of an incident and emitted X-ray beam to the diffraction angle 2θ of a distinct reflection, in addition to the reciprocal lattice of the crystal. Ewald's sphere is an important tool for examining whether the conditions for diffraction are fulfilled for a point in reciprocal space, or in other words, whether the Bragg planes have the correct orientation for diffraction to occur (**Figure 13**).

The Ewald construction is represented by a sphere of $1/\lambda$ radius that contains a crystal at its center O. The origin of reciprocal space is located at point A, which is where the emitted X-rays intercept the sphere. If a set of planes causes Bragg diffraction, then the outgoing X-rays can be represented by the vector OB, where B is a point of reciprocal space with coordinates hkl . In that case, vector AB is named scattering vector and is symbolized as \vec{s} . To experimentally observe a reflection hkl , the crystal should be rotated in such a way that the respective point hkl of the reciprocal space coincides with a point on the Ewald sphere. Using geometry and Bragg's law, it is possible to deduce that \vec{s} is equal to $\frac{1}{d_{hkl}}$, meaning that the recorded distance of a diffraction spot from the center of the detector (or the center of the reciprocal space) is inversely related to the distance between two Bragg planes. In the case that $\vec{s} < \frac{2}{\lambda}$, it is impossible to record that reflection, which is the condition for defining the so-called limiting sphere $(A, \frac{2}{\lambda})$. Only lattice points

contained within the limiting sphere can produce recordable diffraction signal, making the choice of wavelength crucial.

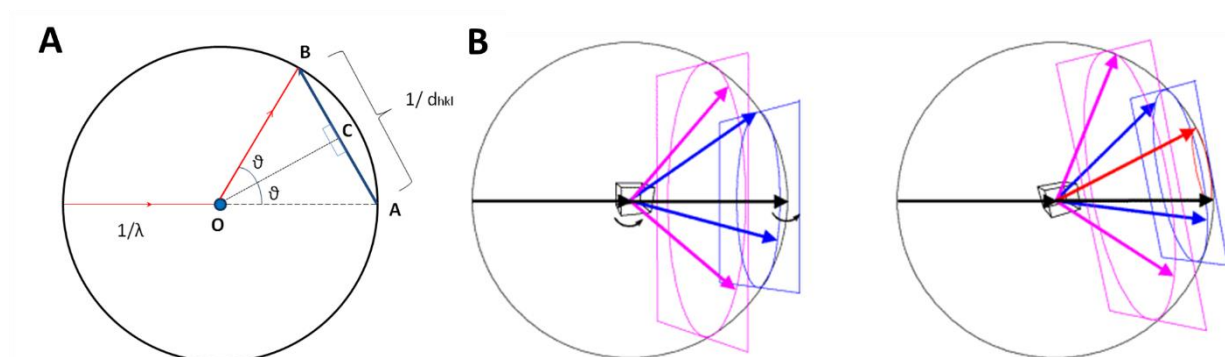


Figure 13: The Ewald construction. (A) Schematic diagram of the Ewald sphere geometry. (B) The left side shows a crystal oriented so that planes in the reciprocal lattice are perpendicular to the X-ray beam. When the crystal is rotated, the reciprocal lattice planes rotate as well (source: <http://www-structmed.cimr.cam.ac.uk/>)

It is important to note the two principles of the Ewald construction; the origin of the direct lattice is located at the center of the sphere, whereas the origin of the reciprocal lattice is located on the same axis as the incoming beam, on the periphery of the sphere. Rotation of the crystal can achieve both the rotation of direct Bragg lattice and reciprocal lattice simultaneously with the same orientation. Thus it is possible to deduce that rotation of the crystal results in rotation of the reciprocal lattice itself. Furthermore, if we consider the detector to be aligned with the initial radiation propagation axis, the diffraction pattern obtained will resemble spots scattered around the center of the detector O, which will correspond to the intersection of the diffracted X-rays with Ewald's sphere and their distance from O will be equal to the measure of the scattering vector \vec{s} , or in other words $\frac{1}{d_{hkl}}$.

2.5.1.7. Atomic form factor and structure factor

Every diffracted ray that impinges on a detector produces a signal that can be described as the sum of the contribution from every electron of a crystal plane hkl that scattered the incident radiation. This summation is otherwise referred to as **structure factor** and is symbolized as F_{hkl} . The structure factor equation can be written as a Fourier transformation, in which each term corresponds to the contribution of an atom to a reflection hkl . On the other hand, the **atomic form factor** (or scattering factor) f_{hkl} describes the contribution of individual atoms to a reflection hkl .

The form factor considers each atom as a sphere of electron density and it varies according to the atomic number of each atom. In effect, a structure factor could otherwise be defined as the sum of all form factors that exist in a certain plane and can be calculated by the following equation:

$$\mathbf{F}_{hkl} = \sum_{j=1}^N \mathbf{f}_j \cdot e^{2\pi(hx_i+ky_i+lz_i)} = \sum_{j=1}^N \mathbf{f}_j \cdot e^{i\varphi_j}, \varphi \text{ being the phase of the wave.}$$

It should further be noted that the intensity of diffraction I is analogous to the squared structure factor:

$$I \approx |\mathbf{F}_{hkl}|^2$$

Of importance is also the Debye-Waller factor (or B factor), which takes into account the thermal displacement of electrons into the calculation of the structure factors. Electron thermal displacement on the electron cloud results in the decrease of scattering ability in wide scattering angles. If this factor is considered, the structure factor can be calculated by the following:

$$\mathbf{F}_{hkl} = \sum_{j=1}^N \mathbf{f}_j \cdot e^{-B\frac{\sin^2\theta}{\lambda^2}} \cdot e^{2\pi(hx_i+ky_i+lz_i)}, \text{ in which } e^{-B\frac{\sin^2\theta}{\lambda^2}} \text{ is the Debye-Waller factor.}$$

Additionally, a structure factor can be written as the integral of the electron density within a unit cell volume. The electron density of a volume element in position (x,y,z) can be calculated as the average of the electron density $\rho(x,y,z)$ in any position:

$$\mathbf{F}_{hkl} = \int_V \rho(x, y, z) \cdot e^{-B\frac{\sin^2\theta}{\lambda^2}} e^{2\pi(hx_i+ky_i+lz_i)} dV$$

This last equation indicates that the structure factor is the Fourier transform of electron density $\rho(x,y,z)$ from a set of planes hkl .

2.5.1.8. *Electron density and the phase problem*

Electron density represents the measure of the probability of existence of an electron in a certain position in an atom:

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l \mathbf{F}_{hkl} e^{-2\pi(hx+ky+lz)}$$

This equation indicates that to calculate the electron density, a series of Fourier transforms of the structure factors is required. As F_{hkl} is a periodic function, it can be described by its amplitude, frequency, and phase. In an X-ray diffraction experiment, the frequency will remain the same as that of the source, while the amplitude is analogous to the square root of the measured intensity of the diffraction I_{hkl} . Nonetheless, the measure of phase ϕ_{hkl} cannot be directly derived experimentally and that is a bottleneck for the calculation of structure factors and subsequently, electron density.

Phase is a term used to describe the position of a wave relative to another wave. Two same waves are **'in phase'** when their maxima and minima coincide, and **'out of phase'** when the maxima of one wave coincide with the minima of the other.

The difficulty of obtaining the correct phases of a diffraction pattern is also known as phase problem, and is a major challenge in crystallography. Resolving the phase problem in most cases equals solution of a crystal structure by determining the correct position of atoms within the unit cell. A variety of methods has been developed for overcoming the phase problem. A powerful phasing method is **Multi-Wavelength Anomalous Dispersion** (MAD), which makes use of the inherent ability of the electrons occupying the inner atomic shell to absorb certain wavelength radiation and re-emit it with time delay, inducing a phase change in all reflections. Analysis of the induced phase alteration of the obtained reflections can aid in solving the phase problem. Another widely used phasing method is **Multiple Isomorphous Replacement** (MIR), which utilizes the property of heavy atoms introduced in a crystal to increase scattering intensity, combined with **Single Wavelength Anomalous Dispersion** (SAD) to calculate phases by comparing Patterson map differences between native and heavy atom derived diffraction patterns. In that case, it is possible to determine the position of heavy atoms in a structure, and consequently the position of protein atoms. Lastly, the widely used phasing method of **Molecular Replacement** exploits protein sequence information, in order to identify similar protein structures whose phases can be used as templates for determining the selected protein structure.

2.5.1.9. *Synchrotron radiation*

X-ray Crystallography as a tool for protein structure solution requires the development of capable X-ray sources, such as Synchrotrons, which can produce high energy and coherence radiation. Synchrotrons are large-scale circular electron accelerators that are crucial for obtaining high resolution and quality crystallographic data. Synchrotron radiation is in fact, ‘light’ emitted from highly accelerated electrons upon application of a magnetic field that will cause a change in their speed and direction of propagation. The wavelength of the radiation depends upon the characteristics of the electrons that are circulating around the central ring of the synchrotron device (**storage ring**). Electrons are initially produced in the center of the synchrotron by an **electron gun** and are accelerated close to the speed of light by a **linear accelerator** (LINAC). Eventually, the electrons are transferred to a circular accelerator, or **booster ring**, where their energy increases, and lastly progress to the outer **storage ring**. In the storage ring, electrons are further accelerated, while passing through magnetic devices that serve to slow them down due to the magnetic field they are producing, causing them to emit electromagnetic radiation. The emitted photons are diverted from their trajectory tangentially to the storage ring and undergo focusing through sets of mirrors, as well as selection of wavelengths by use of devices called **monochromators**.

2.5.2. **Experimental procedures in X-ray crystallography**

2.5.2.1. *Crystallisation*

Crystallisation was performed using the commercial screens Classics_I, PEGS_I, ProComplex and JSCG_Core_I (QIAGEN) available at the Sample Preparation and Characterisation (SPC) facility at EMBL- Hamburg. Crystal drops were set up in sitting drop mode (Swissci, 96-well crystallisation plates, TTP Labtech) using a Mosquito-LCP crystallisation robot (TTP Labtech). Each drop contained 100 nl of protein solution mixed with 100 nl of mother liquor and crystallisation took place at 19°C. In the case of promising crystallisation hits, further optimisation was carried out by refinement of the crystallisation conditions manually. The refinement screens were set up in pre-greased SuperClear™ 24-well plates (Jena Bioscience) using the hanging-drop method.

2.5.2.2. *Data collection*

Synchrotron experiments were performed at beamline ID30A-1 / MASSIF-1 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) which utilises a PILATUS3 2M (Dectris). Crystals of PEX3_{C235S(41-373)} - PEX19_{pep1}- PEX19_{pep2}, PEX3_{C235S(41-373)} - PEX19_{pep2}, Pcs60, Pex5p₍₁₉₈₋₆₁₂₎-Pcs60, Pex5p₍₃₁₂₋₆₁₂₎-Pcs60 and *AtMIF1* were collected using Mounted CryoLoops (Hampton Research) and were cryo-protected with 25% v/v ethylene glycol. The data collection parameters for all crystals are summarised in the table below:

Table 7: Data collection parameters

Parameters	Pcs60 crystal type I	Pcs60 crystal type II	<i>AtMIF1</i>
Wavelength (Å)	0.966	0.966	0.966
Detector distance (mm)	287.8	412.61	196.99
Beam size (µm x µm)	129.8 x 146.83	129.8 x 146.83	50 x 50
Total exposure time (s)	250	90	35
Collection	115° (0.1° oscillation range)	180° (0.1° oscillation range)	332° (0.2° oscillation range)

The PEX3_{C235S(41-373)} - PEX19_{pep1}- PEX19_{pep2}, PEX3_{C235S(41-373)} - PEX19_{pep2}, Pex5p₍₁₉₈₋₆₁₂₎-Pcs60 and Pex5p₍₃₁₂₋₆₁₂₎-Pcs60 crystals were screened at the same beamline, as well as the Petra III P14 beamline (Hamburg, Germany) but produced poor quality diffraction data and no full datasets were collected.

2.5.2.3. *Data processing*

Processing of the diffraction data can provide information about the amplitude of the scattered X-rays. As mentioned previously, the scattering amplitude is proportional to the intensity of the reflections, which can be used for calculation of the structure factors that subsequently contribute to the calculation of the electron density. Furthermore, diffraction data can also supply information about the internal symmetry of the crystal and the unit cell which is important for processing of the data and structure solution. Information about the geometry of the unit cell can help calculate

Matthews coefficient (V_m) (Matthews, 1968), which can estimate the number of molecules in the ASU using the following equation:

$$V_m = \frac{V_{unit\ cell}}{n * (MW)}$$

V corresponding to the volume of the unit cell, n to the number of asymmetric units, and MW to the molecular weight of contents of the ASU.

There are several steps in the data processing procedure, which include indexing, integration and scaling. Indexing makes use of the experimental setup parameters (such as origin of the image, detector distance, wavelength and more) and combines them with information derived from the position of the reflections appearing on a diffraction image. This way it is possible to determine the space group symmetry and unit cell parameters of the crystal. Integration, on the other hand, is used to extract the intensity of each reflection point. It is also important to scale the data, which combines the integrated values of different images into one set of structure factors and normalises them according to symmetry, while the following step of data merging, merges partial and symmetry equivalent measurements into a single value. The produced list of reflections with their Miller indices (hkl), as well as their corresponding signal to noise ratio ($I/\sigma I$) is produced and further quality assessment indicators are calculated for the whole dataset (**Table 8**). Data quality indicators are often used to decide at which resolution to cut the data.

All the acquired datasets were indexed and integrated using the programme XDS (Kabsch, 2010), re-indexed using Pointless (Evans, 2011) for correct identification of the space group, while merging and scaling took place using Aimless (Evans *et al*, 2013) available at the CCP4 suite (Winn *et al*, 2011).

Table 8: Selected crystallographic data quality indicators. I_{hkl} is the intensity of a redundant reflection (hkl), $\langle I_{hkl} \rangle$ is the mean intensity of a measured reflection.

Indicator	Equation	Description
R_{merge}	$R_{merge} = \frac{\sum_{hkl} \sum_j I_{hkl,j} - \langle I_{hkl} \rangle }{\sum_{hkl} \sum_j I_{hkl,j}}$	Widely used to determine the resolution cut-off. However, the higher the multiplicity, the higher is its value, making R _{merge} unreliable.
R_{meas}	$R_{meas} = \frac{\sum_{hkl} \sqrt{\frac{n}{n-1}} \sum_{j=1}^n I_{hkl,j} - \langle I_{hkl} \rangle }{\sum_{hkl} \sum_j I_{hkl,j}}$	Redundancy-independent merging R-factor that gives the precision of individual intensity measurements, independent of multiplicity, showing how broad the distribution is.
CC_{1/2}	$CC = \frac{\sum(x - \langle x \rangle)(y - \langle y \rangle)}{[\sum(x - \langle x \rangle)^2 \sum(y - \langle y \rangle)^2]^{1/2}}$	The dataset is randomly split in half and Pearson's correlation coefficient is determined. Data is usually cut at the point where statistical significance ends (as a rule, data are insignificant if CC _{1/2} < 0.15 or lower) (Karplus & Diederichs, 2012)
I/σI	Indicates the signal-to-noise ratio and is defined as the ratio of the mean intensity and the mean standard deviation of the measurements. As a rule, for useful data the value should be >2.	
Resolution	Resolution is a measure of the quality of the data and can be determined by the position of highest resolvable peak in the diffraction pattern.	
Multiplicity	The average number of observations of each reflection. High multiplicity can indicate better quality data.	
Completeness	The number of crystallographic reflections measured in a data set, expressed as a percentage of the total number of reflections present at the specified resolution	

For the Pcs60 crystal type I, molecular replacement was performed using the Phenix software Phaser (McCoy *et al*, 2007) with the model 3TSY (Wang *et al*, 2011). Phaser can compute the correct molecular replacement solution by means of the Z-score, which can be defined as a signal-to-noise ratio that arises from the comparison of the log likelihood gain (LLG) values from a rotational or translational search. Z-score values of 6-8 indicate possible solution. For AtMIF1, molecular replacement was performed using the automated pipeline BALBES (Long *et al*, 2008). BALBES integrates all components of the molecular replacement procedure that are necessary for structure solution. It is an automated pipeline that requires only the sequence of the protein involved, in addition to an X-ray data file, in order to perform data checking, homologue search in its internal database and generation of template models that will be used for molecular replacement.

2.5.2.4. *Model building and refinement*

Obtaining suitable starting phases will enable the calculation of an initial electron density map, which requires several iterative steps of refinement. The aim of this refinement procedure is to adjust the model to the experimental data and improve the crystallographic R-factors (R_{work} and R_{free}), which are measures of the ability of the models of macromolecular structures to explain the crystallographic data and can be calculated from the following equation:

$$R = \frac{\sum_{hkl} |F_{hkl}^{\text{obs}} - F_{hkl}^{\text{calc}}|}{\sum_{hkl} F_{hkl}^{\text{obs}}}$$

R_{work} and R_{free} differ in the set of reflections they are calculated from; R_{work} is derived from the working set of reflections, whereas R_{free} is calculated from a small subset of reflections that are not used in the refinement process. This way it is possible to monitor the refinement process and ensure that the R-factors are not artificially reduced due to over-refinement. Other parameters commonly used to describe a model is the atomic B-factor, which describes the atomic displacement of atoms due to thermal movement. Furthermore, it is important for the model to make sense chemically, so stereochemical restraints should be applied. These refer to bond lengths and angles as well as torsion angles which should result in < 0.2% of Ramachandran outliers in the refined structure.

Ramachandran plot: enables visualisation of energetically allowed regions for backbone dihedral angles ψ against ϕ of amino acid residues in protein structure.

Refinement of both Pcs60 crystal structures was performed using the software Phenix.Refine available at the Phenix crystallographic suite (Afonine *et al*, 2012; Adams *et al*, 2010). Sequence information provided to Phenix.Refine allowed preliminary structure building which was further improved manually using the software *Coot* (Emsley *et al*, 2010) after each refinement cycle. *Coot* was also used for validation of geometric parameters and assessment of model quality. In the case of *AtMIF1*, the structure was build using ARP/wARP (Langer *et al*, 2008) which was integrated in the BALBES pipeline and further refined using Phenix.Refine similar to Pcs60.

2.5.2.5. *Structure analysis*

In order to analyse the interfaces between the non-crystallographic as well as the symmetry-related molecules within the crystal, the programme PISA was utilised (Krissinel & Henrick, 2007). PISA can analyse crystal structures to identify the component chemical monomers and their interfaces, as well as evaluate the strength of their interactions. That way it can predict whether neighbouring monomers in a crystal form stable multimers that could signify a likely biological form or assembly.

2.5.3. **Small angle X-ray scattering (SAXS)**

Small angle X-ray scattering (SAXS) is a powerful method for the structural characterisation of biological macromolecules in solution. It is a biophysical method used to study the overall shape and structural transitions of proteins, as well as provide information on their shape, conformation and assembly state (Kikhney & Svergun, 2015).

2.5.3.1. *Theoretical background*

A SAXS experiment requires a solution of particles (in this case protein molecules) placed in a quartz capillary, which is illuminated by a collimated monochromatic X-ray beam. The X-rays are scattered by the molecules and their scattered intensity is recorded by an X-ray detector (**Figure 14**). A SAXS experiment is fundamentally different from X-ray crystallography, as crystals make use of the repeating array of molecules in the crystal to amplify the diffraction signal in certain orientations based on the symmetry of the crystal, while SAXS, relying on the scattering of single particles in an isotropic solution, produces lower intensity diffraction patterns concentrated around the primary beam position, which is where the name “small-angle X-ray scattering” was derived.

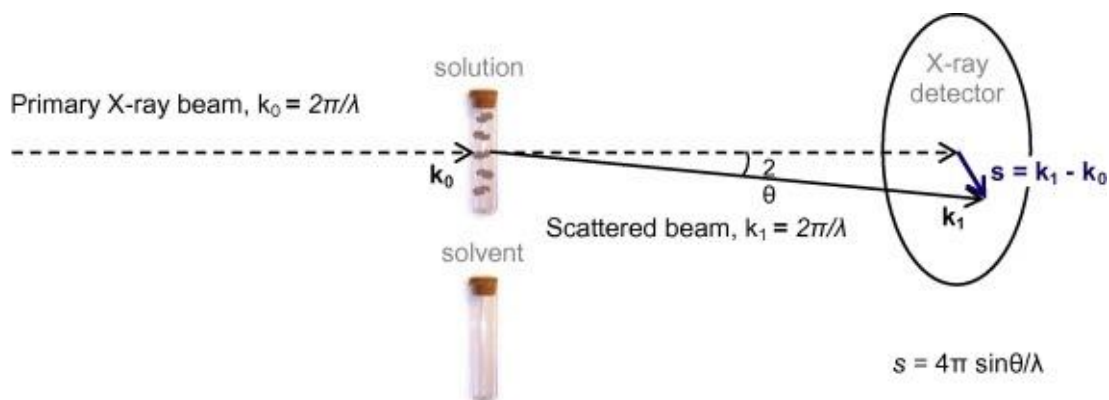


Figure 14: Experimental setup of a SAXS experiment. Proteins in solution are irradiated by an X-ray beam and the diffracted rays are recorded on an X-ray detector. It is important to also measure the contribution of the matching solvent, for subsequent subtraction (Kikhney & Svergun, 2015).

X-ray scattering by a solution of biological macromolecules relies on the concentration of biomolecules being illuminated and the contrast that is produced by excess scattering length density. In the case of X-rays, the contrast, $\Delta\rho(r)$, is derived from the difference in the electron density of the solute and the solvent, which tends to be very small for biomolecules in aqueous solutions and thus, the contribution of the background needs to be minimised. That is usually done after data collection, after a step of background subtraction in which the scattering of the solvent is removed.

Solutions of monodisperse biological macromolecules generate patterns displaying isotropic scattering intensity, which depends on the absolute value of the momentum transfer s ($s = 4\pi\sin(\theta)/\lambda$, where 2θ is the angle between the incident and scattered beam):

$$I(s) = \langle I(s) \rangle_{\Omega} = \langle A(s)A^*(s) \rangle_{\Omega}$$

where the scattering amplitude $A(s)$ is a Fourier transformation of the excess scattering length density, and the scattering intensity is averaged over all orientations (Ω). Following subtraction of the solvent scattering which contributes to the recorded scattering, the background corrected intensity $I(s)$ becomes proportional to the scattering of a single particle averaged over all orientations.

2.5.3.2. SAXS data acquisition

SAXS data are usually acquired using synchrotron sources that provide higher brilliance compared to in-house diffractometers. Practically, for a SAXS experiment, 10-100 μl of highly monodisperse sample per measurement are required, including a necessary concentration series measurement (1-10 mg/ml). To avoid radiation damage, short exposure measurements are performed on the same sample and are compared to detect the radiation-induced differences (Kikhney & Svergun, 2015; Franke *et al*, 2012).

In this thesis, SAXS of the PEX3_{C235S(41-373)} - PEX19₍₁₋₂₉₉₎, PEX3_{C235S(41-373)}- PEX19₍₁₆₁₋₂₈₃₎, PEX3_{C235S(41-373)}- PEX19₍₅₄₋₂₉₁₎, PEX3_{C235S(41-373)}- PEX19 _{Δ (62-148)}, PEX19₍₁₋₂₉₉₎ - PEX14₍₁₋₃₇₇₎, Pex5p₍₁₉₈₋₆₁₂₎-Pcs60, Pex5p₍₃₁₂₋₆₁₂₎-Pcs60 and their separate components was performed at the Petra III beamline P12 (Blanchet *et al*, 2015). The samples were monodisperse in a concentration series of 0.5-5 mg/ml and were illuminated by a monochromatic 1.24 Å beam at 10°C for 0.045 s and the 2D photon counting Pilatus 2M detector (Dectris) was placed at a 2 m distance.

2.5.3.3. SAXS data analysis

SAXS scattering patterns are typically presented as radially averaged one-dimensional curves that can reveal several important parameters, such as size of the particles, oligomeric state and overall shape. Similar to crystals, the scattering intensity is amplified by the number of molecules, signifying that the sample scattering intensity is analogous to its concentration. This, in turn, implies that in higher concentrations the signal to noise ratio improves, but, on the other hand, there are higher possibilities of sample aggregation. To minimize this contribution, the low angle data measured at lower protein concentrations are normally merged with high angle data at higher concentrations, yielding a composite curve with low aggregation effects at low angles and higher signal to noise ratio at high angles. The original manipulation of the data is performed using the ATSAS suite software PRIMUS (Petoukhov *et al*, 2012; Konarev *et al*, 2003).

Further ways of data representation can provide additional information and are summed up in the table below:

Table 9: SAXS plots and experimental parameters.

Plot types	Description
Experimental data	SAXS data are presented as $\log_{10}I(s)$ vs. s , where $I(s)$ is the intensity (in arbitrary units) and s is the modulus of the scattering vector (in inverse nanometres, $s=4\pi\sin(\theta)/\lambda$, where 2θ is the scattering angle, λ is the wavelength).
Guinier plot	Refers to low angle data plotted as $\ln I(s)$ vs. s^2 . The Guinier plot can give information about the radius of gyration (R_g), as well as assess the quality of the data from the linearity of the plot.
Kratky plot	Can be used to estimate the degree of compactness, represented by a $s^2I(s)$ vs. s plot. A bell-shaped peak indicates a globular particle, whereas a plateau can indicate extended or unfolded proteins.
Experimental parameters	Description
Radius of gyration (R_g)	Refers to the average of square center-of-mass distances in the particle weighted by the scattering length density. R_g is a measure for the overall size of the macromolecule.
Forward scattering intensity (I_0)	The intercept of the Guinier plot fit can give the forward scattering intensity I_0 which is proportional to the solution concentration and to the number of atoms in the particle.
Porod volume (V^{porod})	For globular particles of uniform density the excluded volume of the hydrated particle is inversely related to the area below the Kratky plot. It can be used to estimate the molecular weight of the measured protein.
Pair distance distribution function ($p(r)$)	$p(r)$ represents the distribution of distances between all pairs of atoms within the particle, weighted by their respective electron densities. $p(r)$ is the Fourier transform of $I(s)$ (and vice versa).
Maximum intra-particle distance (D_{max})	The point where $p(r)$ is decaying to zero is called D_{max} and represents the maximum size of the particle.

2.5.3.4. *Ab initio* modelling

Advances in computational methods have enabled the determination of three-dimensional structures derived from SAXS data. Low resolution SAXS models (1-2nm) can be produced *ab initio* or through refinement of available high-resolution structures or homology models. In all *ab initio* modelling methods, the χ^2 value is used as a measure of the discrepancy between theoretical and experimental curves.

In this thesis, SAXS *ab initio* modelling was used by means of the following software available on the ATSAS suite:

- DAMMIF (Franke *et al*, 2009): a tool for rapid *ab initio* shape. Using DAMMIF, a particle is represented as a collection of beads inside a search volume. Starting from an arbitrary initial model, DAMMIF utilizes simulated annealing procedures to construct a compact model that fits the experimental data.
- CORAL (Petoukhov *et al*, 2012): CORAL (COmplexes with RANdom Loops) performs SAXS-based rigid body modelling of complexes, of which partial high resolution structures are available. A simulated annealing protocol is used to find the optimum positions and orientations of these structures and the approximate conformations of the missing portions of polypeptide chains.
- BUNCH (Petoukhov & Svergun, 2005): BUNCH performs modelling of multi-domain proteins against SAXS data combining rigid body and *ab initio* modelling. Similar to CORAL, BUNCH uses a simulated annealing protocol to find the optimal positions and orientations of available partial structures and the probable conformations of the dummy residues chains attached, to fit the experimental scattering data.
- EOM (Tria *et al*, 2015): Ensemble Optimisation Method (EOM) fits an averaged theoretical scattering intensity of an ensemble of conformations to experimental SAXS data. A pool of n independent models based on sequence and structural information is generated, and comparison of the averaged theoretical scattering intensity from these models against the scattering data is performed. Finally, the ensemble that best describes the experimental SAXS data is selected.
- Saspy (Panjkovich & Svergun, 2016): a PyMOL plugin for manipulation and refinement of high resolution models against SAXS data. This program represents each input structure as an ensemble of points and attempts to align them.

2.5.4. Negative stain electron microscopy (EM)

Negative stain microscopy can be used to obtain low-resolution structural information on proteins and protein complexes with a molecular weight larger than 100 kDa. The principle of negative stain screening relies on the staining procedure, which enhances the contrast between the particle and its background, making it visible under an electron microscope. In this technique, the background is stained leaving the actual specimen untouched, thus visible. Negative staining involves the addition of heavy metal salt solutions that form electron-dense surroundings around individual macromolecular complexes. In the electron microscope, these surroundings produce high contrast images that are resistant to radiation damage and can be utilised for sample screening and low resolution structural determination of the particle ($>20 \text{ \AA}$). Images produced by negative stain EM should be processed in a way that similar particles will be assigned to the same class. Firstly, a translational and rotational alignment of the particle images is performed and the aligned images are subjected to classification. After specification of the number of classes, the images of each class are averaged, creating class averages.

This method was performed by collaborators at the Max Planck Institute (Pascal Lill, Dr. Bjoern Klink and Dr. Christos Gatsogiannis, MPI Dortmund) who used the microscopes JEOL JEM 1400 EM (JEOL) and a FEI Tecnai Spirit EM (FEI) to produce preliminary images for the PEX19₍₁₋₂₉₉₎- PEX14₍₁₋₃₇₇₎ and Pex5p₍₁₉₈₋₆₁₂₎-Pcs60 complexes. Data acquisition was performed on both instruments with a TVIPS F416 4K camera (TVIPS).

3. Results

3.1. Structural studies of human PEX19 complexes

Despite previous structural information on the cytosolic domain of the peroxisomal membrane anchor PEX3 with an N-terminal PEX19 peptide, little is known about its interaction with full length PEX19. PEX19 is a flexible protein that accommodates PMP binding and docking into the peroxisomal membrane, as well as facilitating PMP insertion into the membrane. Elucidation of the interactions between PEX19 and PEX3 can provide insight into the mechanism of action of PEX19. Furthermore, PEX19 is known to act as a chaperone and import receptor for a wide range of PMPs. Although PEX14 is a PMP itself, it does not require PEX19 for its insertion into the peroxisomal membrane, making it thus an unusual PMP. To date, the interaction of these two proteins has not been well-characterised, possibly due to difficulties in producing the transmembrane-domain-containing, full-length human PEX14. In the first part of the thesis, the interaction of PEX19 with its various partners is examined by means of structural and biophysical methods.

3.1.1. Structural and biophysical characterisation of the human PEX3-PEX19 complex

The constructs used in this project include the full-length human PEX19₍₁₋₂₉₉₎ and the C-terminal cytosolic domain of PEX3_{C235S} (41-373) (**Figure 15**), with the C235S mutation shown to prevent non-native oxidation (Schmidt *et al*, 2010). These two proteins were expressed, purified and characterised, as described in the following chapters. Additionally to the full length PEX19 protein, shorter constructs were also expressed and purified: (1) PEX19₍₅₄₋₂₉₁₎, which is missing the N-terminal binding site to PEX3, (2) PEX19₍₁₆₁₋₂₈₃₎, which contains only the folded C-terminal domain of PEX19, and lastly (3) the deletion construct PEX19_{Δ(62-148)} which contains the N-terminal PEX3 binding site and the folded C-terminal domain, but the disordered region (51-149) has been replaced by the ovalbumin linker –GSGSGS– as a means to reduce flexibility.

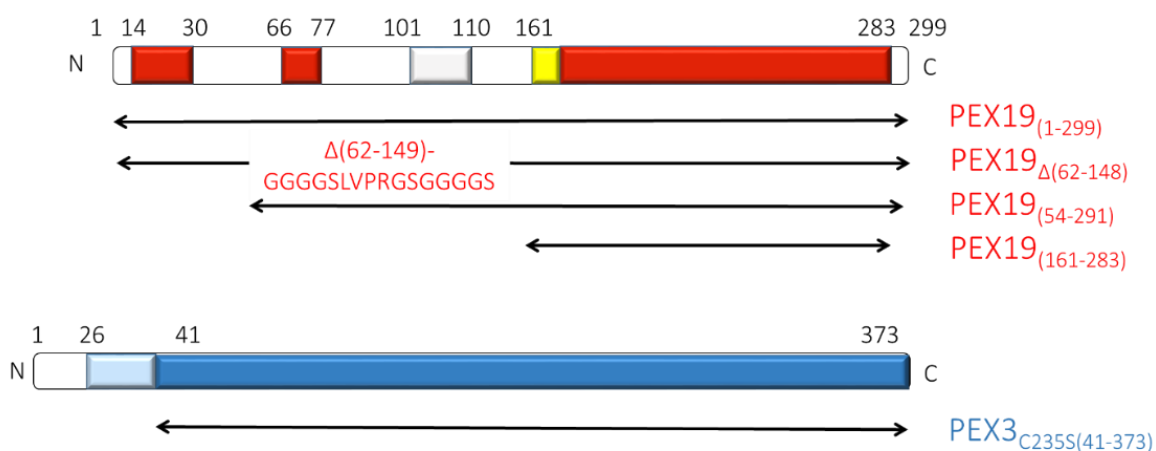


Figure 15: Schematic representation of PEX3 and PEX19 constructs used. The red rectangles on PEX19 correspond to the PEX3 binding site (14-30), the PEX14 binding site (66-77) and the structured PEX19 C-terminal domain (161-283), the grey rectangle corresponds to its predicted secondary PEX3 binding site (101-110), while the yellow one corresponds to the PMP-binding helix (171-182). In the case of PEX3, the light blue and darker blue (41-373) rectangles correspond to its cytosolic domain.

3.1.1.1. Expression, purification and biophysical characterisation of the human PEX3_{C235S(41-373)}-PEX19₍₁₋₂₉₉₎ complex

In order to biophysically characterise PEX19₍₁₋₂₉₉₎ and PEX3_{C235S(41-373)}, as well as to investigate their interaction and complex formation, the two proteins were initially expressed individually using auto-induction media (Studier, 2005) and purified *via* affinity purification and size exclusion chromatography.

PEX19₍₁₋₂₉₉₎ is an extremely soluble 32.8 kDa protein and its expression yields large amounts of very stable protein (~60 mg per litre of expression), which can reach very high concentrations (~100 mg/ml) without any aggregation effect observed. Due to the flexible and disordered nature of its N-terminus, the full length PEX19₍₁₋₂₉₉₎ exhibits the elution and migration profile of a higher molecular weight protein, as indicated by SEC and SDS-PAGE (**Figure 16A,B**). However, molecular weight estimation by SEC coupled with Right Angle Light Scattering (RALS) confirms that the molecular weight of this protein is in fact 35 kDa +/- 10%, which correlates with the theoretical mass of 32.8 kDa that is expected for the monomeric species (**Figure 16A**). Further characterisation of PEX19₍₁₋₂₉₉₎ was performed by Circular Dichroism (CD), which is used to provide predictions with regards to the secondary structure content of a protein. Unsurprisingly, CD estimated ~42% random coil and ~48.5% helical content for the full length PEX19₍₁₋₂₉₉₎ protein

(Figure 16C, Table 10), which is in accordance with previous structural information (Shibata *et al.*, 2004).

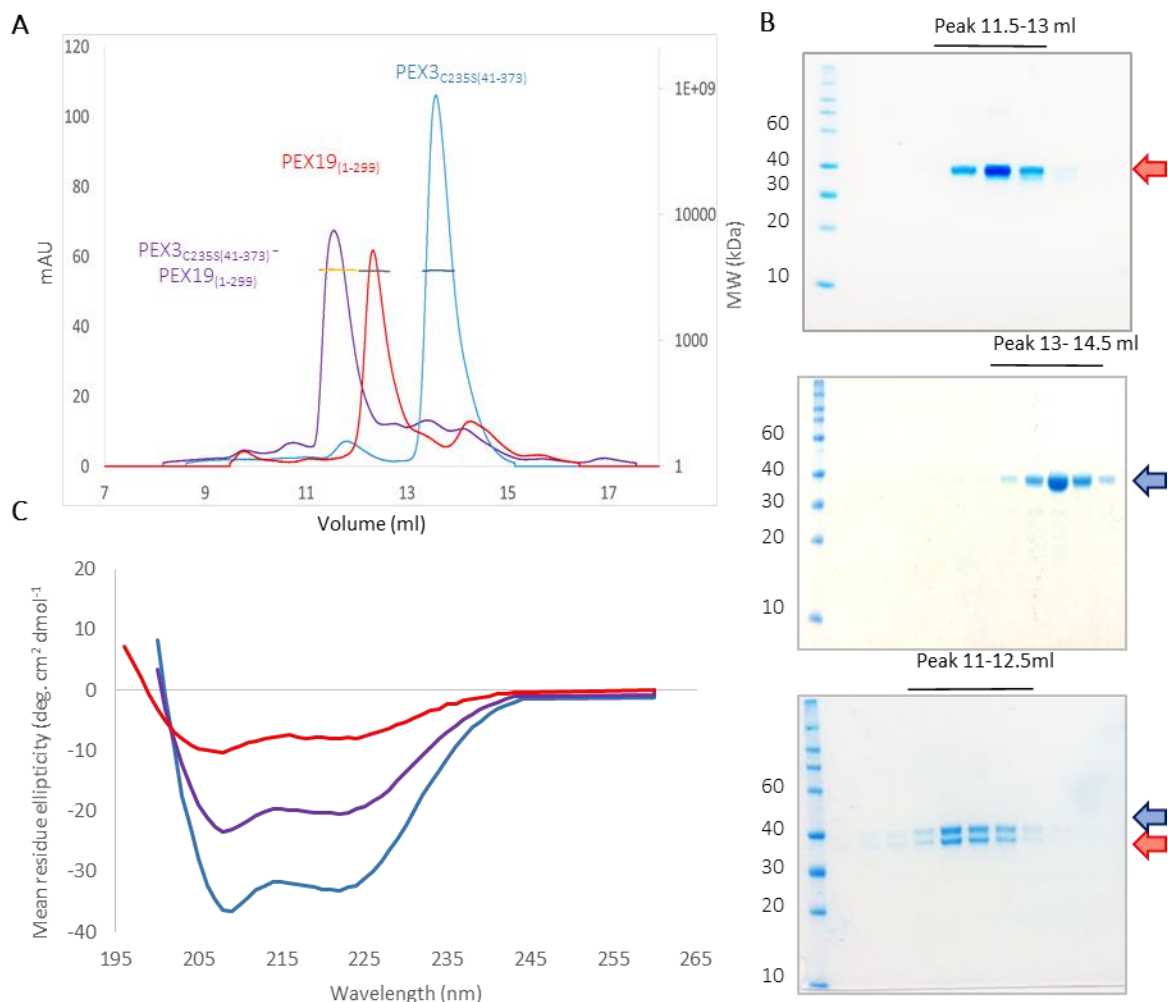


Figure 16: Purification and characterisation of PEX19₍₁₋₂₉₉₎, PEX3_{C235S(41-373)} and their complex. (A) SEC-RALS analysis of PEX19₍₁₋₂₉₉₎ (red), PEX3_{C235S(41-373)} (blue) and PEX19₍₁₋₂₉₉₎-PEX3_{C235S(41-373)} (purple), using a Superdex 200 increase 10/300GL column, (B) their corresponding SDS-PAGE gels, and (C) CD analysis of the complex and its components. PEX19 is indicated with red arrows and curves, PEX3 with blue arrows and curves, while their complex is shown in purple. The horizontal yellow, black and blue lines correspond to SEC-MALS calculated masses for the complex, PEX19₍₁₋₂₉₉₎ and PEX3_{C235S(41-373)}, respectively.

PEX3_{C235S(41-373)} is another soluble protein, the purification of which yielded ~18 mg per litre of expression. Despite the C235S mutation, the remaining cysteine residues in PEX3 have the capacity to form disulphide bonds, causing the protein to aggregate in high concentrations. For that reason, 1mM Tris(2-carboxyethyl)phosphine (TCEP) was added in all buffers, allowing the protein to reach concentrations such as 8 mg/ml. Although PEX3_{C235S(41-373)} is predominantly monomeric in solution, as can be assessed from its elution profile using SEC-RALS (Figure 16A),

it exhibits a peculiar migration profile on SDS-PAGE, in which protein bands corresponding to dimers appeared even under reducing conditions (data not shown). CD evaluation of its secondary structure is in accordance with the known structures (PDB entries: 3MK4, 3AJB), and is predicted to be 95% helical (**Figure 16C**, **Table 10**).

In a similar fashion, the PEX3_{C235S(41-373)}-PEX19₍₁₋₂₉₉₎ complex was formed after equimolar mixing, incubation in ice for 60 min, and an additional step of size exclusion chromatography (**Figure 16B (bottom panel)**). The complex was stable at relatively high concentrations (~6-8 mg/ml) in the presence of TCEP, and SEC-RALS analysis indicated that its elution volume and estimated molecular weight (71 kDa \pm 10%) are comparable to the theoretical molecular weight of the hetero-dimer (70 kDa) (**Figure 16A**). CD experiments on this complex show an intermediate degree of secondary structure formation with 55% helical content and 29.9% random coil, which can be attributed to the N-terminal disorder of PEX19₍₁₋₂₉₉₎.

Table 10: Distribution of secondary structure elements based on CD measurements.

Secondary structure content	PEX3 _{C235S(41-373)}	PEX19 ₍₁₋₂₉₉₎	PEX3 _{C235S(41-373)} -PEX19 ₍₁₋₂₉₉₎
Helix	95%	48.5%	55%
Anti-Parallel	-	5.4%	3.1%
Turn	5%	4.1%	11.6%
Random Coil	-	42%	29.9%

3.1.1.2. Identification of PEX3_{C235S(41-373)} - PEX19₍₁₋₂₉₉₎ binding sites via XL-MS

In order to verify pre-existing hypotheses that the N-terminus of PEX19 contains auxiliary binding sites, which might likely be involved in PMP binding and/or release, chemical crosslinking coupled with mass spectrometry (XL-MS) was employed for their identification. The PEX3_{C235S(41-373)} - PEX19₍₁₋₂₉₉₎ complex was chemically attached by the zero-length crosslinkers EDC/sulfo-NHS that are used to link primary amine side chains (K) to carboxyl side chains (E, N). From SDS-PAGE analysis it is possible to deduce that the crosslinked complex forms a 1:1 hetero-dimer (**Figure 17A**).

Although there are no detected crosslinked peptides for the monomeric PEX19₍₁₋₂₉₉₎, several intramolecular crosslinked peptides were identified for the PEX3_{C235S(41-373)} crosslinked dimers, most of which can be attributed to the hydrophobic patches present on the surface of PEX3 that are responsible for its homo-dimerisation at high concentrations. This homo-dimerisation effect brings the reacting side-chains (amino-groups and carboxylic groups) in close proximity and crosslinking becomes possible. The crosslinked PEX3_{C235S(41-373)} dimers are visible in **Figure 17A**, in the presence of crosslinkers EDC/sulfo-NHS. Conversely, the control lanes of PEX3_{C235S(41-373)} with crosslinker BS³, used to crosslink primary amines (K), resulted in a monomer band of reduced molecular weight, as shown by SDS-PAGE analysis (**Figure 17A**). This could be an effect of a high number of intramolecular crosslinks that cause a decrease of the hydrodynamic radius of PEX3_{C235S(41-373)} and affect its migration profile on SDS-PAGE.

Additional to the intramolecular crosslinks and the already known N-terminal binding site between PEX3_{C235S(41-373)} - PEX19₍₁₋₂₉₉₎, another intermolecular interaction was detected within the middle region of PEX19₍₁₋₂₉₉₎ (residues: 90-110) with an α -helix located at the N-terminus of PEX3_{C235S(41-373)} (residues: 80-100)(**Figure 17B,C**). The PEX3 helix involved in both binding events with PEX19 displays an amphipathic motif, with its hydrophobic surface interacting with the PEX19 amphipathic helix formed of residues 14-30. On the other hand, the predicted amphipathic helix that composes the newly-detected additional binding site, would likely be placed in such manner to allow interaction with the hydrophilic surface of the PEX3 helix (**Figure 17E**). It is noteworthy to mention that this region of PEX19 corresponds to one of the predicted amphipathic α -helices that are presumably important for PEX19 function, helix alpha-c, as referred to by Chen *et al*, 2014.

Moreover, the secondary binding site is mapped adjacent to the known primary binding site on the apex of PEX3_{C235S(41-373)}, suggesting that the N-terminus of PEX19 might first bind *via* its first 14-30 residues and then likely bend in order to allow the interaction of the helices involved in the secondary binding (**Figure 17F**).

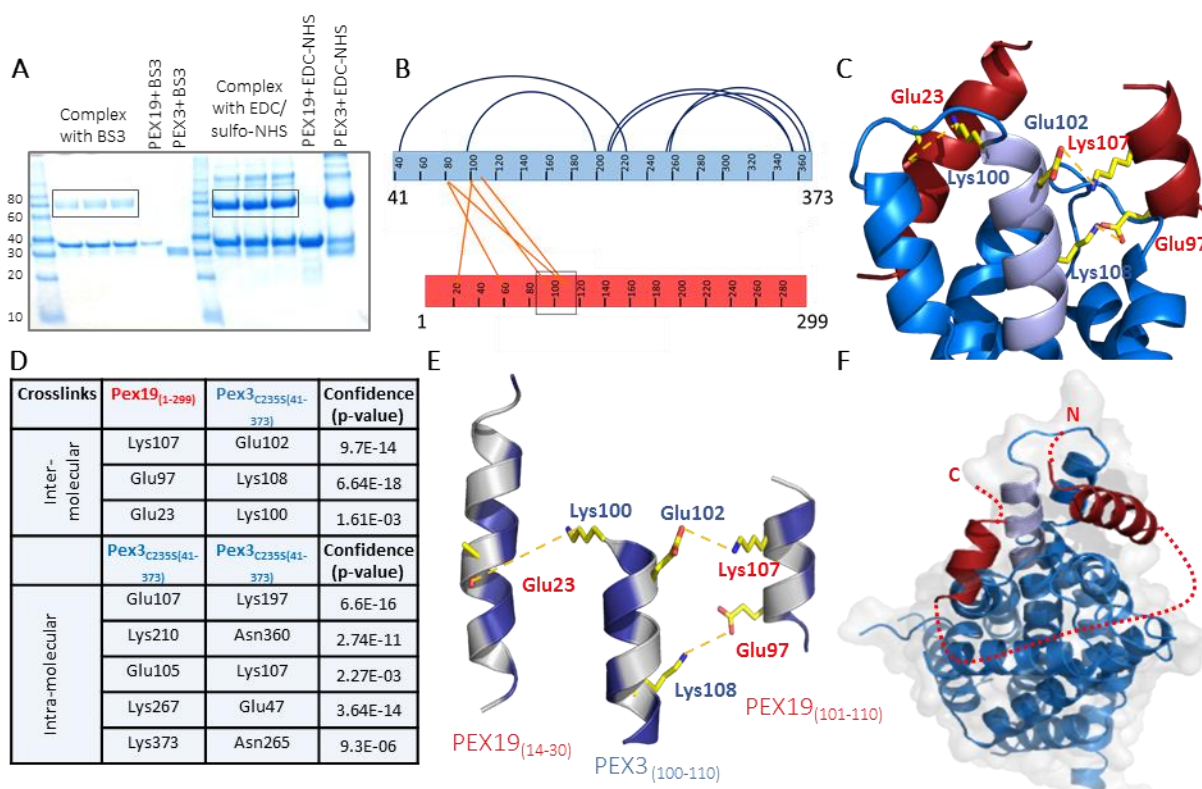


Figure 17: Crosslinking of PEX3_{C235S(41-373)} - PEX19₍₁₋₂₉₉₎. (A) SDS-PAGE analysis of the crosslinked complex with crosslinkers BS³ and EDC/sulfo-NHS. The crosslinked product band is indicated with a black box. (B) Illustration of the intramolecular PEX3_{C235S(41-373)} crosslinks (dark blue) as well as the intermolecular PEX3_{C235S(41-373)} - PEX19₍₁₋₂₉₉₎ crosslinks (orange), (C) Model displaying the intermolecular PEX3_{C235S(41-373)} - PEX19₍₁₋₂₉₉₎ (blue and red respectively). The crosslinked residues are indicated with yellow sticks, connected by orange dashed lines, while the helix of PEX3 that is involved in the interaction is coloured light blue, (D) Table summarising the crosslinked residues detected, with corresponding confidence scores. (E) Amphipathic helices involved in PEX3-PEX19 interaction, as detected by XL-MS. The cartoon is coloured with dark blue to indicate hydrophobicity, grey to indicate hydrophilicity and the crosslinked residues are represented by yellow sticks. (F) Model of PEX19 binding on PEX3. Potential bending of the disordered domain indicated by dashed red lines. The models of the novel PEX19 binding site were modelled using the software Chimera (Pettersen *et al*, 2004).

Sequence analysis using the ConSurf server (Ashkenazy *et al*, 2016; Celniker *et al*, 2013; Landau *et al*, 2005; Glaser *et al*, 2003) shows that the newly detected PEX19 secondary binding site (101-110) displays intermediate conservation (**Figure 18**) and contains the predicted buried residues L101 and F105, whereas all the other residues are exposed. In addition to this sequence, the K107 residue is predicted to be functionally important, as it is highly conserved and exposed. Mutagenesis on this amino acid can be used to examine *in vivo* effects on the phenotype. Extra crosslinked peptides were detected between PEX3-K100 and PEX19-E23, which correspond to the known primary binding site, as well as an additional crosslink between PEX19-K60 and PEX3-

N85, but as the level of conservation of this PEX19 site is very low, they were considered unspecific.



Figure 18: Sequence conservation of human PEX19 and PEX3, as illustrated by the ConSurf server. The letters b, e, f and s, correspond to buried, exposed, functional and structurally important residues respectively.

To validate the XL-MS results experimentally, peptides of the two binding sites ADRELEELLESALDDFDKAK (PEX19_{pep1}) and EPHLVEQFQKLSE (PEX19_{pep2}) were purchased and used for co-crystallisation with PEX3_{C235S(41-373)}, as will be discussed in the following chapters.

3.1.1.3. SAXS analysis for PEX3_{C235S(41-373)}-PEX19₍₁₋₂₉₉₎ indicates a flexible multi-domain system

Small angle X-ray scattering (SAXS) is a powerful technique that can provide *ab initio* and rigid body low-resolution models of a protein or protein complexes, as well as additional information about their oligomeric state, flexibility, and dimensions. Low resolution models of on the PEX3_{C235S(41-373)} - PEX19₍₁₋₂₉₉₎, PEX3_{C235S(41-373)} - PEX19₍₁₆₁₋₂₈₃₎, PEX3_{C235S(41-373)} - PEX19₍₅₄₋₂₉₁₎ and PEX3_{C235S(41-373)} - PEX19_{Δ(62-148)} complexes and their separate components in solution were obtained by SAXS. These additional PEX19 constructs were used as controls for investigation of complex formation and conformation in solution, as they either lack both PEX3 binding sites (PEX19₍₁₆₁₋₂₈₃₎), the primary binding PEX3 binding site (PEX19₍₅₄₋₂₉₁₎), or the secondary PEX3 binding site (PEX19_{Δ(62-148)}). The overall experimental parameters derived from experimental scattering curves are presented in **Table 11**. The experimental scattering curves are shown in **Figure 19A**.

The values of the radius of gyration were computed from the Guinier plot and the maximum particle size (D_{\max}) was obtained from the pair-distance distribution function (**Figure 19C, Table 11**). From this analysis it can be derived that the flexibility of PEX19₍₁₋₂₉₉₎ is contributing to the general particle size, which seems to be rather comparable in presence or absence of PEX3_{C235S(41-373)}. The complexes containing shorter PEX19 constructs also displayed an elongated conformation in the cases of PEX3_{C235S(41-373)} - PEX19₍₅₄₋₂₉₁₎ and PEX3_{C235S(41-373)} - PEX19_{Δ(62-148)}, which indicates that the secondary binding site is sufficient for complex formation. In the case of PEX3_{C235S(41-373)} - PEX19₍₁₆₁₋₂₈₃₎, which lacks the main PEX3 binding site, as well as the secondary binding site, the smaller R_g could be attributed to the possibility that the complex does not form, but instead scattering from a mixture is observed (**Table 11**).

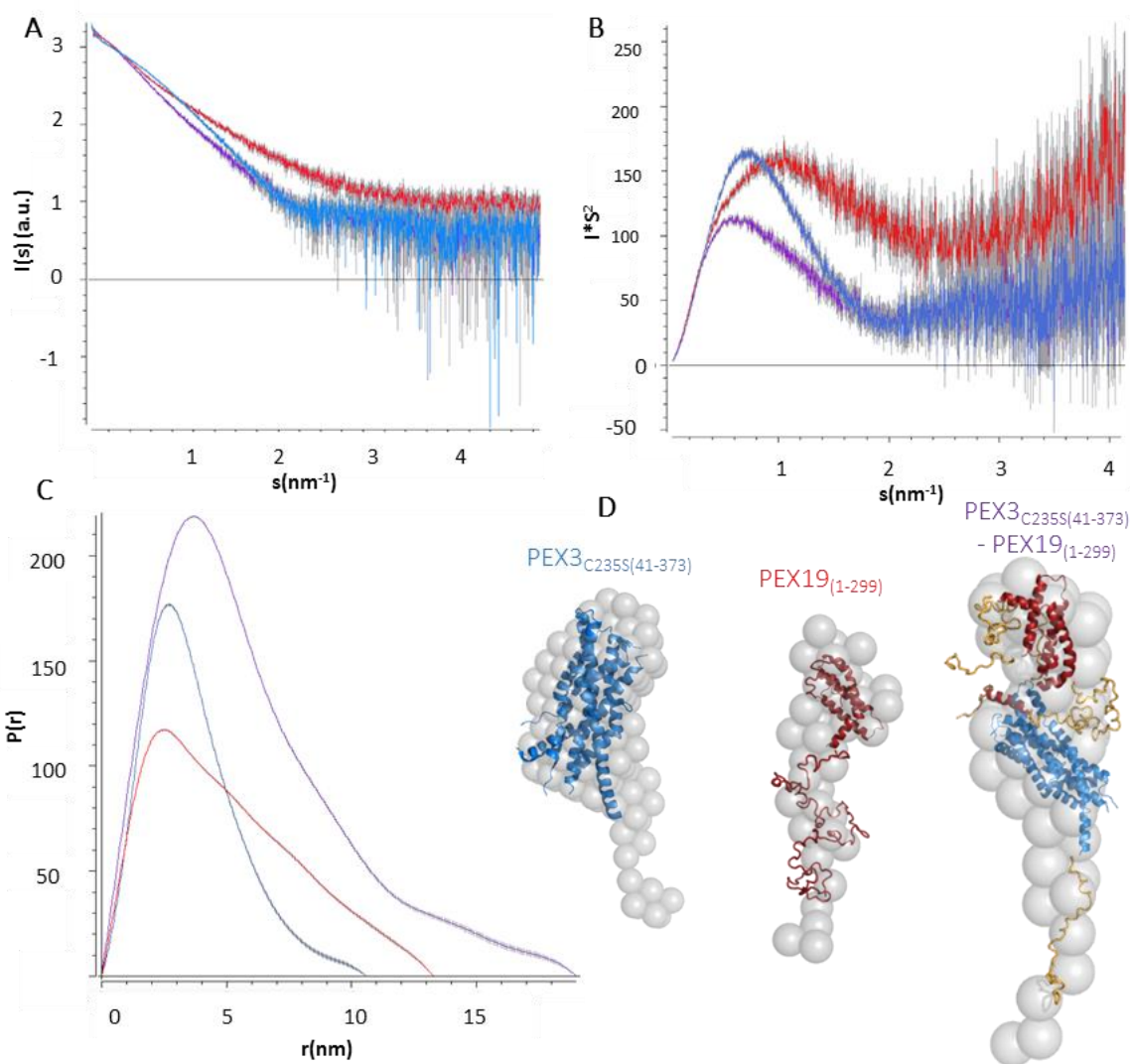


Figure 19: SAXS analysis of the $PEX3_{C235S(41-373)} - PEX19_{(1-299)}$ complex and its separate components. (A) Scaled plotted intensity of the SAXS measurements for $PEX3_{C235S(41-373)} - PEX19_{(1-299)}$ (purple), $PEX3_{C235S(41-373)}$ (blue) and $PEX19_{(1-299)}$ (red). (B) Kratky plots illustrating the degree of compactness. (C) Distance distribution function for the above samples. $PEX3_{C235S(41-373)}$ displays a globular conformation indicated by the bell-shape of its corresponding peak, while $PEX19_{(1-299)}$ shows a partly globular/elongated profile, which is visible by the asymmetry of its peak. (D). Ab initio models using the programme DAMMIF, in sphere representation. The crystal structures of $PEX3_{C235S(41-373)}$ (3MK4), as well as ab initio models made using BUNCH ($PEX19_{(1-299)}$ in red) and CORAL ($PEX19_{(1-299)}$ in red, $PEX3_{C235S(41-373)}$ in blue and the disordered loops in orange) were used to fit into the DAMMIF models.

Furthermore, Kratky analysis was used to assess the compactness of the system (**Figure 19B**) and indicates that PEX3_{C235S(41-373)} displays the characteristics of a globular and folded protein, PEX19₍₁₋₂₉₉₎ is mostly unfolded, while their complex can be described as a partly unfolded system. As expected, the SAXS-derived low resolution *ab initio* model for PEX19₍₁₋₂₉₉₎ displays an extended and elongated conformation (**Figure 19D**).

Table 11: Summary of SAXS-derived parameters for several PEX19 constructs and their complexes with PEX3.

	PEX3 C235S (41-373)	PEX19 (1-299)	PEX19 (161-283)	PEX19 (54-291)	PEX19 Δ(62-148)	PEX3 C235S (41-373)	PEX3 C235S (41-373)	PEX3 C235S (41-373)	PEX3 C235S (41-373)
						PEX19 (1-299)	PEX19 (161-283)	PEX19 (54-291)	PEX19 Δ(62-148)
Data collection parameters									
Instrument	EMBL P12 (PETRA-III)								
Beam geometry (mm) ²	0.2 x 0.12								
Wavelength (Å)	1.24								
s range (Å ⁻¹) ^a	0.003-0.45								
Exposure time (s)	0.045-20								
Sample concentration (mg/mL)	0.5-5.0								
Temperature (K)	283								
Structural parameters									
I ₀ (0) (cm ⁻¹) [from P(r)]	10760 ±39.3	10480 ±118.4	225.9 ±3.0	695.3 ±7.2	581.7 ±8.87	21310 ±235	1816 ±15.88	2596 ±26.48	2209 ±12.56
I ₀ (0) (cm ⁻¹) [from Guinier]	10718.9 0 ±35.7	9914.37 ±29.75	256.85 ±4.02	716.12 ±7.89	592.98 ±8.59	20002.40 ±81.7	1803.55 ±9.45	2467.67 ±12.31	2213.07 ±13.25
R _g (nm) [from P(r)]	2.85 ±0.02	4.65 ±0.12	1.73 ±0.09	3.27 ±0.04	3.71 ±0.06	5.39 ±0.15	2.48 ±0.04	4.14 ±0.09	4.26 ±0.02
R _g (nm) [from Guinier]	2.82 ±0.40	3.96 ±0.02	1.76 ±0.13	3.34 ±0.42	3.61 ±0.39	4.44 ±0.1	2.41 ±0.02	3.56 ±0.04	4.13 ±0.05
D _{max} (nm)	8.3	13.2	4.8	10.2	11.83	19	10.85	17.34	14.34
Porod volume estimate (nm ³)	66.41	68.62	12.95	29.2	39.5	130.86	48.07	56.48	106.25
Molecular mass MW (kDa) [from Porod volume (V /1.7)]	39.06	40.36	7.6	17.1	23.23	76.97	28.3	33.2	62.5
Calculated monomeric MW(kDa) ^b	38	32	14	26.5	24.5	70	46	58.5	56.5
Software employed									
Primary data reduction	Automated pipeline (Franke et al., 2012)								
Data processing	PRIMUS								
<i>Ab initio</i> analysis	DAMMIF, GASBOR, EOM, CORAL, BUNCH								
Validation and averaging	DAMMIN, DAMAVER								
^a Momentum transfer $ s = 4\pi\sin(\theta)/\lambda$.									
^b calculated from the sequence using the http://www.basic.northwestern.edu/biotools/proteincalc.html server.									

By using the ATSAS suite programme BUNCH to perform *ab initio* modelling for PEX19₍₁₋₂₉₉₎ via simulated annealing, it was possible to determine its three-dimensional domain model based on the experimental scattering curve and the existing crystal structure of its C-terminus (residues 161-283, PDB entry 2WL8). The fit of the BUNCH-derived models to the experimental data had a fit of $\chi^2=0.8$, indicating that despite the unfolded nature of the missing N-terminal region, the general conformation adopts the same elongated arrangement (**Figure 20A**).

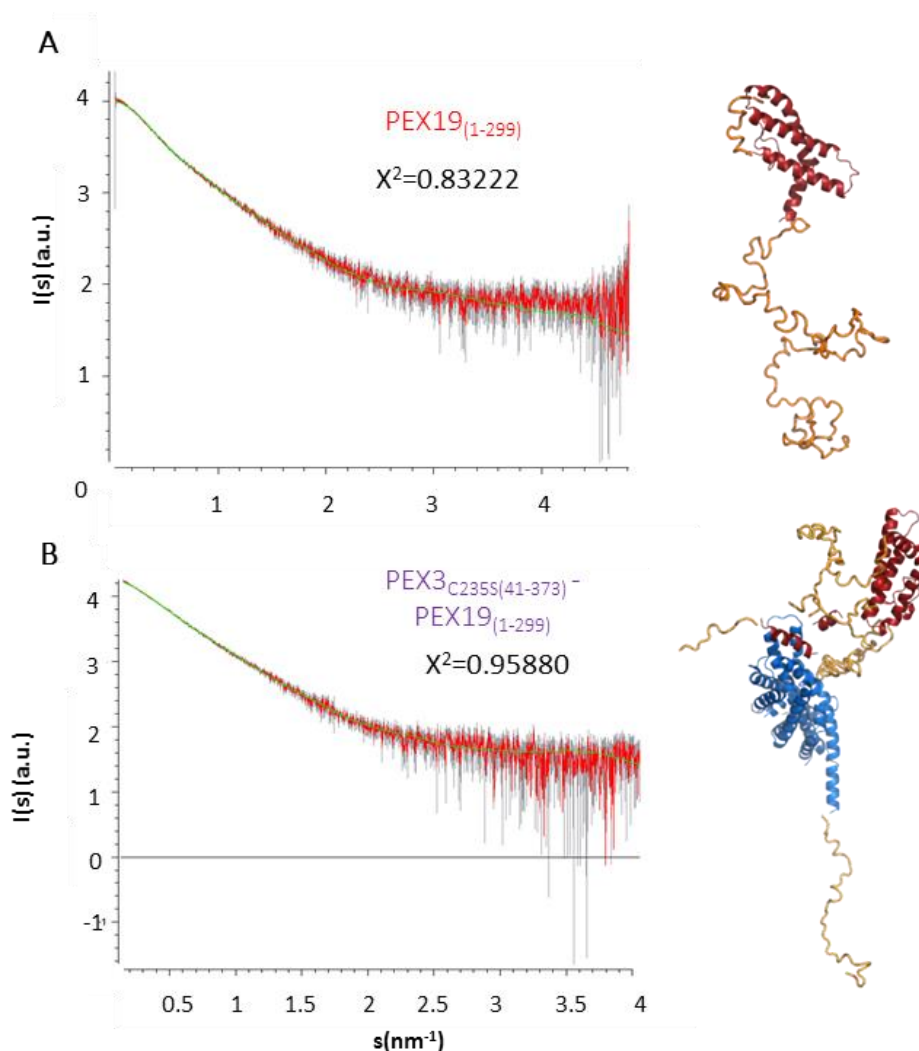


Figure 20: *Ab initio* models of PEX19₍₁₋₂₉₉₎ and the PEX3_{C235S(41-373)} - PEX19₍₁₋₂₉₉₎. (A) BUNCH was used for *ab initio* modelling of PEX19₍₁₋₂₉₉₎ (red, unstructured loops in orange) with the fit shown in the left panel. (B) CORAL was used to model the complex (PEX19₍₁₋₂₉₉₎ shown in red, PEX3_{C235S(41-373)} shown in blue, and unstructured regions in orange) with the fit shown at the left panel. The experimental data is coloured red, while the fit is in green.

Additionally, the programme CORAL was used to model the full-length complex of PEX3_{C235S(41-373)} - PEX19₍₁₋₂₉₉₎, using structures 3MK4 and 2WL8, respectively. Distance restraints of 5 Å were imposed for the PEX19 region 101-110, based on our finding of the secondary PEX19 binding site, which would disallow the free movement of this helix, while keeping it proximal to its binding site on PEX3. The model produced had a good fit to the experimental data and indicated that although the C-terminal PEX19 domain was able to assume any position, it remained in close proximity to PEX3 (**Figure 20B**).

Moreover, the ATSAS programme EOM was used to find plausible models that would explain the flexibility of the PEX3_{C235S(41-373)} - PEX19₍₁₋₂₉₉₎ complex, by fitting an averaged theoretical scattering intensity derived from an ensemble of conformations to the experimental SAXS data. Four conformations displayed acceptable fits (**Figure 21**), all of which were in agreement with the Kratky analysis, showing that this complex is neither fully flexible/unfolded, nor globular, but instead can be described as a multi-domain flexible system that tends towards some compactness.

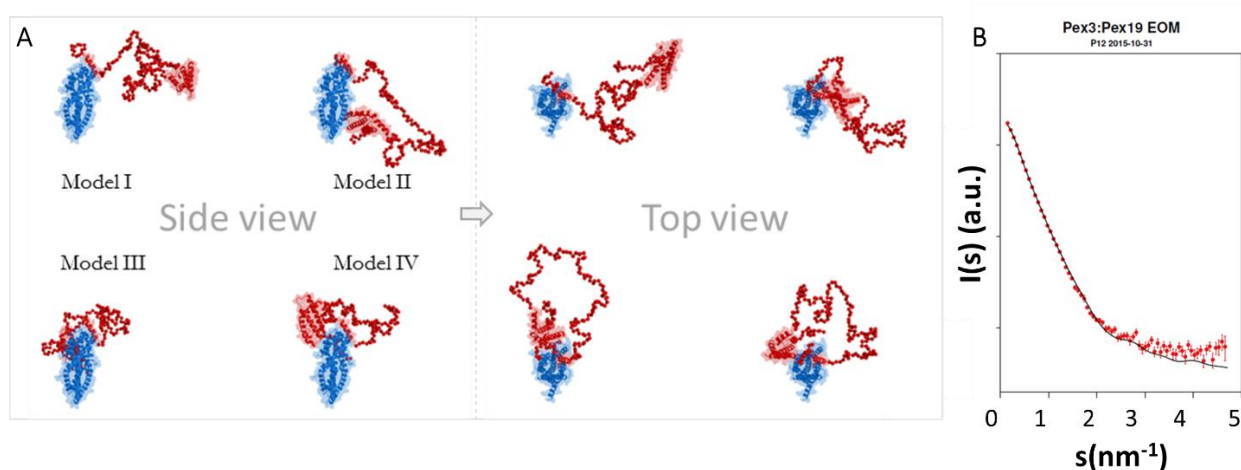


Figure 21: EOM models of PEX3_{C235S(41-373)} - PEX19₍₁₋₂₉₉₎. (A) The complex forms a flexible multi-domain system, with Model III being the most probable (50% of the conformations). PEX3_{C235S(41-373)} is indicated in blue, while PEX19₍₁₋₂₉₉₎ is depicted in red. (B) The fit of EOM model III (black) to the experimental data (red).

3.1.1.4. Crystallisation of PEX3-PEX19 constructs

Several PEX3-PEX19 constructs were used for crystallisation and are summarised in **Table 12**. Due to the high flexibility and disorder of the PEX19 N-terminus, crystallisation trials were not successful for most constructs, even after deletion of the N-terminal 50-149 disordered region.

Table 12: List of PEX19 constructs used for crystallisation with PEX3_{C235S(41-373)}

Construct Name	Crystallisation conditions	Crystals
PEX3 _{C235S(41-373)} - PEX19 ₍₁₋₂₉₉₎	-	No
PEX3 _{C235S(41-373)} - PEX19 _{Δ(62-148)}	-	No
PEX3 _{C235S(41-373)} - PEX19 _{pep1} - PEX19 _{pep2}	QIAGEN_ProComplex_C12	Single crystals, needles
PEX3 _{C235S(41-373)} - PEX19 _{pep2}	QIAGEN_Classics_I_H04	Needles

Crystallisation trials of PEX3_{C235S(41-373)} with PEX19 peptides PEX19_{pep1} and PEX19_{pep2} or PEX3_{C235S(41-373)}- PEX19_{pep2} yielded crystals of different macroscopic properties that appeared within 3 to 5 days (**Figure 22**). Optimisation of the crystals was performed and attempts were made for diffraction experiments on the ESRF ID30A-1 / MASSIF-1 beamline, but the diffraction data was very weak in both cases. Additionally, PEX3_{C235S(41-373)}- PEX19_{pep2} crystals produced diffraction images containing overlapping multiple lattices, making data processing impossible.

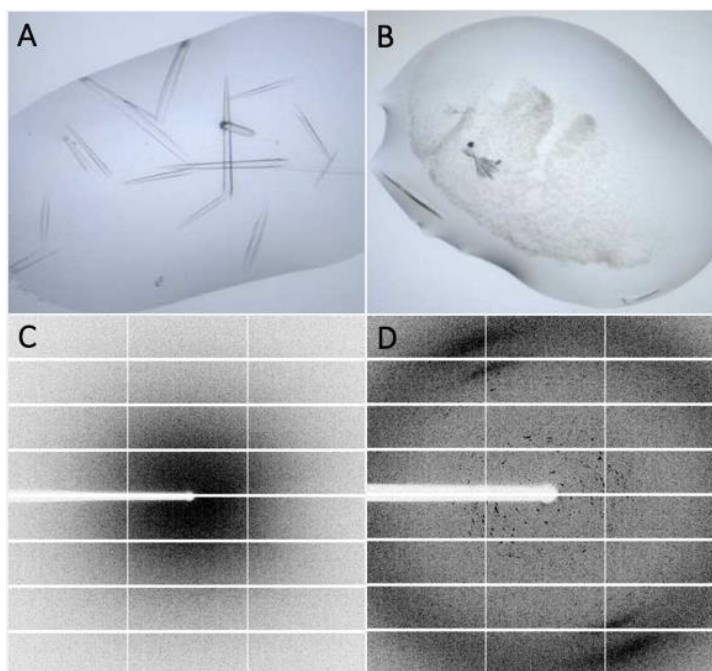


Figure 22: Crystals of PEX3_{C235S(41-373)} with PEX19 peptides. (A) PEX3_{C235S(41-373)} - PEX19_{pep1}- PEX19_{pep2} crystals, (B) PEX3_{C235S(41-373)}- PEX19_{pep2} crystals. (C-D) Diffraction patterns of the above crystals of A and B respectively. No diffraction was observed in the case of PEX3_{C235S(41-373)} - PEX19_{pep1}- PEX19_{pep2}, while weak diffraction was observed in the case of PEX3_{C235S(41-373)}- PEX19_{pep2} crystals.

3.1.2. Structural and biophysical characterisation of the human PEX19-PEX14 complex

In this thesis it is demonstrated that co-expression of full-length PEX19₍₁₋₂₉₉₎ and PEX14₍₁₋₃₇₇₎ can be used to produce small but significant amounts of the complex, further indicating that despite PEX14 being an unusual PMP, PEX19 acts as its chaperone, shielding it and maintaining its solubility in the absence of detergents.

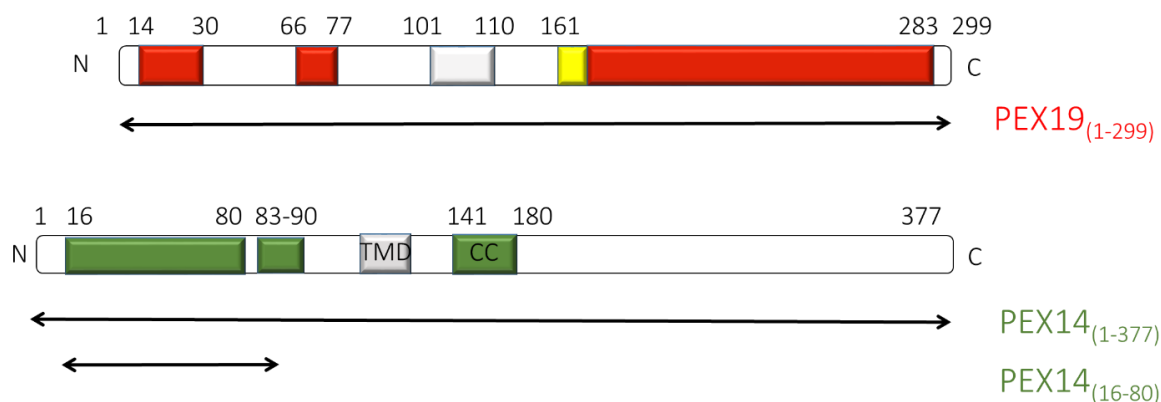


Figure 23: List of PEX19 and Pex14 constructs used. The PEX19 construct has been previously described (Figure 15). The green rectangles on PEX14 correspond to its conserved N-terminal domain (16-80), its PEX13 binding site (83-90) and coiled-coil region (CC, 141-180). The grey rectangle indicates the predicted transmembrane domain (TMD) of PEX14.

3.1.2.1. Expression, purification and biophysical characterisation of human PEX19-PEX14

Individual expression and purification of PEX14₍₁₋₃₇₇₎ resulted in insoluble and aggregating protein, as an effect of its hydrophobicity. However, upon co-expression with PEX19₍₁₋₂₉₉₎, it was possible to isolate a small but sufficient amount of PEX14₍₁₋₃₇₇₎ in complex with PEX19 (Figure 24A). The two proteins were expressed in *E. coli* cells using auto-induction media and the complex was purified with affinity chromatography, followed by two steps of SEC. The first step of SEC served to isolate the complex from the excess of unbound PEX19 (Figure 24B), while the second SEC was performed to ensure purity and mono-dispersity of the PEX19-PEX14 complex (Figure 24D).

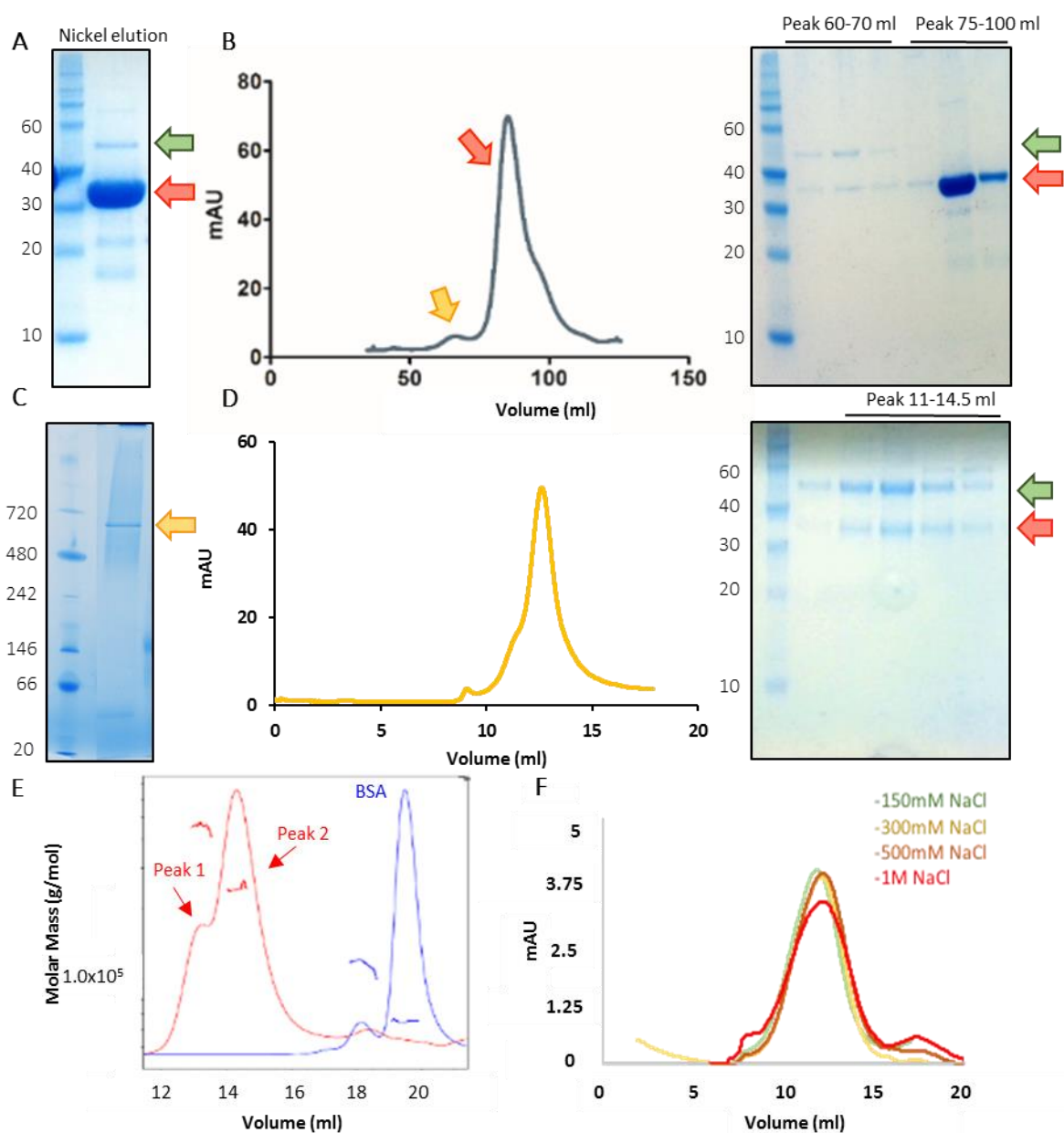


Figure 24: Purification of PEX19₍₁₋₂₉₉₎-PEX14₍₁₋₃₇₇₎. (A) SDS-PAGE analysis of nickel-affinity purified PEX19₍₁₋₂₉₉₎ (red arrow)-PEX14₍₁₋₃₇₇₎ (green arrow) complex (B) SEC elution from a Hiprep Superdex 200 16/60 GL column with its corresponding SDS-PAGE. The yellow arrow indicates the PEX19₍₁₋₂₉₉₎-PEX14₍₁₋₃₇₇₎ complex, while the red arrow indicates the excess PEX19₍₁₋₂₉₉₎ (C) Blue native PAGE of the PEX19₍₁₋₂₉₉₎-PEX14₍₁₋₃₇₇₎ complex (yellow arrow). (D) SEC elution from a Superose 6 increase 10/300 GL column (yellow) with its corresponding SDS-PAGE. (E) SEC-MALS characterisation of the complex (red) and the standard BSA (blue). Two major peaks are visible for the complex, indicating two oligomeric species corresponding to 457 kDa \pm 2.9% (Peak 1) and 245.8 kDa \pm 8.1% (Peak 2) respectively. (F) SEC was performed for the complex with increasing NaCl concentrations, but no change was observed, displaying its stability.

In contrast with the membrane protein PEX14₍₁₋₃₇₇₎ that is challenging to purify, PEX19₍₁₋₂₉₉₎ is a very soluble protein that can be purified in large excess of the other protein (**Figure 24A, B**). PEX14₍₁₋₃₇₇₎ on the other hand, is always purified in association with PEX19₍₁₋₂₉₉₎ and never individually, indicating that it requires PEX19 for becoming soluble. Purification of the complex yielded comparatively small protein amounts (1.5 mg per litre of expression) and aggregation is caused upon high concentration (>5 mg/ml). Furthermore, the PEX19₍₁₋₂₉₉₎-PEX14₍₁₋₃₇₇₎ complex forms higher oligomers. This, in addition to the flexibility of both of its components, cause the elution profile of the complex to corresponds to higher molecular weight proteins, something that was also observed in blue native electrophoresis (**Figure 24C**). These oligomers are very stable to ionic strength variation and cannot be separated (**Figure 24F**).

As the stoichiometry of the complex is not known and was difficult to determine by its size exclusion profile, native Mass Spectrometry (native-MS) was used (in collaboration with Dr. F. Drepper and J. Bender, Warscheid group, Freiburg). The native-MS results indicated a heterotrimeric complex of 108 kDa that contains two molecules of PEX19₍₁₋₂₉₉₎ and one of PEX14₍₁₋₃₇₇₎ (**Figure 25**).

Additionally, in combination with SEC-MALS data that showed two major oligomeric species of 457 kDa \pm 2.9% and 245.8 kDa \pm 8.1% (**Figure 24E**), we can deduce that the observed oligomers consist of two (216 kDa) or four (432 kDa) heterotrimers, respectively. The formation of the heterotrimer is presumed to rely on the coiled-coil region of PEX14₍₁₋₃₇₇₎ (Will *et al*, 1999; Otera *et al*, 2002; Itoh & Fujiki, 2006), ranging residues 141-180, downstream of its putative transmembrane domain.

Similarly, the small N-terminal region of PEX14₍₁₆₋₈₀₎ was also expressed and purified in large amounts (60 mg per litre of culture). Co-purification with PEX19₍₁₋₂₉₉₎ also displayed a 2:1 stoichiometry of two PEX19 molecules binding to one PEX14 (data not shown), but native-MS analysis indicated only a 1:1 hetero-dimer (**Figure 25**).

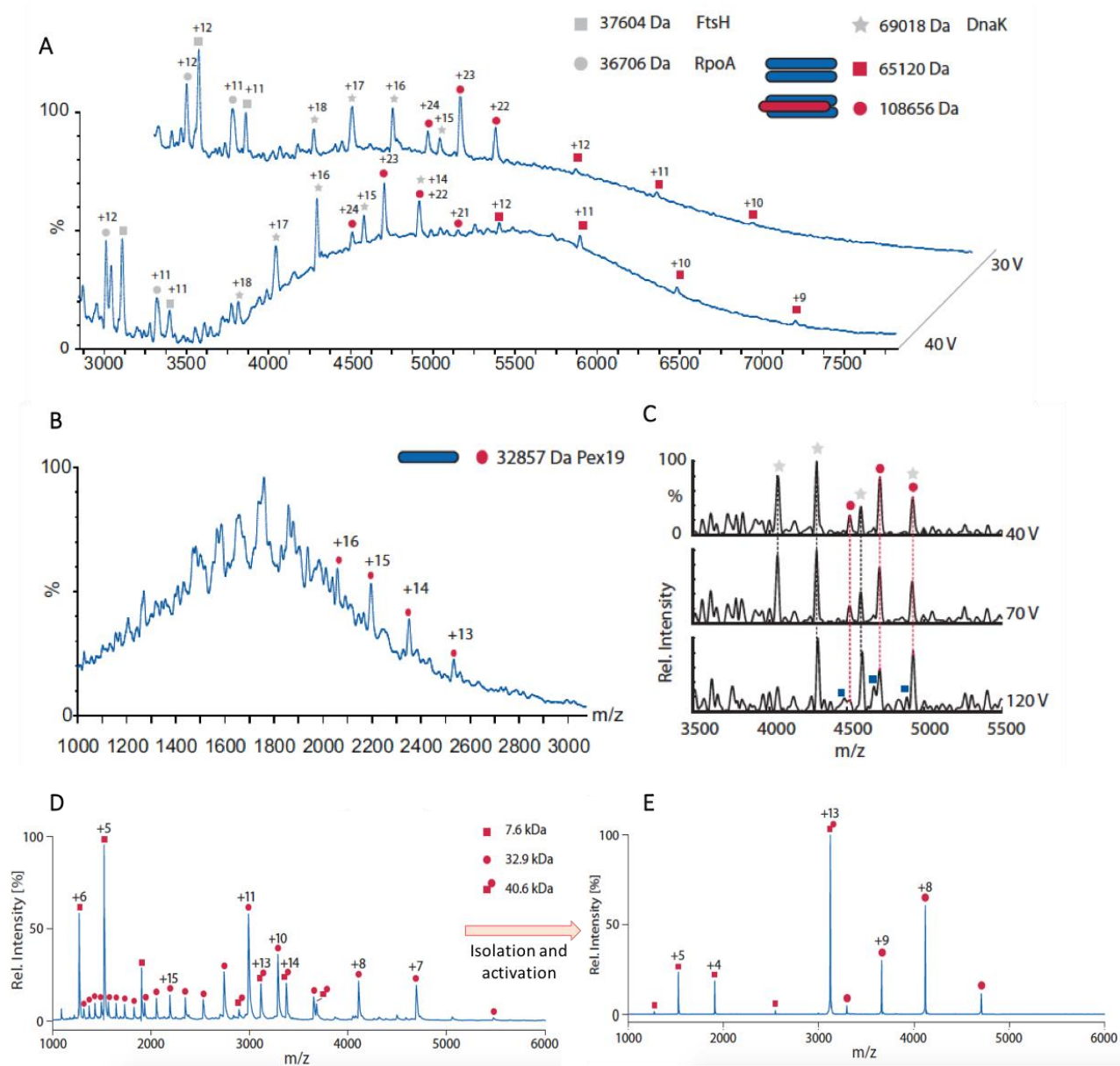


Figure 25: Native MS analysis indicates a 108 kDa complex. (A) Spectra at 30-40V are indicative of masses of 36.7 kDa, 37.6 kDa and 68 kDa (likely *E. coli* contaminants), as well as 65 kDa and 108 kDa. The last two masses correspond to dimeric PEX19₍₁₋₂₉₉₎ and a hetero-trimer with a 2:1 PEX19₍₁₋₂₉₉₎: PEX14₍₁₋₃₇₇₎ stoichiometry, respectively, (B) Higher collisional activation of 120V reveals a low-*m/z* peak series corresponding to a mass of 32.5 kDa, indicative of PEX19₍₁₋₂₉₉₎ monomers. (C) With increasing collisional activation energy, new sub-peaks (blue squares, corresponding to a mass of 107.5 kDa) arise for the peak series assigned as 108 kDa complex. (D) Native-MS analysis of PEX19₍₁₋₂₉₉₎- PEX14₍₁₆₋₈₀₎ indicates a 1:1 complex (shown in the range between *m/z* 2900 and 4000). Furthermore, both monomers were identified: PEX14₍₁₆₋₈₀₎, as well as PEX19₍₁₋₂₉₉₎ in a broad charge state series. (E) Upon isolation and collisional activation of the +13 charged ion of the putative 1:1 complex, the spectrum consists of a peak series in the high *m/z* range indicating a mass of 32.9 kDa and corresponding to monomers of PEX19₍₁₋₂₉₉₎ as well as a series in the low *m/z* range corresponding to the 7.6 kDa MW of PEX14₍₁₆₋₈₀₎.

3.1.2.2. Identification of secondary PEX19-PEX14 binding sites via XL-MS

Chemical crosslinking coupled with mass spectrometry (XL-MS) was employed in order to identify novel binding sites between PEX19 and PEX14. The same protocol was followed for two complexes (PEX19₍₁₋₂₉₉₎ - PEX14₍₁₋₃₇₇₎ and PEX19₍₁₋₂₉₉₎ - PEX14₍₁₆₋₈₀₎) using the homo-bifunctional crosslinker BS³, which can link primary amine side chains of lysine residues (K) that are in close proximity (crosslinker spacer arm: 11.4 Å).

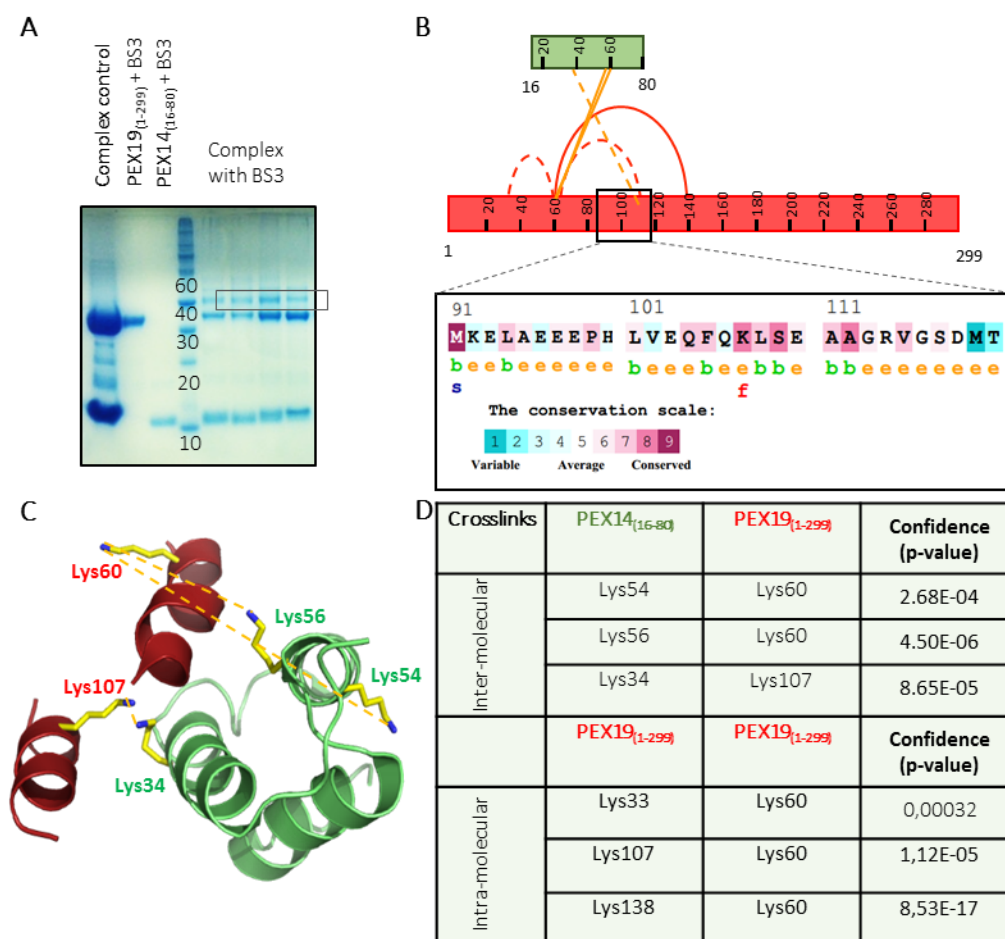


Figure 26: Crosslinking of PEX19₍₁₋₂₉₉₎ - PEX14₍₁₆₋₈₀₎. (A) SDS-PAGE analysis of the crosslinked complex with crosslinker BS³. The crosslinked product band is indicated with a black box. (B) Illustration of the intramolecular PEX19₍₁₋₂₉₉₎ crosslinks (red) as well as the intermolecular PEX19₍₁₋₂₉₉₎ - PEX14₍₁₆₋₈₀₎ (yellow). The sequence corresponding to the boxed area in PEX19₍₁₋₂₉₉₎ where the majority of crosslinks is concentrated is shown in the inset, coloured by conservation score (ConSurf server). The letters b, e, f and s, correspond to buried, exposed, functional and structurally important residues respectively. (C) Model displaying the intermolecular PEX19₍₁₋₂₉₉₎ - PEX14₍₁₆₋₈₀₎ crosslinks. The crosslinked residues are indicated with sticks. (D) Table summarising the crosslinked residues that were detected by MS analysis.

Crosslinked complexes were identified only in the case of PEX19₍₁₋₂₉₉₎ - PEX14₍₁₆₋₈₀₎ (**Figure 26**), while no crosslinks appeared in the individual PEX19₍₁₋₂₉₉₎ and PEX14₍₁₆₋₈₀₎ control samples, indicating that these two proteins do not homo-oligomerise. This is to be expected for both PEX19, which is in principle monomeric, and PEX14₍₁₆₋₈₀₎, which is missing its putative dimerisation-capable coiled-coil region. Mass spectrometry analysis revealed crosslinks for the already known binding site between highly conserved PEX14-K54/K56 and PEX19-K60, as determined by NMR (PDB code: 2W85). In addition to that abundantly crosslinked site, there was indication of a novel crosslinked region between the conserved and predicted functionally important PEX19-K107 with PEX14-K34 located upstream of the highly conserved and structurally important F35 (**Figure 27**). It is important to note that this is the very same binding site discovered for PEX3 binding (Chapter 3.1.1.2), suggesting that it is shared between PEX3 and PEX14.

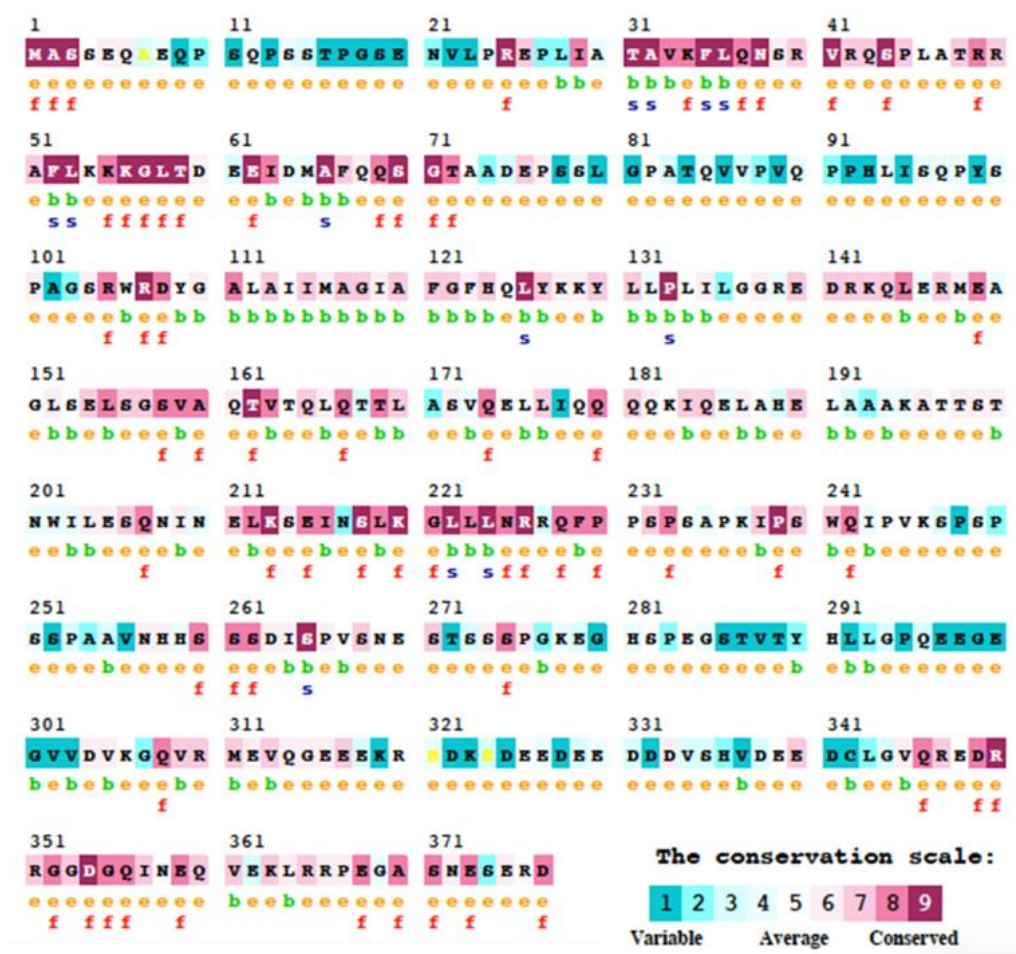


Figure 27: Sequence conservation of human PEX14, as illustrated by the ConSurf server. The letters b, e, f and s, correspond to buried, exposed, functional and structurally important residues respectively.

3.1.2.3. Structural characterisation of PEX19-PEX14 via negative stain EM

The PEX19₍₁₋₂₉₉₎ - PEX14₍₁₋₃₇₇₎ complex was further characterised by negative stain electron microscopy (in collaboration with Pascal Lill, Dr. Bjoern Klink and Dr. Christos Gatsogiannis, Max Planck Institute, Dortmund). Low resolution characterisation *via* negative stain EM showed that this complex forms elongated particles of 20 nm as observed after preparation on the negative stain grid (**Figure 28**).

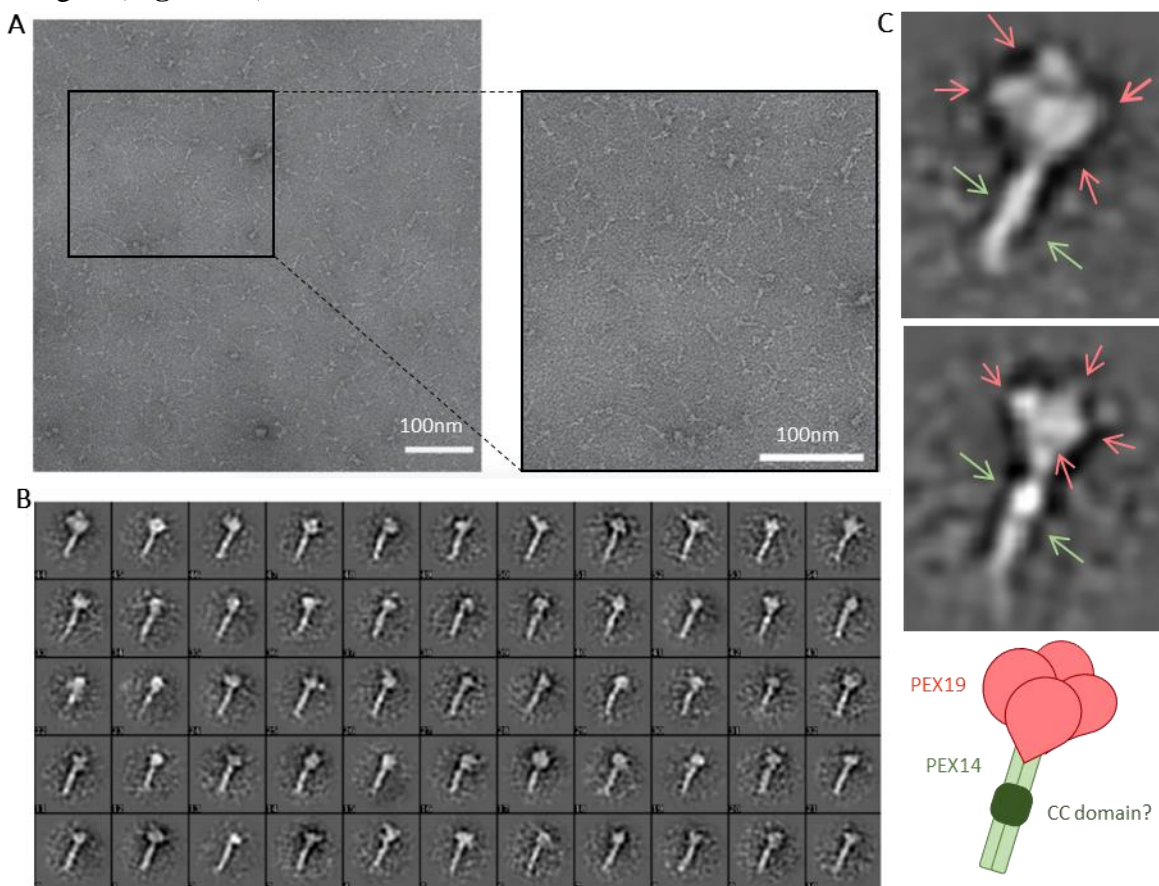


Figure 28: Negative stain EM of PEX19₍₁₋₂₉₉₎ - PEX14₍₁₋₃₇₇₎. (A) Electron micrograph of the PEX19₍₁₋₂₉₉₎ - PEX14₍₁₋₃₇₇₎ complex. The inset presents a magnified region of the micrograph where particles are more clearly visible. (B) Class averages of the complex, which can be described as elongated with extensions protruding from its top region. (C) Magnified class averages of the PEX19₍₁₋₂₉₉₎ - PEX14₍₁₋₃₇₇₎ complex. The potential CC domain of PEX14 is indicated with green arrows, while PEX19 with salmon. Graphical representation of the EM particles is shown in the right, with four globular extensions, likely PEX19 (salmon), and the stem region, which is likely PEX14 (green). The brighter region of the stem potentially corresponds to the CC domain (dark green).

Class averages of the initial particles provided clearer information into their structure. The particles consist of an elongated stem region and four globular extensions on one end, which, taking the native MS data into account, likely correspond to PEX14 and the C-termini of PEX19,

respectively. An optically brighter region on the stem, could indicate the coiled-coil domain where PEX14 dimerisation can occur. Due to the elongated shape of this particle, multiple orientations were not observed, making a possible three-dimensional reconstruction challenging, but further analysis by negative stain EM and cryo-EM is ongoing.

3.1.2.4. *Low resolution structural characterisation of PEX19-PEX14 via SAXS suggests flexible elongated structure*

SAXS was used to analyse and provide low-resolution structural information on the PEX19₍₁₋₂₉₉₎ - PEX14₍₁₋₃₇₇₎ complex. The acquired SAXS data did not exhibit any aggregation effects and the measurements of a concentration series of the complex were merged and scaled. The experimental scattering curve is shown in **Figure 29A**. The values of the radius of gyration were computed from the Guinier plot and the corresponding D_{\max} was evaluated from probable real-space atom-pair distance distribution, $p(r)$ vs r (**Table 13**).

Table 13: Summary of SAXS-derived parameters for the PEX19₍₁₋₂₉₉₎ - PEX14₍₁₋₃₇₇₎ complex

	PEX19 ₍₁₋₂₉₉₎ -PEX14 ₍₁₋₃₇₇₎
Data collection parameters	
Instrument	EMBL P12 (PETRA-III)
Beam geometry (mm ²)	0.2 x 0.12
Wavelength (Å)	1.24
s range (Å ⁻¹) ^a	0.003-0.45
Exposure time (s)	0.045-20
Sample concentration (mg/mL)	0.5-5.0
Temperature (K)	283
Structural parameters	
Io (0) (cm ⁻¹) [from P(r)]	17710 ±114.5
Io (0) (cm ⁻¹) [from Guinier]	17437.70 ±80.16
R _g (nm) [from P(r)]	10.4 ±0.11
R _g (nm) [from Guinier]	9.86 ±0.11
D _{max} (nm)	40
Porod volume estimate (nm ³)	1216
Molecular mass MW (kDa) [from Porod volume (V _p /1.7)]	715.3
Calculated monomeric MW (kDa) ^b	73
Software employed	
Primary data reduction	Automated pipeline (Franke et al., 2012)
Data processing	PRIMUS
Ab initio analysis	DAMMIF, GASBOR, CORAL
Validation and averaging	DAMMIN, DAMAVER
^a Momentum transfer $ s = 4\pi\sin(\theta)/\lambda$.	
^b calculated from the sequence using the http://www.basic.northwestern.edu/biotools/proteincalc.html server.	

The obtained parameters are consistent with a solution of elongated monomeric particles, which exhibit a large degree of flexibility, as can be assessed by the Kratky plot (**Figure 29B**). Kratky analysis further indicates a multi-domain system, which is in agreement with the multiple peaks visible in the distance distribution function (**Figure 29C**). The distance distribution profile has a defined peak centred between 5-10 nm that could indicate a globular domain, and a secondary overlapping asymmetric peak between 10-40 nm leading to an elongated tail indicating a rod-like feature of the structure.

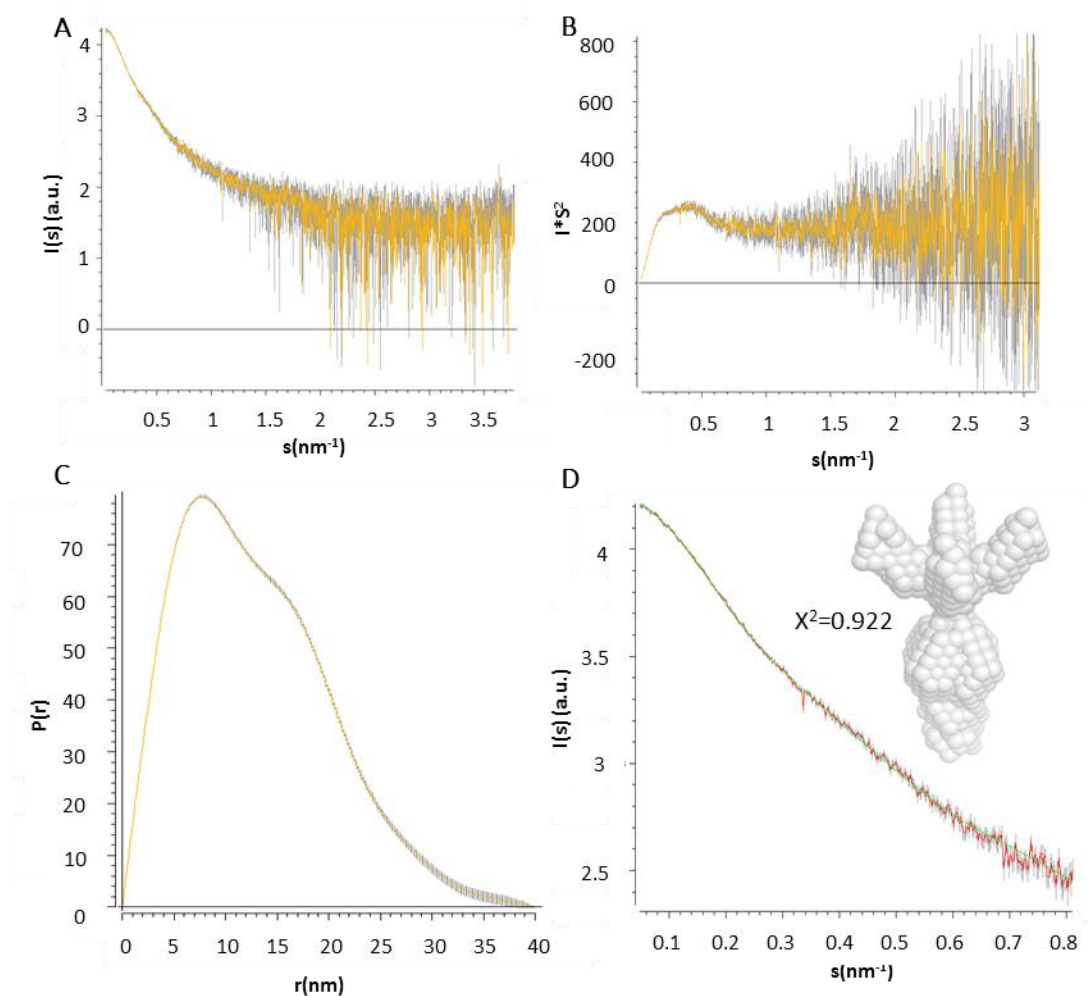


Figure 29: SAXS analysis of the PEX19₍₁₋₂₉₉₎ - PEX14₍₁₋₃₇₇₎ complex. (A) Scaled plotted intensity of the SAXS measurements for the complex (yellow), (B) Kratky plot illustrating the degree of compactness. (C) Distance distribution function indicating a multi-domain elongated conformation. (D) *Ab initio* model and fit to the experimental data using the programme DAMMIF with imposed $P4$ symmetry.

Ab initio modelling using the ATSAS suite software DAMMIF with no symmetry restrictions produced a number of elongated structures, which however, did not fit well with the observed

negative stain EM particles (**Figure 28**). Taking the features of the particles observed by EM and the stoichiometry derived from native-MS, several types of symmetry was imposed, with the P4 models producing very good fits to the experimental data ($\chi^2= 0.922$) (**Figure 29D**). The P4 DAMMIF models display an elongated stem region, followed by four extensions at the upper region, which visibly correlate with the negative stain EM particle structure.

3.1.2.5. Crystallisation of PEX19-PEX14 constructs

In order to obtain high-resolution structural information on the interaction between PEX19 and PEX14, crystallisation trials were performed for the constructs summarised in **Table 14**.

Table 14: List of PEX19 constructs used for co-crystallisation with PEX14

Construct Name	Crystallisation conditions	Crystals
PEX19 ₍₁₋₂₉₉₎ - PEX14 ₍₁₋₃₇₇₎	QIAGEN_Classics_I_A03	Needle bunches
PEX19 ₍₁₋₂₉₉₎ - PEX14 ₍₁₆₋₈₀₎	-	No
PEX19 _{pep2} - PEX14 ₍₁₄₋₈₀₎	-	No

Unfortunately, these crystals did not produce good quality single crystals that could be used for X-ray diffraction experiments (**Figure 30**). Optimisation of the complex crystals was performed, but the crystal morphology and diffraction could not be improved.

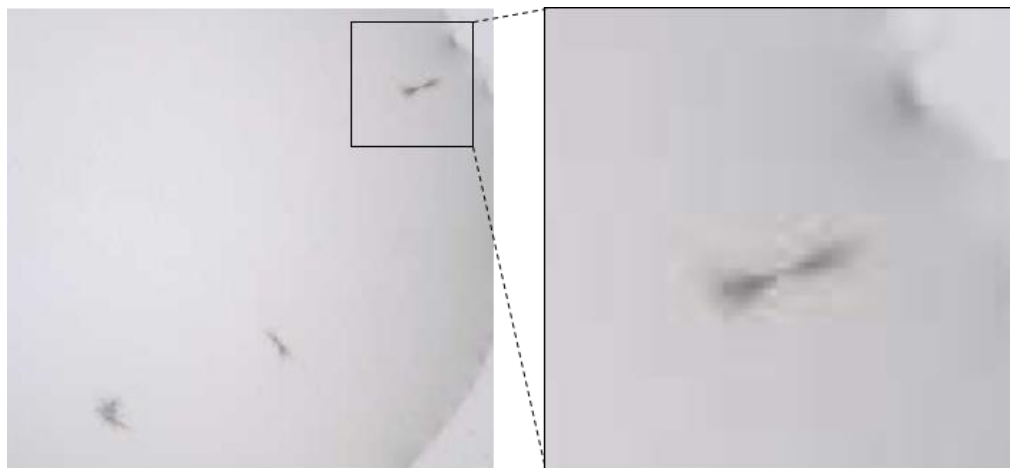


Figure 30: PEX19₍₁₋₂₉₉₎ - PEX14₍₁₋₃₇₇₎ crystallisation. Needle-like crystals appeared within one month of crystallisation.

3.1.3. Interaction of human PEX19 with PMPs

In addition to its interaction with PEX14, human PEX19 is also known to interact with a variety of PMPs, including integral membrane proteins PEX13 and PEX11 β , involved in peroxisome protein import and proliferation, respectively, as well as ACBD5, a tail-anchored protein localising to peroxisomes. The affinity of the interaction of the C-terminus of PEX19 (residues 161-283) with PMP peptides of PEX13 and PEX11 β has been quantified as 8.5 and 23.4 μ M, respectively (Schueller *et al*, 2010). Moreover, extensive research has been performed recently on the interaction of PEX19 with the tail-anchored protein ACBD5 (Costello *et al*, 2017). According to *in vivo* and *in vitro* data, PEX19 is responsible for the recognition of the mPTS domain of ACBD5 located at its C-terminus and is dependent on the existence of a transmembrane domain and a short luminal charged residue tail. In this part of the project, the interaction of PEX19 with various PMP peptides was explored (Table 15).

Table 15: List of peptides used

Peptide Name	Sequence
PEX11 β _{pep}	(FITC)-LALKLRLQVLLLARV
PEX13 _{pep}	(FITC)-FTKVFSALVVRTIR
ACBD5 _{wt}	(FITC)-SPGVLTFAIWPPIAQWLVYLYYQRRRRKL
ACBD5 _{mut1}	(FITC)-SPGVLTFAIWPPIAQWLVYLYYQRARAKL
ACBD5 _{mut2}	(FITC)-SPGVLTFAIWPPIAQWLVYLYYQAAAAL

3.1.3.1. The interaction of PEX19 with ACBD5 depends on the charge of its C-terminus

The C-terminus of ACBD5 contains the necessary information for its intracellular localisation. Its mPTS signal contains a transmembrane helix, followed by an array of luminal polar amino acids, which have been shown to be essential for peroxisomal targeting. Substitution of those polar residues with non-polar residues such as alanine, can result in differential targeting of ACBD5 to other organellar membranes such as mitochondria or the ER (Costello *et al*, 2017). This can be explained by the difference in the affinity of the mutants for PEX19, which is responsible for their peroxisomal membrane import.

Three ACBD5 C-terminal peptides (**Figure 31A**) were tested for their binding with PEX19₍₁₋₂₉₉₎ by means of fluorescence anisotropy polarisation and exhibited distinct binding affinity for it, which decreased with increasing number of mutations (**Figure 31B, C**). This correlates well with the *in vivo* localisation results which showed (1) peroxisomal targeting for ABCD5_{wt}, (2) partial peroxisomal and mitochondrial targeting for ABCD5_{mut1} and (3) ER-targeting for ABCD5_{mut2}.

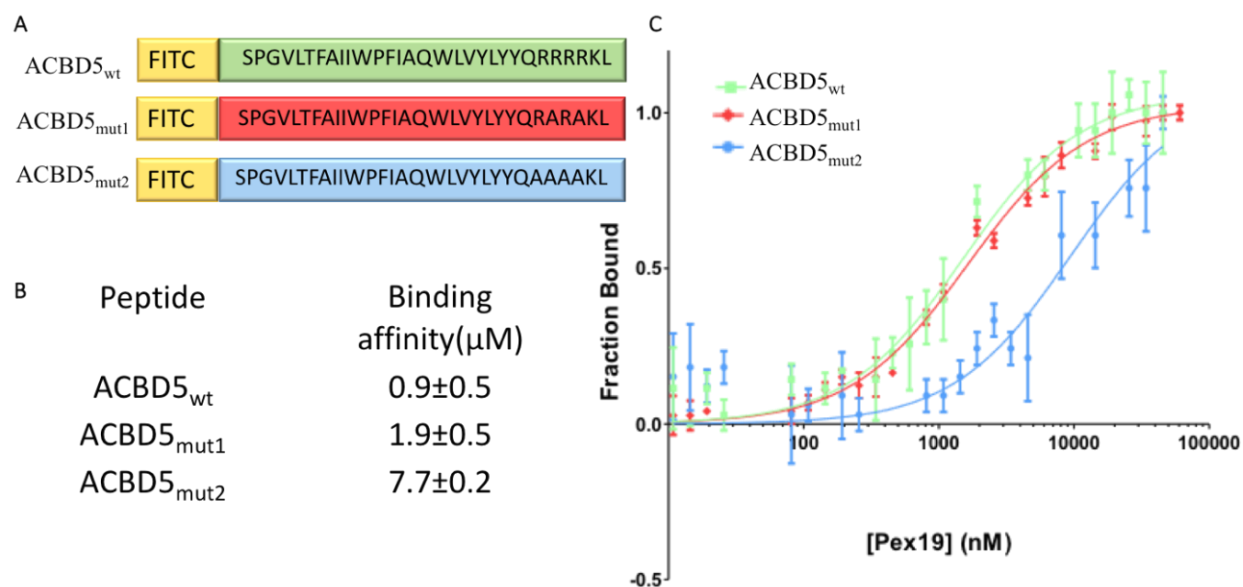


Figure 31: Binding studies of PEX19 with fluorescently labelled PMP peptides. (A) Schematic overview of peptides used for the fluorescence anisotropy assays. (B) Quantified binding affinities, and (C) Fluorescence anisotropy measurements of ACBD5 wild-type and mutant peptides.

3.1.3.2. Crystallisation of PEX19 with PMP peptides

Several constructs of PEX19 were complexed with PMP peptides in presence and absence of PEX3 (summarised in **Table 16**). However, PMP peptides are highly hydrophobic and insoluble in aqueous buffers, resulting in precipitation and making crystallisation of the complexes difficult.

Table 16: List of PEX19 and PEX3 constructs used for crystallisation with PMP peptides

Construct Name	Crystals
PEX19 ₍₁₆₁₋₂₈₃₎ -PEX11β _{pep}	No
PEX19 ₍₁₆₁₋₂₈₃₎ -PEX13 _{pep}	No
PEX19 ₍₁₆₁₋₂₈₃₎ -ACBD5 _{wt}	No
PEX19 _{Δ(62-148)} - PEX3 _{C235S(41-373)} -PEX11β _{pep}	No
PEX19 _{Δ(62-148)} - PEX3 _{C235S(41-373)} - PEX13 _{pep}	No
PEX19 _{Δ(62-148)} - PEX3 _{C235S(41-373)} - ACBD5 _{wt}	No

3.2. Structural and biophysical characterisation of PTS1 import proteins

Peroxisomal PTS1 import is a process mediated by the cycling receptor Pex5, which ensures the import of enzymes that are required for proper peroxisomal function. Two cargo proteins, Pcs60 from yeast, and MIF1 from the plant *A. thaliana* were selected as model peroxisomal cargoes. Pcs60 is an enzyme involved in oxalic acid metabolism, the structure of which was determined in this study *via* X-ray crystallography. As Pcs60 is a PTS1-containing protein, it can interact via its C-terminus with the receptor Pex5p, an affinity that has been previously quantified to be 0.19 μ M (Hagen *et al*, 2015). In this study, we performed structural characterisation of the Pex5p-Pcs60 complex by means of SAXS, X-ray crystallography and negative stain EM, in order to further aid our understanding of the molecular interaction between Pex5p and Pcs60, and delineate the complicated process of peroxisomal matrix protein import. The protocols for expression, purification and crystallization of Pcs60 and its complexes with Pex5p, were initially optimised by Dr. D. Passon and Dr. N. Hanna (EMBL-Hamburg). Lastly, MIF1, a peroxisomal matrix protein from plants that has been implicated in stress response, was also structurally characterised by X-ray crystallography.

3.2.1. Structural characterisation of Pcs60

3.2.1.1. *Expression, purification and characterisation of Pcs60*

Pcs60 is a soluble 60 kDa peroxisomal protein which can be expressed in relatively high amounts (6mg per litre of expression) in high purity (**Figure 32A**). Pcs60 can form dimers and higher oligomers in solution, a concentration-dependant tendency, made visible by its SEC elution profile which corresponds to the size of a dimeric protein, as well as crosslinking experiments in which the dimer appears as a 120 kDa and a >150 kDa band, prominent on SDS-PAGE analysis (**Figure 32B**). Pcs60 is fairly soluble and stable with a melting temperature of \sim 53 $^{\circ}$ C, as indicated by Thermofluor analysis (**Figure 32C**).

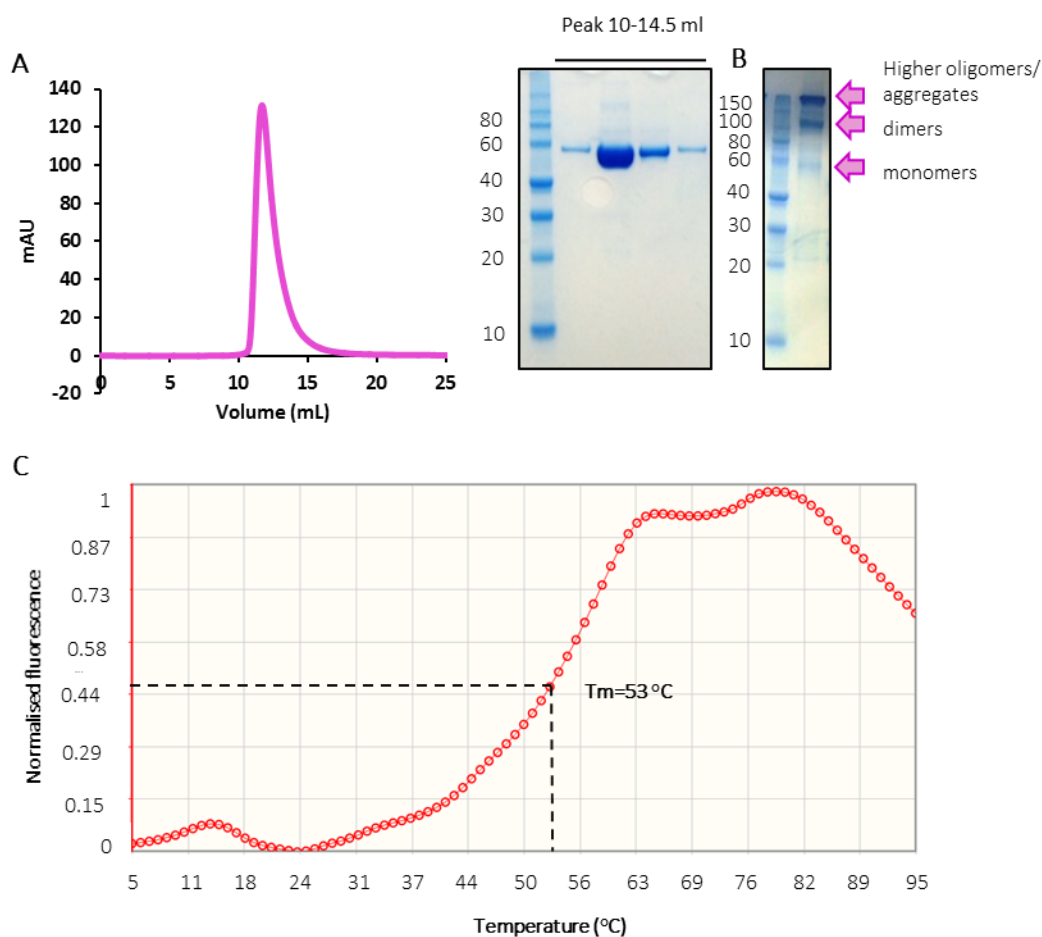


Figure 32: *Pcs60* purification and characterisation. (A) SEC elution profile of *Pcs60* from a Superdex 200 increase 10/300GL column. According to the elution volume, *Pcs60* elutes at a volume corresponding to a dimer. (B) Crosslinking with BS³ produced *Pcs60* dimers and oligomers, while the monomer fraction is nearly depleted. (C) Thermofluor analysis of *Pcs60* in SEC2 buffer indicates moderate stability with a melting temperature T_m of 53 °C.

3.2.1.2. Crystallisation and structure determination of *Pcs60* via X-ray crystallography

Full-length *Pcs60* was successfully crystallised after optimisation of initial crystal hits obtained from the PEGS Suite I screen in two conditions (Table 17). Crystals from both conditions had similar morphologies that belonged in the P2₁ space group (Figure 33), and produced diffraction data spanning to about 3 Å. As a high number of molecules could be traced in the asymmetric unit (ASU) of these crystals, phase determination was challenging.

Table 17: *Pcs60* initial hits and optimised crystallisation conditions

Screen	Protein Concentration (mg/ml)	Salt	Buffer	Precipitant
PEGs Suite I_E04	5	0.2 M LiCl	-	20% w/v PEG 3350
E04 optimised conditions	6	0.2 M LiCl	0.1M HEPES, pH 8.0	10-16% w/v PEG 3350
PEGs Suite I_F03	5	0.2 M LiNO ₃	-	20% w/v PEG 3350
F03 optimised conditions	6	0.2 M LiNO ₃	0.1M MES, pH 6.0	15-18% w/v PEG 3350

However, in the case of crystals grown for more than 3 months in the optimised conditions containing MES buffer, a degradation product of the full-length Pcs60 gave rise to a distinct crystalline polymorph of P2₁2₁2₁ symmetry. These crystals diffracted to a resolution of 2.4 Å and contained 6 molecules/ASU, making it fully possible to resolve the structure of Pcs60. The structure was solved by molecular replacement, using the phases of PDB entry 3TSY, which has a 27% sequence identity with Pcs60 (Wang *et al*, 2011). During these three months, degradation of the full length Pcs60 occurred, and the degradation product contained an N-terminal 50 kDa fragment, which is bulkier and displays reduced flexibility compared to the full-length protein, allowing thus the formation of better quality crystals. This N-terminal 50 kDa fragment, corresponds to an adenylation domain, conserved amongst enzymes of the AMP-binding family.

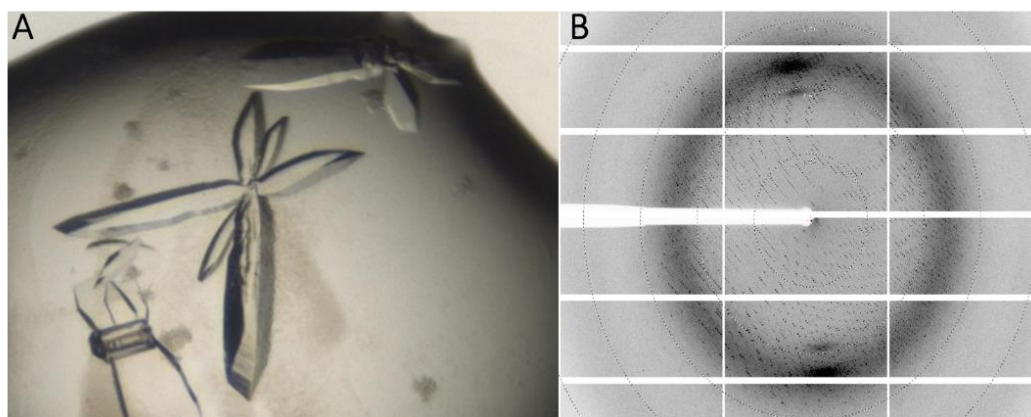


Figure 33: *Pcs60* crystals. (A) Crystals appearing at the optimised condition F03. (B) Diffraction pattern of *Pcs60*, collected at the MASSIF-1/ ID30A-1 beamline.

The newly-solved partial 50 kDa structure derived from the $P2_12_12_1$ crystals was used as a template for processing of the original $P2_1$ dataset, which contained 12 molecules/ASU and gave diffraction data of 2.87 Å resolution. Out of these 12 molecules, we could only trace one full-length Pcs60 molecule in the electron density, making it possible to reveal the whole structure, although all other molecules contained the 50 kDa fragment. Information on the crystallographic parameters for both crystal structures is summarised in **Table 18**.

Table 18: Pcs60 data collection and refinement statistics (statistics for the highest resolution shell are shown in parentheses).

Parameters	$P2_12_12_1$ crystals	$P2_1$ crystals
Data Collection		
Beamline	MASSIF-1/ID30A-1 (ESRF)	MASSIF-1/ID30A-1 (ESRF)
Cryoprotectant	25% Ethylene glycol	25% Ethylene glycol
Temperature	100K	100K
Wavelength (Å)	0.966	0.966
Molecules/ASU	6	12
Space group	$P2_12_12_1$	$P2_1$
Cell dimensions		
a,b,c (Å)	141.2, 163.4, 206.1	109.03, 93.72, 356.49
α,β,γ (°)	90, 90, 90	90, 93.81, 90
Solvent content (%)	62	51
Resolution range (Å)	49.14-2.40	49.20-2.87
R_{merge}	11.0 (128)	12.2 (94)
R_{meas}	0.071	
$I/\sigma I$	8.3 (1.1)	8.6 (1.3)
Completeness (%)	99.4 (99.8)	99.2 (99.2)
$CC_{1/2}$	0.99 (0.5)	0.99 (0.5)
Multiplicity	2.0 (2.0)	1.9 (1.9)
Wilson B-factor	48.69	61.21
Refinement		
Resolution (Å)	2.4	2.87
Unique Reflections	184612 (18355)	163072 (16216)
R_{work}/R_{free}	0.19/ 0.21	0.25/ 0.28
No. of atoms		
Residues	2564	5205
Water	624	--
B-Factors (Å ²)		
Protein	53.60	58.00
Solvent	49.40	--
r.m.s. deviations		
Bond lengths (Å)	0.003	0.006
Bond angles (°)	0.78	1.21

For the first time, the structure of the full-length Pcs60 monomer was revealed. Pcs60 forms a two-domain, helix-rich structure, with a larger N-terminal subdomain (T10-R432) linked to a smaller C-terminal subdomain (E448-K523) *via* a highly flexible linker (I433-I447, hinge loop).

The N-terminal adenylation domain adopts a complicated α/β -fold with alternating α -helices and β -sheets interconnected by elongated loop regions. An interesting feature of the N-terminal domain is that it contains a variation of the Walker-B (P-loop) motif which is known to be important for AMP binding (residues 355-370) formed between β -strands β 16- β 17, which are proximal to the interface with the C-terminal domain (**Figure 35D**). The smaller C-terminus, contains a short α -helix (α 13) followed by a β -hairpin (β 21- β 22) and another short α -helix (α 14).

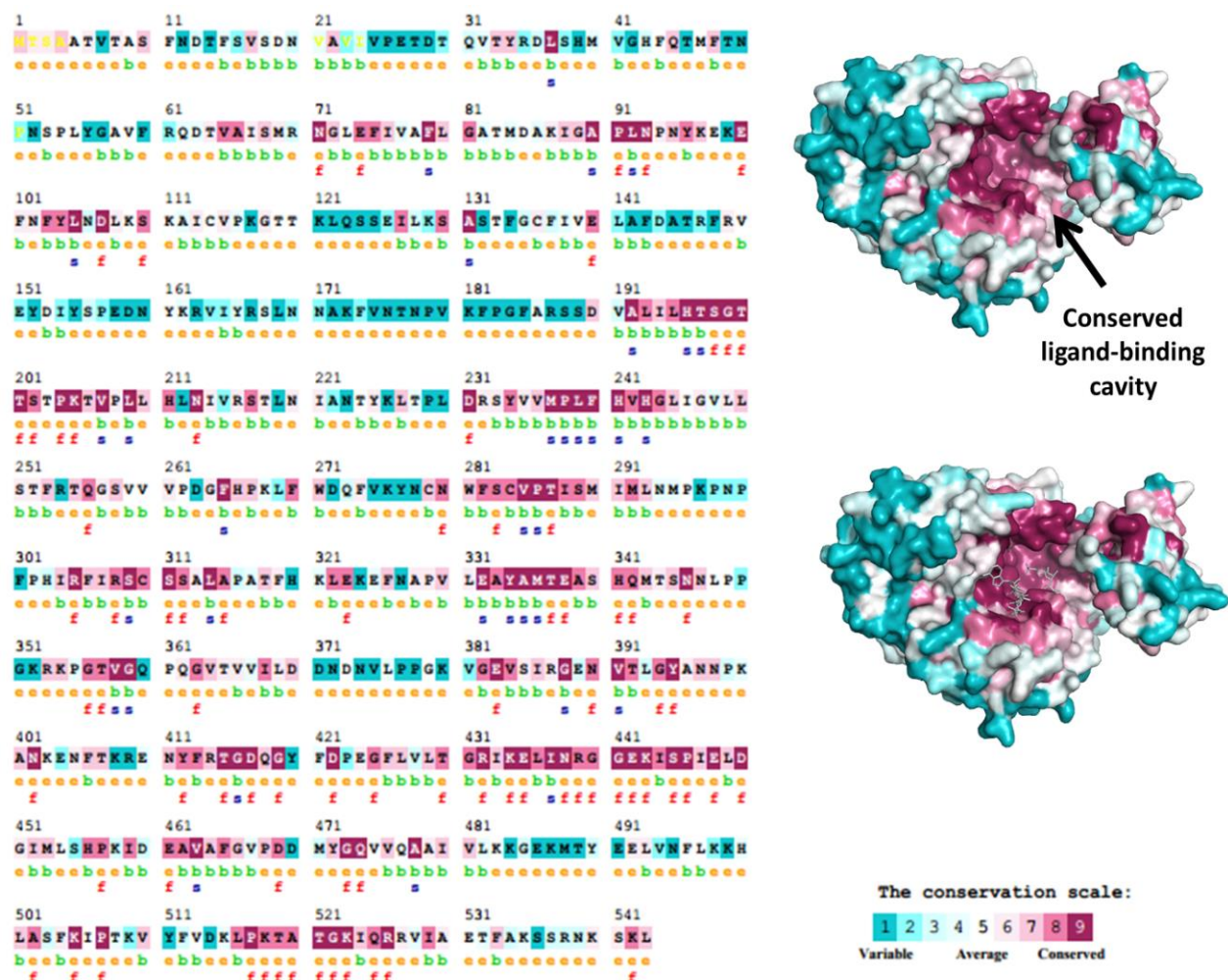


Figure 34: Amino acid sequence of Pcs60 coloured according to conservation scores. Right panel: (Top) Structure of Pcs60 monomer indicating high conservation in the ligand binding cavity, (Bottom) AMP and CoA co-crystallised with structure 5BSR mapped onto the Pcs60 structure. The letters b, e, f and s, correspond to buried, exposed, functional and structurally important residues respectively.

It has been reported that in similar enzymes of the AMP-binding superfamily, which catalyse biosynthetic adenylation reactions, the C-terminal region is dynamic and can rotate up to 140°, upon substrate binding (Li & Nair, 2015). Intriguingly, the AMP molecule can be mapped (using PDB structure 5BSR) into the interface between the N- and C-terminal domain and is buried in the highly conserved cavity formed by the two subdomains (**Figure 34**). Similarly, CoA from the same structure can easily fit into a groove on the N-terminal domain located close to the interface between the two Pcs60 subdomains.

Based on analysis *via* the PISA server (Krissinel & Henrick, 2007), Pcs60 can form homo-tetramers in the crystal, which can otherwise be described as dimers of dimers (**Figure 35A, B**). In accordance with its SEC elution profile, Pcs60 can homo-dimerise, with each dimer containing a surface area of 1027.6 Å² (**Table 19**). A number of hydrogen bonds contributes to this dimer formation (**Table 19, Figure 35C**). The dimer homo-dimerisation, on the other hand, forms a smaller interface area of about 409.8 Å², which is likely formed by crystal contacts (**Table 19, Figure 35C**). The interacting residues are mainly located in loop regions at the N-terminus of Pcs60 and the two monomers are placed in antiparallel orientation, allowing their respective active AMP-binding sites to be exposed to potential ligands (**Figure 35B**). Furthermore, the smaller C-termini that contain the PTS1 motif SKL for Pex5p recognition, although not visible in this structure, would also be placed in favourable positions for Pex5 binding.

Table 19: Residues involved in hydrogen bond formation in the Pcs60 homo-tetramer

Homo-dimerisation involved residues (interface area: 1027.6 Å²)		
Chain A	Distance (Å)	Chain E
R189[NH1]	2.92	G390[O]
S190[OG]	2.67	E391[OE1]
S190[N]	2.87	E391[OE2]
E391[OE1]	2.76	S190[OG]
Homo-tetramerisation involved residues (interface area: 409.8 Å²)		
Chain A	Distance (Å)	Chain B
K354[NZ]	2.89	D38[OD2]

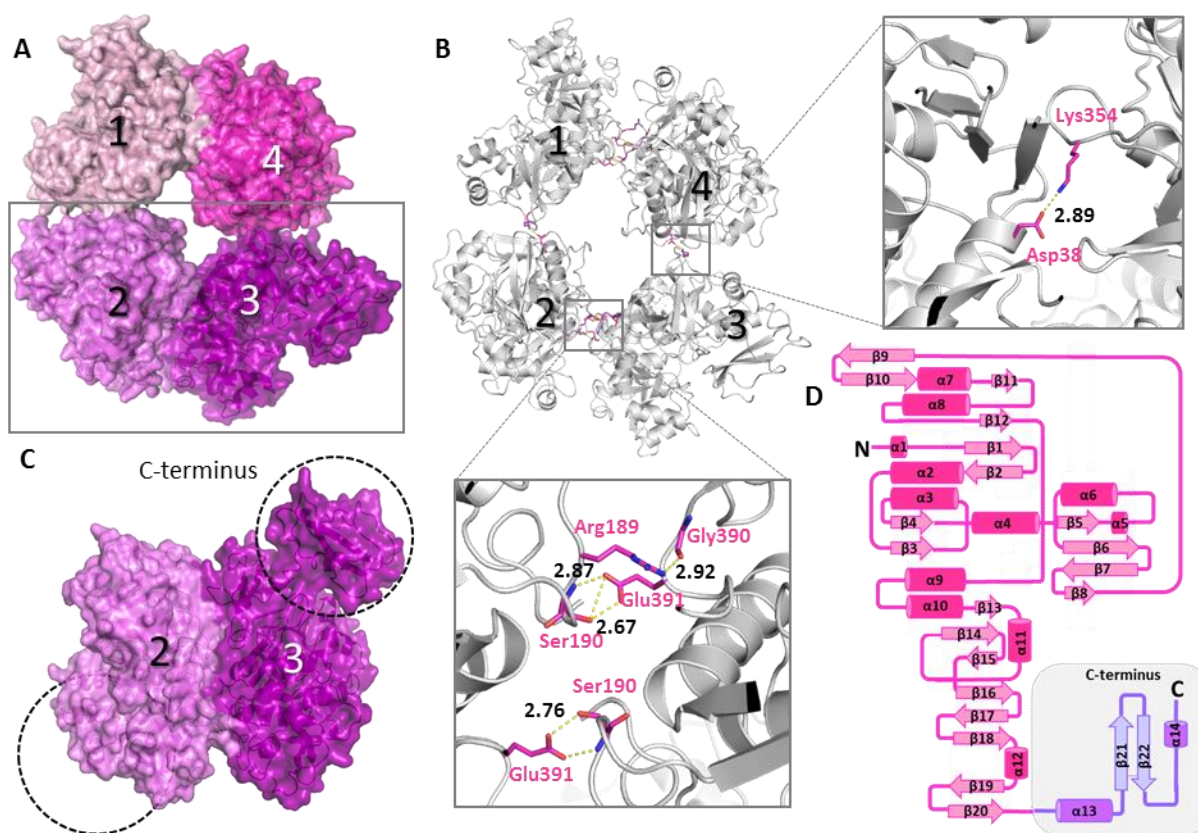


Figure 35: Structure of Pcs60. (A) Surface representation of Pcs60 homo-tetramer. The different monomers are indicated in separate colours. (B) Hydrogen bonds contributing to the homo-tetramerisation of Pcs60. Right insert: Hydrogen bonds between molecules 1-2 and 3-4. Bottom inset: Hydrogen bonds between molecules 1-4 and 2-3. (C) The Pcs60 dimer is formed by antiparallel binding between two monomers. Molecule 3 contains the C-terminal domain which is missing from all the other copies in the asymmetric unit. The position of the missing C-terminus of molecule 2 is indicated with a dashed circle. (D) Schematic representation of the secondary structure of full-length

3.2.2. Structural characterisation of the Pex5p-Pcs60 complex

3.2.2.1. Expression and purification of Pex5p/Pcs60 complexes

To improve the probability for crystallisation, two Pex5 constructs were designed containing the TPR region (Pex5p₍₃₁₂₋₆₁₂₎) and the TPR region (Pex5p₍₁₉₈₋₆₁₂₎), plus the first Pex14p binding site (extended TPR), in addition to the full length Pex5p (Pex5p_{FL}) (**Figure 36A**).

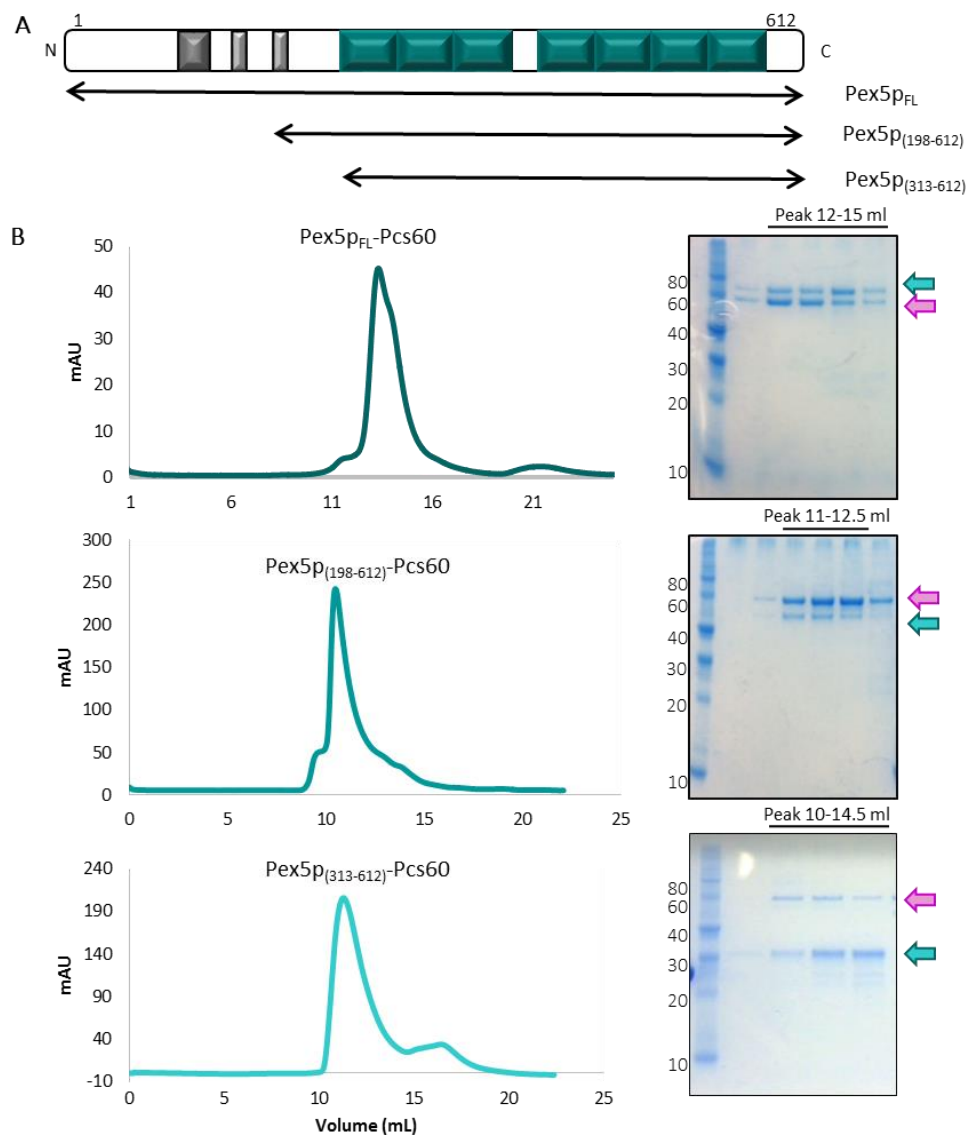


Figure 36: Purification of Pex5p-Pcs60 complexes. (A) Construct design for Pex5p. The dark grey rectangle indicates its Pex13p binding site, the light grey represents the WxxxF/Y motifs required for Pex14p binding and the teal rectangles represent the number of TPR repeats. (B) SEC elution profiles for Pex5p_{FL}-Pcs60 (using a Superose 6 increase 10/300 GL column), Pex5p₍₁₉₈₋₆₁₂₎-Pcs60 and Pex5p₍₃₁₃₋₆₁₂₎-Pcs60 (using a Superdex 200 increase 10/300GL) with their corresponding SDS PAGE analysis. Pex5p constructs are indicated with teal arrows, while Pcs60 with pink.

The different Pex5p constructs were expressed separately and mixed with Pcs60 in equimolar ratio for complex formation (**Figure 36B**). The Pex5p_{FL}-Pcs60 complex is known to form higher oligomers of 440kDa according to static light scattering measurements (Hagen *et al*, 2015), a behaviour that can also be observed with the smaller Pex5p constructs, and can possibly be attributed to the homo-tetramerisation of Pcs60.

All Pex5p constructs are soluble and easy to purify, yielding protein quantities of 30mg per litre of expression. Pex5p_{FL} can form potential dimers in solution as can be observed from its SEC elution profile, whereas the shorter constructs act as monomers (**Figure 36B**). Furthermore, the purification of Pex5p constructs with Pcs60 indicates unstable oligomers, which seem to be more prominent only at higher concentrations, but dissociate upon dilution.

3.2.2.2. *Low resolution SAXS analysis of Pex5p/Pcs60*

SAXS analysis of the Pex5p₍₁₉₈₋₆₁₂₎-Pcs60 and Pex5p₍₃₁₂₋₆₁₂₎-Pcs60 complexes was performed at the Petra III beamline P12, using a series of increasing concentrations for both samples (0.5- 3 mg/ml). The experimental scattering curve is shown in **Figure 37A**. Primary SAXS analysis was performed in collaboration with Dr. C. Jeffries (EMBL-Hamburg) and reveals that both Pex5p-Pcs60 complexes likely form oligomeric concentration-dependent complexes in solution. The Kratky plot indicates a multi-globular particle formation with some degree of flexibility (**Figure 37B**).

The distance distribution function profile (**Figure 37C**) contains a defined peak, centred between 5-10 nm, and the overall shape of the complex has features characteristic of a modular/flattened particle. The D_{max} can vary between 28-34 nm, suggesting either the influence of higher molecular weight species, or that the protein components located at the periphery of the complex are displaying certain degree of flexibility. What can also be deduced from this SAXS analysis is that the complexes do not form 1:1 hetero-dimers, but instead form tetramers consisting of four Pex5p-Pcs60 hetero-dimers (or eight protomers). This hypothesis is further validated by the estimated molecular weight values of the complexes that are comparable to their theoretical molecular weight (440 kDa for Pex5p₍₁₉₈₋₆₁₂₎-Pcs60 and 376 for Pex5p₍₃₁₂₋₆₁₂₎-Pcs60)(

Table 20).

Table 20: Extrapolated zero concentration of SAXS structural parameters

	Pex5p ₍₁₉₈₋₆₁₂₎ -Pcs60	Pex5p ₍₃₁₂₋₆₁₂₎ -Pcs60
Data collection parameters		
Instrument	EMBL P12 (PETRA-III)	
Beam geometry (mm ²)	0.2 x 0.12	
Wavelength (Å)	1.24	
s range (Å ⁻¹) ^a	0.003-0.45	
Exposure time (s)	0.045-20	
Sample concentration (mg/mL)	0.5-5.0	
Temperature (K)	283	
Structural parameters		
Io (0) (cm ⁻¹) [from P(r)]	99050 ±146.1	28760 ±28.6
Io (0) (cm ⁻¹) [from Guinier]	99575.6 ±235.05	28818 ±57.14
R _g (nm) [from P(r)]	7.28 ±0.016	5.45 ±0.06
R _g (nm) [from Guinier]	7.21 ±0.2	5.44 ±0.04
D _{max} (nm)	28-34	19
Porod volume estimate (nm ³)	1192	680
Molecular mass MW (kDa) [from Porod volume (V _p /1.7)]	460	340
Calculated monomeric MW (kDa) ^b	440	376
Software employed		
Primary data reduction	Automated pipeline (Franke et al., 2012)	
Data processing	PRIMUS	
Ab initio analysis	DAMMIF, GASBOR	
Validation and averaging	DAMMIN, DAMAVER	

^aMomentum transfer $|s| = 4\pi\sin(\theta)/\lambda$.

^bcalculated from the sequence using the <http://www.basic.northwestern.edu/biotools/proteincalc.html> server.

Taking into account the possibility of hetero-tetrameric complexes, the ATSAS suite programme GASBOR was used to calculate *ab initio* models in P4 symmetry, which had an acceptable fit to the experimental data ($\chi^2 = 2.5$). The resulting models resemble flattened disks with flexible extensions for both complexes (**Figure 37E**). On the assumption that the central core of the complex corresponds to Pcs60, Pex5p can be placed on the periphery of the assembly (**Figure 37D**).

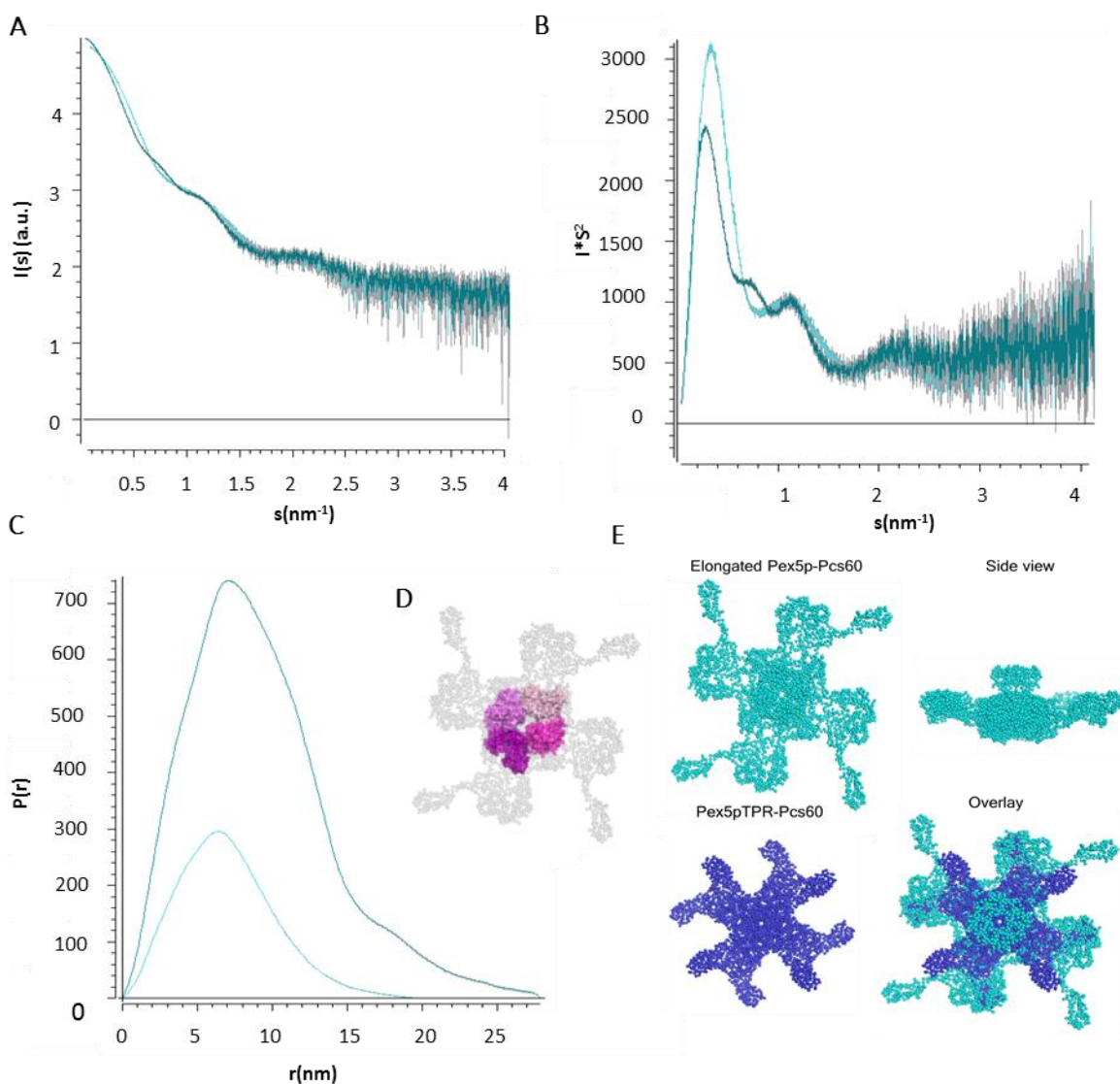


Figure 37: SAXS analysis of Pex5p-Pcs60 complexes. (A) Scaled plotted intensity of the SAXS measurements for Pex5p₍₁₉₈₋₆₁₂₎-Pcs60 (dark teal) and Pex5p₍₃₁₂₋₆₁₂₎-Pcs60 (light teal). (B) Kratky plots, indicating compactness of the system as well as multiple domains in both cases. (C) Distance distribution function overlay, (D) Overlay of Pex5p₍₁₉₈₋₆₁₂₎-Pcs60 GASBOR model with the Pcs60 tetramer crystal structure. (E) GASBOR models for both complexes.

3.2.2.3. Crystallisation of Pex5p/Pcs60 complexes

The complexes Pex5p₍₁₉₈₋₆₁₂₎-Pcs60 and Pex5p₍₃₁₂₋₆₁₂₎-Pcs60 were crystallised after optimisation of initial crystal hits obtained from the ProComplex and JCSG_Core_I crystallisation screens, respectively (**Table 21**). For the Pex5p₍₁₉₈₋₆₁₂₎-Pcs60 complex, two macroscopically different crystal types were obtained; the original crystal hits from ProComplex screen A12 condition, which resembled thin overlapping plates (crystal type I), and the new crystals that changed dramatically in morphology, which resembled hexagonal plate crystals (crystal type II) upon lowering of the pH to 5.0 (**Figure 38A,C**).

Table 21: Initial hits and optimised crystallisation conditions for Pex5p-Pcs60 complexes

Screen	Construct	Protein Concentration (mg/mL)	Salt	Buffer	Precipitant
ProComplex_A12	Pex5p ₍₁₉₈₋₆₁₂₎ -Pcs60 (Crystal type I)	2.8	0.2M Sodium acetate	0.1M Sodium citrate pH5.5	5% w/v PEG 4000
A12 optimised conditions	Pex5p ₍₁₉₈₋₆₁₂₎ -Pcs60 (Crystal type II)	3	0.2M Sodium acetate	0.1M Sodium citrate pH5.0	5-8.4% w/v PEG 4000
JCSG_CORE_I_B01	Pex5p ₍₃₁₂₋₆₁₂₎ -Pcs60	6.6	0.2M NaCl	0.1M HEPES pH7.5	20% w/v PEG 3000

Crystal type I crystals were tested on the MASSIF-1/ID30A-1 beamline at ESRF, but due to their plate-like morphology, they diffracted anisotropically to 4.5-5 Å resolution (**Figure 38B**). The crystals belonged to the tetragonal P4 space group ($a = b = 80.106 \text{ \AA}$, $c = 186.262 \text{ \AA}$, $\alpha = \beta = \gamma = 90^\circ$), but further analysis was not possible due to lack of a complete dataset. Matthews coefficient probability calculation (Weichenberger & Rupp, 2014; Matthews, 1968) predicts one hetero-dimer per unit cell based on the molecular weight of the complex and the volume of the unit cell ($V = 1,194,661.1 \text{ \AA}^3$) with 54.7% solvent content. In light of observations that the complex consists of four hetero-dimers, the internal four-fold symmetry of the crystal would comply with the four-fold symmetry of the hetero-tetramer. Crystal content analysis was performed *via* SDS-PAGE analysis and indicated that both Pex5p₍₁₉₈₋₆₁₂₎ and Pcs60 were present in the crystal (**Figure 38D**).

Crystal type II crystals only gave weak diffraction and require further optimization. The crystals that were acquired from the Pex5p₍₃₁₂₋₆₁₂₎-Pcs60 complex contained only Pcs60, according to SDS-PAGE analysis of the crystal content (not shown). The latter finding indicates that the

stability of the complex is reduced when the shorter TPR region is used. This could point to additional interaction sites located upstream of the Pex5p TPR region.

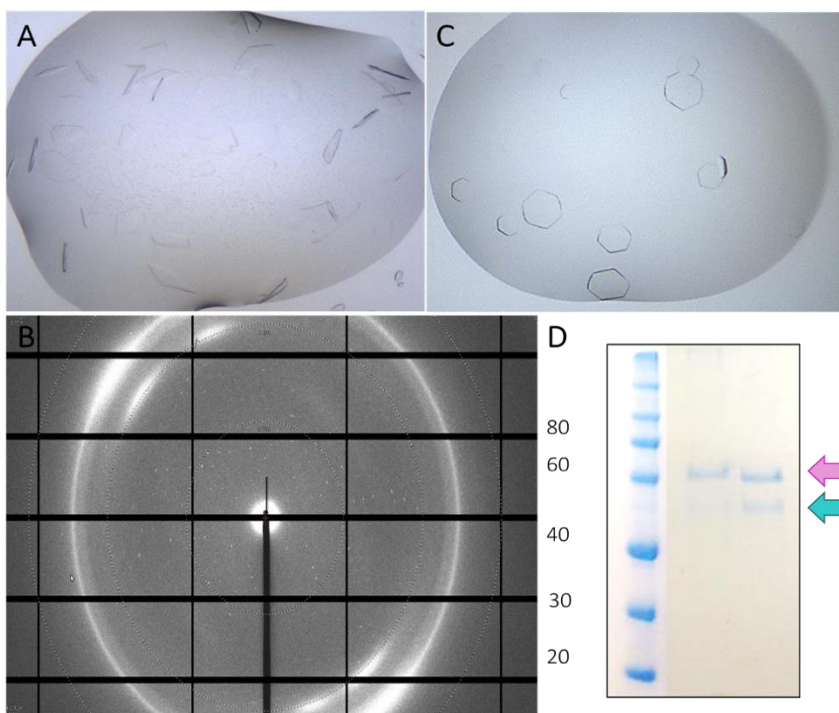


Figure 38: *Pex5p₍₁₉₈₋₆₁₂₎-Pcs60* crystals. (A) Crystal type I (ProComplex_A12 Conditions), (B) Limited diffraction of Crystal type I, (C) Crystal type II (A12 optimised conditions), (D) Crystal content of crystal type I. Both *Pex5p₍₁₉₈₋₆₁₂₎* (teal arrow) and *Pcs60* (pink arrow) are present.

3.2.2.4. Structural analysis of *Pex5p_{FL}/Pcs60* via negative stain EM

Protein samples of the *Pex5p₍₁₉₈₋₆₁₂₎-Pcs60* complex were screened by means of negative stain EM (in collaboration with Pascal Lill, Dr. Bjoern Klink and Dr. Christos Gatsogiannis, Max Planck Institute, Dortmund). Despite hetero-tetrameric complexes having been detected by SAXS, the low concentrations used for negative stain EM experiments prevented their formation under native conditions. For this reason, prior to screening, the complexes were crosslinked, allowing hetero-tetramerisation to occur (**Figure 39A, B**). Imaging of the complex without use of crosslinking resulted in its dissociation and formation of smaller hetero-dimeric complexes (not shown). This further validates the observation that the hetero-tetramer formation is unstable and concentration dependent. Preliminary EM results indicate a hetero-tetrameric oligomer, similar to the predicted SAXS model, in which a tetrameric core correlates well with the *Pcs60* tetramer surrounded by four *Pex5p* molecules (**Figure 39C, D**).

However, unlike the SAXS model, not all four Pex5p positions seem to be occupied in all occasions, likely due to limitations of the method and the staining techniques that were used. Making use of the increased size offered by the Pex5p_{FL}-Pcs60 complexes and a potentially more stable interaction, these complexes, although not suitable for crystallography due to the flexibility present at the Pex5p N-terminus, can offer additional structural insight.

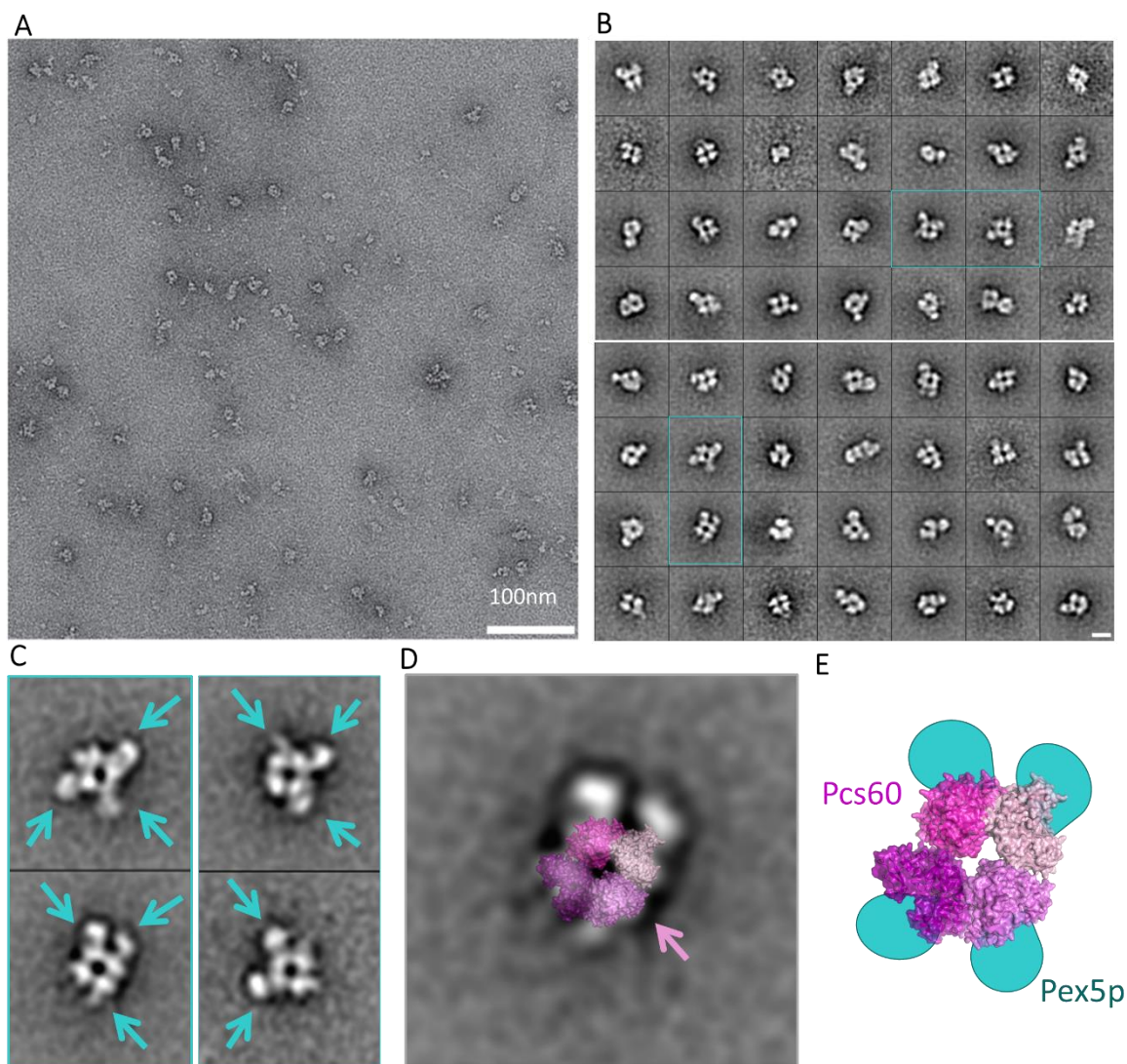


Figure 39: Negative stain EM analysis of the Pex5p₍₁₉₈₋₆₁₂₎-Pcs60 complex. (A) Electron micrograph of the crosslinked complex. (B) Class averages of the complex, which clearly contain a tetrameric core, flanked by 1-3 globular extensions in the periphery, likely to be Pex5p₍₁₉₈₋₆₁₂₎. The general shape of the particle correlates well with the SAXS-derived model. (C) Magnified class averages, in which Pex5p₍₁₉₈₋₆₁₂₎ is indicated by arrows (teal). (D) Magnified class average, superimposed with the crystal structure of the Pcs60 tetramer (pink). (E) Graphical representation of EM particles, with the Pcs60 tetramer in the core of the complex. Four Pex5p₍₁₉₈₋₆₁₂₎ domains (teal) are located in the periphery of the complex.

3.2.3. Structural characterisation of *AtMIF1*

3.2.3.1. *Expression, purification and biophysical characterisation of AtMIF1*

Full length *AtMIF1* is a very soluble 12 kDa plant peroxisomal protein that was structurally characterised as a potential *AtPEX5* target protein. This *AtMIF1* construct contained a non-cleavable N-terminal His_{6x} tag which was used for purification and could be purified at high yield (60 mg per litre of expression) (**Figure 40**).

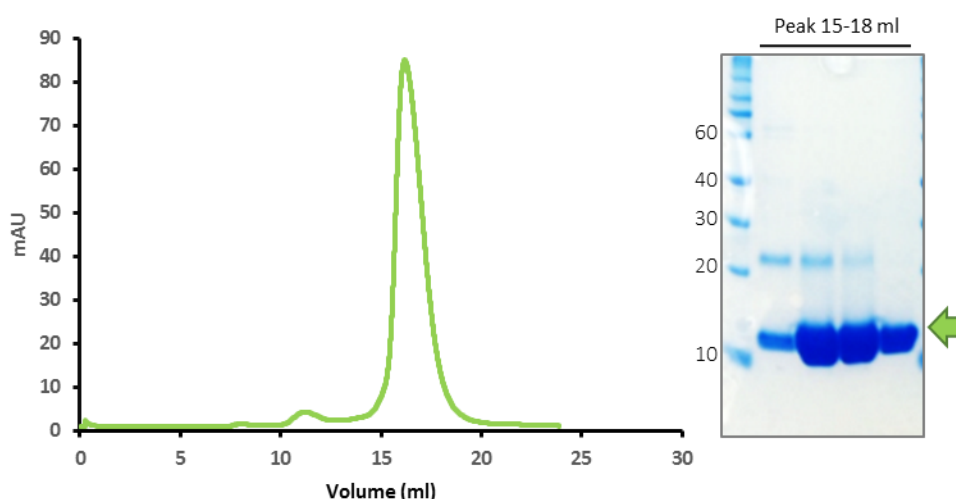


Figure 40: *AtMIF1* purification (green) and corresponding SDS-PAGE analysis (green arrow).

3.2.3.2. *Structural characterisation of AtMIF1 via X-ray crystallography*

AtMIF1 was crystallised using the commercial crystallisation screens QIAGEN_Classics_I and QIAGEN_PEGS_I (conditions summarised in **Table 22**).

Table 22: Crystallisation conditions for *AtMIF1*

Screen	Protein Concentration (mg/mL)	Salt	Buffer	Precipitant
Pegs_I_E03	15	0.1M ammonium fluoride	-	20% w/v PEG 3350
Pegs_I_H08	15	0.2 M di-ammonium sulfate	-	20% w/v PEG 3350
Classics_I_E11	15	1.4 M lithium sulfate	0.1 M HEPES pH 7.5	-
Classics_I_F07	15	0.2 M calcium acetate	0.1 M sodium cacodylate pH 6.5	18% w/v PEG 8000

Macroscopically similar crystals appeared in several conditions in both screens within 1 hour after setting up the experiment (**Figure 41A-D**). The crystals were tested at the MASSIF-1/ID30A-1 beamline at the ESRF and gave good quality diffraction data that enabled determination of the crystal structure of *At*MIF1 at 1.9 Å resolution (**Figure 41E**). The three-dimensional structure was resolved by molecular replacement using PDB model 3T5S (MIF homologue from *Giardia lamblia*, 24% sequence identity). Information on the data collection and refinement parameters can be found in **Table 23**.

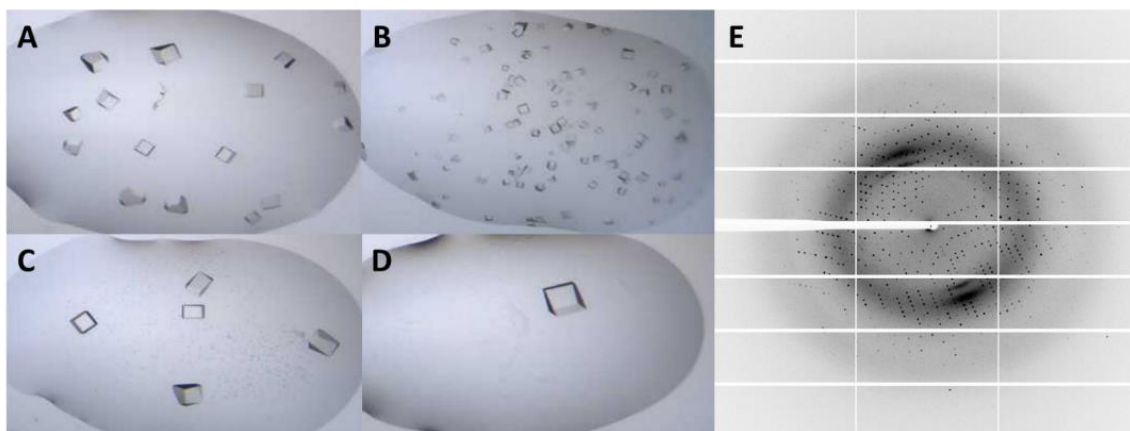


Figure 41: *At*MIF1 crystals corresponding to the four different crystallisation conditions, E03, H08, E11 and F07, summarised in **Table 22**.

The *At*MIF1 crystals contain one molecule in the asymmetric unit but forms a functional trimer, with an interface area of 720.8 Å² between each monomer (**Figure 42A, C**). A number of hydrogen bonds contribute to trimer formation indicated in **Figure 42C**. The *At*MIF1 monomer adopts an α/β -fold comprising two anti-parallel α -helices facing a four-stranded β -sheet. An additional short β -strand, aligned with the β 4-strand from the adjacent subunit, is involved in the stabilisation of the homotrimer (**Figure 42D, E**).

Table 23: *AtMIF1* data collection and refinement statistics (statistics for the highest resolution shell are shown in parentheses).

Data collection parameters	
Beamline	MASSIF-1/ ID30A-1
Cryoprotectant	25% Ethylene Glycol
Temperature	100K
Wavelength (Å)	0.966
Molecules/ASU	1
Space group	I23
Cell dimensions	
a,b,c (Å)	89.943, 89.943, 89.943
α,β,γ (°)	90, 90, 90
Solvent content (%)	50.5
Resolution range (Å)	63.59-1.89
R_{merge}	4.5 (76)
R_{meas}	0.03439
$I/\sigma I$	12.7 (1.5)
Completeness (%)	98.76 (99.29)
$CC_{1/2}$	0.99 (0.7)
Multiplicity	1.8 (1.8)
Wilson B-factor	36.79
Refinement parameters	
Resolution (Å)	1.90
Unique Reflections	9743 (974)
R_{work}/R_{free}	0.18/ 0.21
No. of atoms	
Residues	103
Water	58
B-Factors (Å ²)	
Protein	48.40
Solvent	75
r.m.s. deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.73

Due to the presence in the electron density of an extra proline (Pro-1) belonging to the uncleavable purification tag, in addition to the initial methionine (Met-2), which is normally post-translationally removed, the ligand binding cavity becomes obscured and unavailable for binding. This renders the catalytic Pro-3 unavailable and the protein inactive. In order to detect bound ligands, new constructs of *AtMIF1* containing cleavable or C-terminal purification tags have been produced and their expression is being optimised. Due to flexibility in the C-terminus of *AtMIF1*, there is no density to support the PTS1 –SKL sequence. However, from the formation of the trimer, it is possible to derive that the C-termini are exposed, allowing *AtPEX5* to bind.

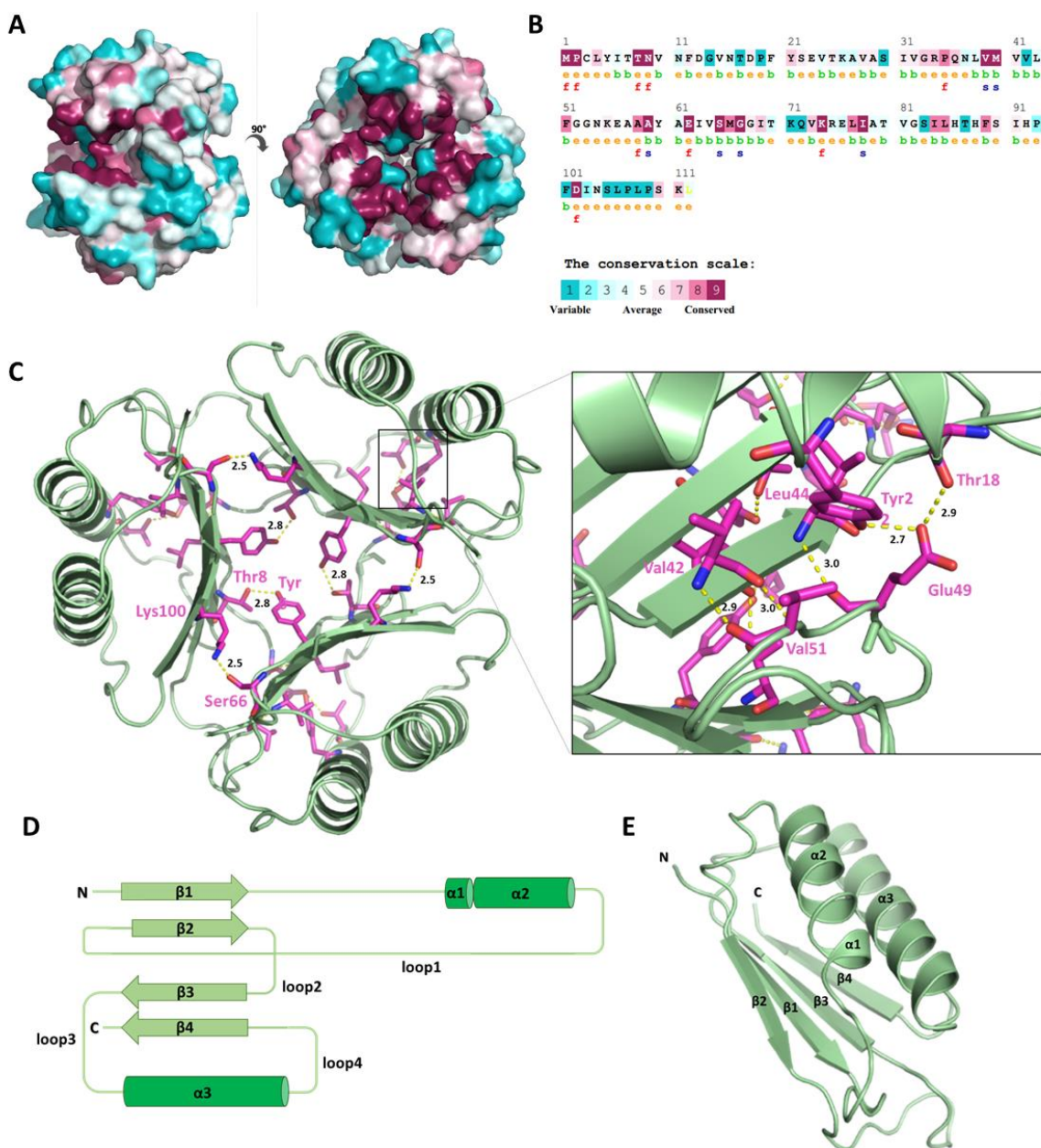


Figure 42: AtMIF1 structure. (A) Surface area of AtMIF1 trimer indicating conservation (produced using the ConSurf server), (B) Residue sequence of AtMIF1 coloured according to conservation score. The letters b, e, f and s, correspond to buried, exposed, functional and structurally important residues respectively. (C) Hydrogen bonds forming between AtMIF1 monomers, (D) Schematic representation of secondary structure features of AtMIF1, (E) Structure of AtMIF1 monomer.

4. Discussion

4.1. Structural insight into peroxisomal membrane complexes

In this thesis, the structure of PEX19 complexes was studied. PEX19 is a soluble receptor involved in peroxisomal membrane protein (PMP) import and consequently, peroxisome biogenesis. PEX19 can interact with a range of PMPs by a number of amphipathic helices, predicted at its unstructured N-terminus. The best characterised of these PMP interactions concerns its binding to the membrane anchor protein PEX3, which is crucial for PMP insertion into the peroxisomal membrane. Currently, two crystal structures of the cytosolic domain of PEX3 with an N-terminal PEX19 peptide are available (PDB entries: 3MK4, 3AJB), but structural insight on the interaction of PEX3 with the full-length PEX19 is scarce. For this reason, the first part of the thesis aimed to provide structural insight into the above interaction. PEX19 is also known to interact with the essential peroxisomal protein PEX14, which is a PMP involved in the formation of the PTS1 import pore. As the peroxisomal localisation of PEX14 has been shown to rely mainly upon PEX13, rather than PEX19, the basis of its interaction with PEX19 is still not well understood. In this thesis, the structural properties of the PEX19-PEX14 interaction were investigated, using a hybrid approach. Furthermore, the interaction of PEX19 with a subset of other PMPs was also examined.

4.1.1. Structural characterisation of the PEX3_{C235S(41-373)}-PEX19 complex

The PEX3_{C235S(41-373)}-PEX19 complex was expressed, purified and characterised using a variety of biophysical methods. From its SEC-RALS profile it can be deduced that it forms a soluble hetero-dimer, while individual PEX19 and PEX3_{C235S(41-373)} are monomers in solution. PEX19, despite its smaller molecular weight, displays the elution profile of a protein with larger hydrodynamic radius, due to the disorder of its N-terminus, which can bind PEX3 and is possibly involved in PMP import and release. Secondary structure analysis by CD indicated a highly helical content for PEX3_{C235S(41-373)}, a largely unstructured random coil content for PEX19, and a significantly decreased random coil content for the complex, which can be attributed to the presence of α -helical PEX3_{C235S(41-373)}. Despite binding to PEX3_{C235S(41-373)}, PEX19 retains its

unstructured nature in solution (and in absence of PMPs), implying that this unstructured region requires additional binding partners to adopt a specific structural conformation.

To further investigate the flexibility of PEX19 in solution and in complex with PEX3_{C235S(41-373)}, SAXS analysis was performed. SAXS provided information about the overall shape and compactness of the system and contributed to the computation of *ab initio* models that can explain the experimental data. Consistent with prior knowledge, the modelling was performed using the α -helical C-terminal domain of PEX19 containing the PMP binding helix and the PEX3_{C235S(41-373)} structure (PDB entry: 3MK4) that provides restraints on the binding of PEX19. On the other hand, the connecting region between these two subdomains can be characterised as disordered (**Figure 43A**). Using the programme CORAL, multiple models were generated, indicating that the C-terminal domain of PEX19 can move freely while in complex with PEX3_{C235S(41-373)}, despite binding restraints imposed on its N-terminus. We assume that the flexibility of the linking region can accommodate a number of PMPs, as it contains several exposed hydrophobic and charged residues that are predicted to be involved in PMP binding and release (Chen *et al.*, 2014). Furthermore, it was possible to reconstruct a model of PEX19 in solution, based on SAXS data. The model correlates well with the experimental data and indicates flexibility of the N-terminus that adopts an elongated conformation (**Figure 43B**). Taking all SAXS data into account we can conclude that the PEX3_{C235S(41-373)}- PEX19 complex can form a multi-domain flexible system that can be characterised by a level of compactness.

As an ultimate goal of this thesis, the crystal structure of PEX3_{C235S(41-373)}- PEX19- PMP complexes was pursued. Although crystallisation of the complex was attempted using several conditions, the inherent flexibility of PEX19 hindered any crystal formation. Based on the assumption that PEX19 can act as a chaperone for PMPs, PMP peptides that contain the PEX19-interacting mPTS region were used for co-crystallisation with PEX19 and PEX3_{C235S(41-373)}. However, due to the high hydrophobicity of the peptides, solubilisation in aqueous buffers was impossible and subsequent complex formation was prevented. An approach of co-expression of PEX19 with PMPs was also followed, but the resulting PMPs were not soluble and thus were not pursued further.

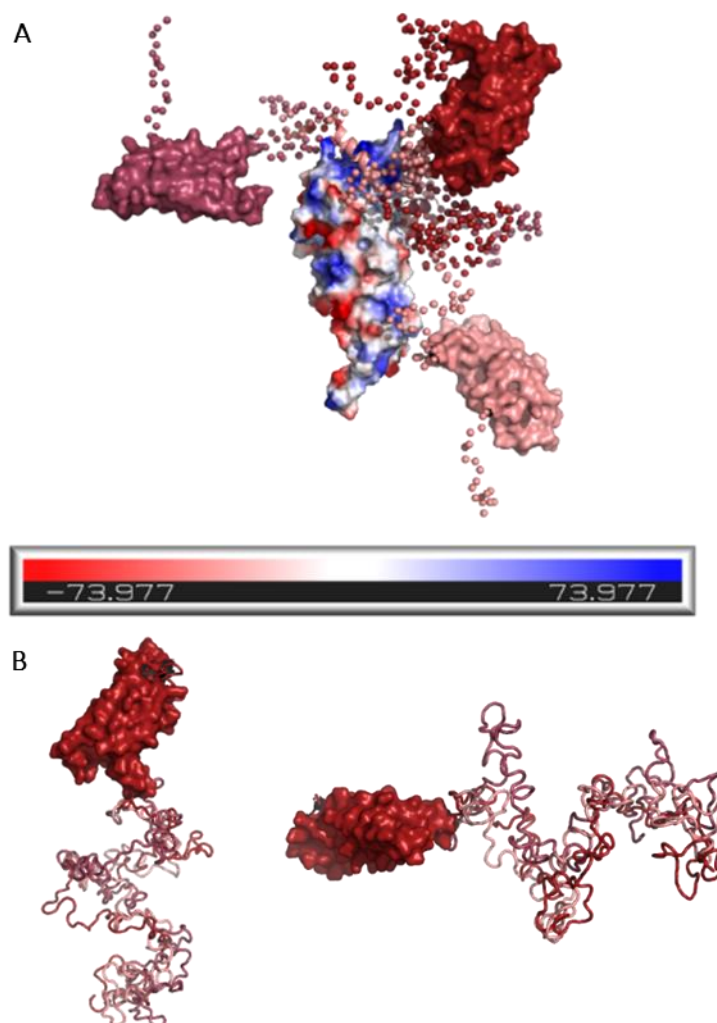


Figure 43: SAXS models using CORAL. (A) PEX19 (red, raspberry and light pink colour) can freely move around PEX3 (electrostatic potential coloured surface), (B) Ab initio modelling of PEX19₍₁₋₂₉₉₎ (left) side view and (right) upper view.

4.1.2. The mPTS charge can affect PEX19 binding

In collaboration with the Schrader group (University of Exeter), we explored the importance of mPTS charge for PEX19 binding and peroxisomal localisation. Using ACBD5, a tail-anchored PMP protein, quantitative binding affinity measurements were performed and verified by *in vivo* data from our collaborators, which suggested that peroxisomal targeting can be influenced by the charge of the region following the transmembrane domain of ACBD5. Indeed, substitution of the

positively charged arginine residues on the ACBD5 C-terminus with non-polar alanine residues, resulted in a decrease in binding affinity with PEX19, which could also be verified *in vivo*, by mislocalisation of the mutated proteins to other cellular compartments (Costello *et al.*, 2017). These findings propose a model, in which, the highly charged tail of ACBD5, or other tail-anchored proteins, is required for peroxisomal import *via* binding with the import receptor PEX19.

4.1.3. Novel overlapping binding site on PEX19 can accommodate both PEX3 and PEX14

Following successful isolation and characterisation using biophysical methods, we investigated the presence of additional binding sites between PEX3 and PEX19, as has been previously suggested (Fransen *et al.*, 2005; Matsuzono *et al.*, 2006; Schmidt *et al.*, 2010, 2012). For this purpose, PEX3_{C235S(41-373)} was chemically crosslinked with PEX19 and the resulting dimers were analysed by mass spectrometry.

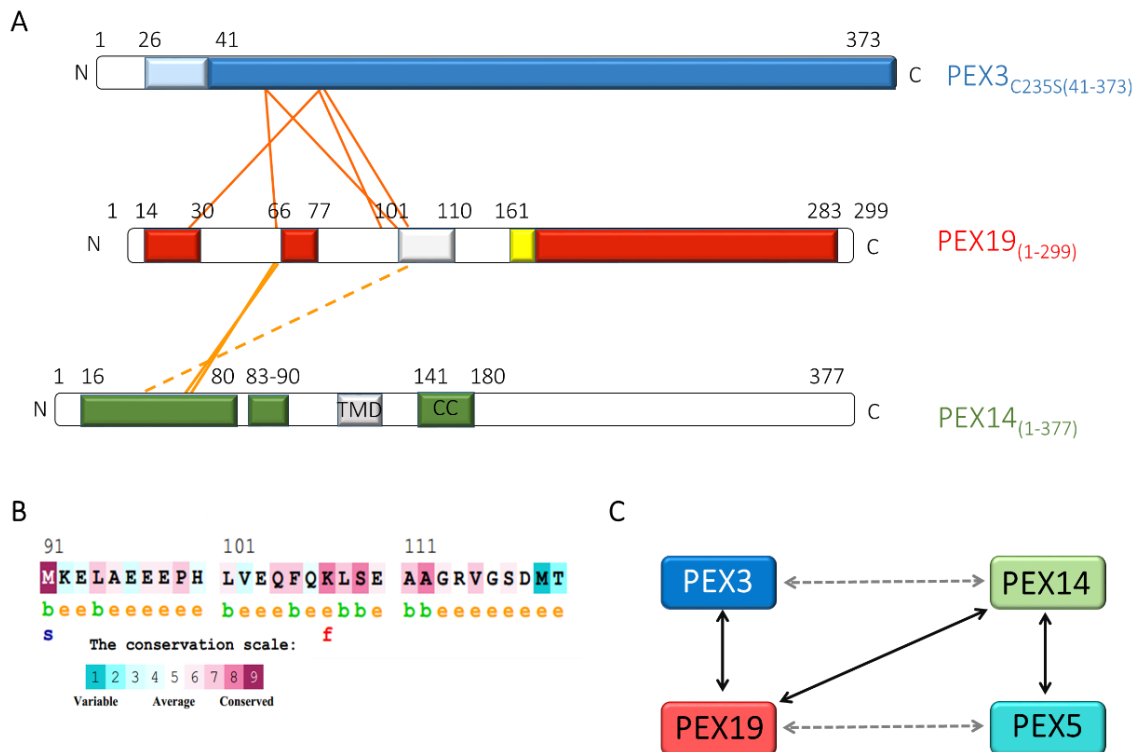


Figure 44: Auxiliary PEX19 binding site shared between PEX3 and PEX14. (A) Graphical illustration of intermolecular crosslinking results (orange lines). (B) Sequence conservation of the PEX19 novel binding site, containing the LVEQF segment. (C) Interaction map between PEX3, PEX19, PEX14, and PEX5. Proven interactions are indicated with solid black arrows, while possible competitive interactions with dashed grey arrows.

In addition to the already well-characterised N-terminal PEX19 binding site, a number of crosslinks revealed an auxiliary binding region located in the middle of the disordered domain (**Figure 44A**). Closer investigation of this region showed that it is conserved amongst species and has been previously annotated as one of the predicted amphipathic helices, alpha-c, corresponding to residues 101-112 (**Figure 44B**) (Chen *et al*, 2014). The binding between PEX19 and PEX3 relies on two amphipathic segments of PEX19, defined as its primary binding site (residues 14-30) and its secondary binding site (residues 101-110). Both of these segments were shown to be proximal with another amphipathic segment of PEX3 (residues 100-110). Although it is not known whether binding occurs simultaneously or sequentially, binding of both PEX19 segments on the apex of PEX3 would likely result in a ‘bent’ conformation of the PEX19 N-terminus (**Figure 45**). Interestingly, the connecting N-terminal chain (residues 31-100) contains an FxxxF motif shown to bind PEX14 (PDB entry: 2W85), as well as another predicted amphipathic helix, denoted as alpha-b previously (Chen *et al*, 2014), which has not been well characterised so far. Taking the above into account, this bent conformation of the PEX19 N-terminus could, in turn, allow the connecting region between the two binding sites to accommodate PEX14 (**Figure 45**), although attempts to isolate the ternary PEX19-PEX3-PEX14 complex *in vitro* have not been successful so far.

Furthermore, the novel PEX19 binding site contains the sequence LVEQF, which bears resemblance to the LVAXF motif that was discovered on PEX5 and allows PEX14 binding (Neuhaus *et al*, 2014). Based on this finding, we investigated the presence of additional sites on PEX19 for PEX14 binding. Using the same approach, our data showed that indeed, the majority of novel crosslinks points towards the same region of PEX19 (residues 101-112) (**Figure 44A**). Taking our native-MS data, which show a 1:1 binding of PEX19 with the N-terminus of PEX14 (residues 16-80), we can conclude that most likely one PEX19 molecule interacts with PEX14 on multiple sites, rather than multiple PEX19 molecules binding to that region of PEX14. However, whether there are additional sites further downstream of that region is still under investigation.

This shared binding site builds a complex interaction map between PEX3, PEX19, PEX14 and PEX5 (**Figure 44C**). It is known that PEX14 and PEX5 are required for protein import into peroxisomes, whereas PEX3 and PEX19 are involved in membrane protein import and peroxisome

biogenesis, while further interactions can provide insight into alternative functions for these proteins.

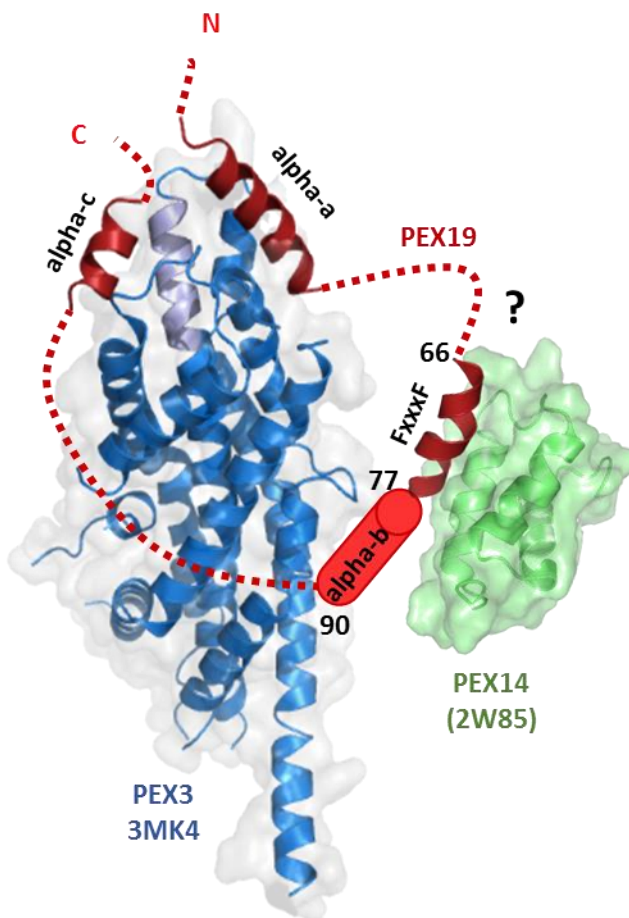


Figure 45: Model representing the potential bent conformation of the PEX19 N-terminus upon PEX3 binding. The names of the binding sites are in accordance with Chen *et al*, 2014. Amphipathic PEX19 helix alpha-a contains the primary PEX3 binding site, while helix alpha-c was indicated as a second binding site by XL-MS. The linking region contains a previously characterised FxxxF motif that can interact with PEX14 (green) and another predicted amphipathic helix, alpha-b (red cylinder). PEX3 is indicated with blue (PDB entry 3MK4), while its PEX19 binding helix is shown in light blue. PEX19 is shown in red with the missing chain represented by dashed lines and PEX14 in green.

4.1.4. PEX19 facilitates PEX14 solubility

Previous attempts at co-expression of PEX19 with PMPs were not successful, except for the case of PEX14. PEX14 contains a putative transmembrane domain (112-121) and has been classified as a PMP, even though it does not contain a conventional mPTS motif (Rottensteiner *et*

al, 2004). Yeast Pex14p, which is predicted to be membrane-associated and not a fully integral membrane protein, can be purified with the help of detergents in absence of Pex19p. On the other hand, human PEX14 is predicted to be integrally inserted to the peroxisomal membrane, making its expression and purification in full-length challenging.

In this work, we successfully expressed and purified the full length PEX14 after co-expression with PEX19 without use of detergents or any other solubilising means. We assume that PEX19 enhances PEX14 solubility in accordance with its chaperone-like capabilities. It is possible that the amphipathic helices of the PEX19 N-terminus form a protective “shield” around the transmembrane domain of PEX14, concealing its hydrophobicity from the aqueous environment of the buffer (in this case) or the cytosol (*in vivo*).

As PEX14 does not contain a canonical mPTS signal and its targeting to the peroxisomal membrane relies upon PEX13 rather than PEX19 (Girzalsky *et al*, 1999; Hashimoto *et al*, 2005), the basis of the PEX19-PEX14 interaction is still not well understood. The reported binding site has been mapped to the conserved N-terminus of PEX14 (residues 16-80) which also accommodates PEX5 binding. Further downstream interaction sites could explain how PEX19 is involved in preserving PEX14 solubility and are under investigation.

4.1.5. Structure analysis of the full-length PEX19-PEX14 complex using a hybrid approach

As the purification of the full length PEX19-PEX14 complex is possible *via* co-expression, structural analysis was performed. For this part of the project, a hybrid structural approach utilising SAXS, negative stain EM, and native MS was followed, yielding information about the overall structure of the complex.

The PEX19-PEX14 complex forms higher oligomeric assemblies that elute at a volume corresponding to higher molecular weight (~700 kDa) than the theoretical molecular weight expected for the complex. This can be attributed to the elongated shape of the particles, as well as the disorder present at the N-terminus of PEX19 and the C-terminus of PEX14. SEC-MALS analysis combined with native MS, revealed a lower molecular weight of ~240 kDa for the most prominent peak, in addition to a stoichiometry corresponding to two PEX19 molecules bound to

one PEX14. How the actual binding occurs is not known, but based on the ability of PEX19 to solubilise PEX14, an additional binding site proximal to the transmembrane domain could be possible. Furthermore, it is noteworthy that the complex is very stable to ionic strength variation and does not change its behaviour in a variety of buffers.

Initially, SAXS measurements were performed for the complex, which indicated a multi-domain, flexible, and elongated complex. As there was no prior information about the shape of the particle, several internal symmetry restrictions were imposed and all *ab initio* models resulted in elongated structures. To further validate our modelling results, negative stain EM was employed for visualisation of the PEX19-PEX14 complex. In accordance with our assumption, the complex displays an elongated shape that can be described by an internal P4 or P2 symmetry. Indeed, generation of SAXS models with imposed P4 symmetry results in shapes similar to the EM particles. Ultimately, the shape of the complex can be divided into two main domains: the elongated stem region, and the multi-globular upper region, within which 3-4 globular subdomains are visible (**Figure 46**).

Based on the derived stoichiometry, the hypothesis is as follows: two molecules of PEX14, which can dimerise *via* their coiled coil region, form the stem. Each PEX14 molecule can bind two PEX19 molecules, which are visible as globules at the upper part of the structure. Further structural analysis is required in order to delineate the complicated interactions that contribute to the assembly of this oligomer and increase our understanding of the molecular mechanisms that this complex is involved in.

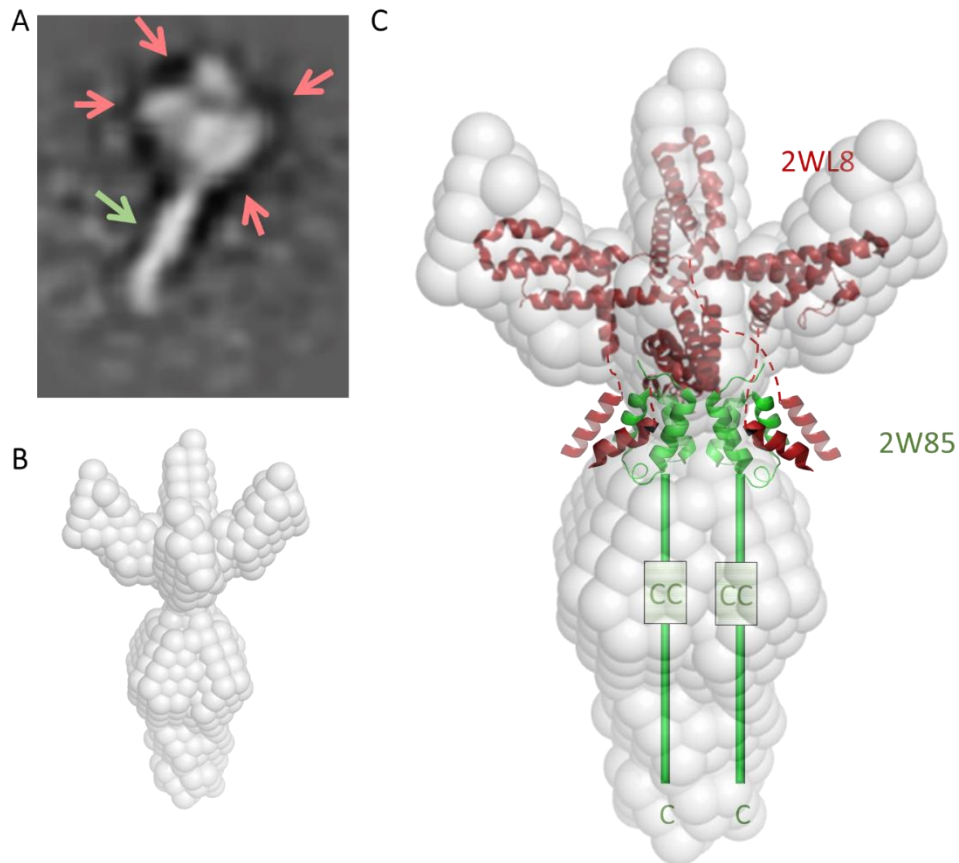


Figure 46: PEX19-PEX14 model. (A) Negative stain EM class average. Globular subdomains (PEX19) and stem region (PEX14) are indicated with salmon and green arrows, respectively. (B) P4 symmetry generated ab initio SAXS model. The shape of the model correlates well with the observed EM particle. (C) Potential model built using available structures of 2WL8 and 2W85. An additional PEX19 helix corresponding to the novel binding site has also been modelled on structure 2W85. PEX14 is coloured green while PEX19 is red.

4.2. Structural insight into peroxisomal matrix protein import

In this project, we investigated the structure of proteins and complexes involved in peroxisomal matrix translocation. Pcs60 is a peroxisomal matrix protein from yeast, which is involved in oxalate metabolism, while MIF1 is a peroxisomal enzyme from plants, involved in stress response. Peroxisomal matrix proteins contain a tripeptide sequence on their C-terminus, denoted as PTS1, which serves as a recognition and binding motif for the cycling receptor Pex5. In this thesis, the two novel crystal structures of Pcs60 and MIF1 were obtained by X-ray crystallography. Furthermore, the complex of Pcs60 with the peroxisomal receptor Pex5p from yeast was also investigated.

4.2.1. Structure of the novel peroxisomal enzyme Pcs60

In this thesis, Pcs60, a novel cargo protein from yeast, was expressed, purified, and its structure was resolved by means of X-ray crystallography. Pcs60 shows a concentration dependent oligomerisation and elutes as a dimer during SEC.

The crystal structure of this cargo protein was determined for the first time, using two obtained crystal types that had distinct symmetry ($P2_12_12_1$ and $P2_1$) and different number of molecules per asymmetric unit (6 and 12 respectively). Both crystal types contained a degraded 50 kDa fragment of Pcs60, which corresponds to its larger and conserved N-terminal domain, while one full-length copy was present in the second crystal type. The number of molecules in the asymmetric unit can be further decomposed to three dimers in the case of the $P2_12_12_1$ (which can form tetramers with their symmetry mates), and three tetramers (or six dimers) in the case of the $P2_1$ crystal type.

The full length Pcs60 contains a highly conserved AMP- and CoA- binding cavity and is located at the interface between the N-terminal and the smaller C-terminal domain. The structure can be described as an alternating α/β bundle connected by elongated loop regions, which accommodate the residues involved in dimer (and tetramer) formation. Dimers are held together by a number of hydrogen bonds and adopt an antiparallel orientation, with the ligand binding cavity exposed to the solvent. Tetramers can be characterised as dimers of dimers and are associated by a smaller number of hydrogen bonds. It is unclear at this point whether Pcs60 forms dimers in physiological conditions, but the homo-tetramerisation is likely an effect of high protein

concentrations. In any case, it is important to note that the missing C-termini of the tetramer which contain the Pex5p recognising motif SKL are exposed on the periphery of the tetramer, enabling Pex5p binding.

4.2.2. Structural analysis of Pex5p-Pcs60 complexes

To better understand the molecular mechanism of peroxisomal matrix protein import, we continued with the structural characterisation of Pex5p-Pcs60 complexes. Pex5p, similar to PEX19, contains an unstructured N-terminal domain followed by the TPR region which is known to recognise and bind to PTS1 signal carrying proteins. For that reason, shorter constructs containing only the TPR domain (Pex5p₍₃₁₃₋₆₁₂₎), as well as constructs containing an elongated region upstream of the TPR domain (Pex5p₍₁₉₈₋₆₁₂₎) were selected for crystallisation purposes.

Crystallisation of both constructs with Pcs60 yielded crystals, although only in the case of the longer Pex5p₍₁₉₈₋₆₁₂₎ construct it was possible to detect a hetero-dimer in the crystal content. The crystals of this complex were indexed with a P4 crystal symmetry, while the volume of the unit cell would allow one hetero-dimer per asymmetric unit.

Further analysis was performed for the Pex5p₍₃₁₃₋₆₁₂₎-Pcs60 and Pex5p₍₁₉₈₋₆₁₂₎-Pcs60 complexes by means of SAXS and indicates that both complexes likely exist as hetero-tetramer and higher-order oligomers in solution, with an observed concentration-dependant oligomerisation. The tetramers appear as flattened disks when imposing P4 symmetry in the *ab initio* modelling, with the extensions to the main core of the complex located on the outer edge of the disk. The predicted SAXS models were further validated *via* negative stain EM screening, after crosslinking of the sample. The

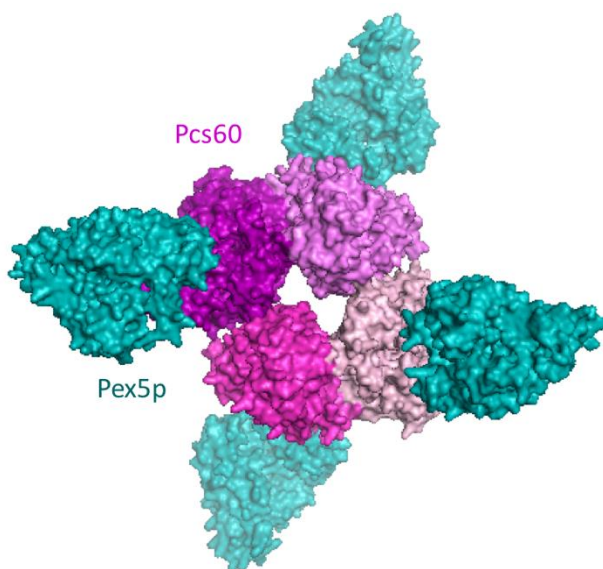


Figure 47: Model based on the Pcs60 crystal structure (pink), and homology models of the TPR domain of Pex5p (teal), calculated by SWISS-MODEL (Arnold *et al*, 2006; Biasini *et al*, 2014; Bordoli *et al*, 2008).

requirement of crosslinking for the hetero-tetramers to appear displays the concentration dependence of this oligomer formation. Indeed, in the EM images, it was possible to distinguish a central tetrameric core, which in all likelihood corresponds to the Pcs60 homo-tetramer, flanked by extensions on its edges, which correlate with our hypothesis of a peripheral Pex5p binding (**Figure 47**).

As this hetero-tetramer forms at higher concentrations only, we hypothesise that *in vivo*, complexes containing either a Pcs60 monomer bound to one Pex5p, or a Pcs60 dimer bound to two Pex5p molecules are likely to be involved in the import process. Alternatively, in cases of high metabolic requirement for Pcs60, its higher expression could result in homo-dimerisation or homo-tetramerisation that would allow faster and more efficient import. In that case, extending our hypothesis by our EM observations, which indicate that the complex does not contain Pex5p molecules occupying all Pcs60 C-termini, one Pex5p could be sufficient for oligomeric Pcs60 import (**Figure 48**).

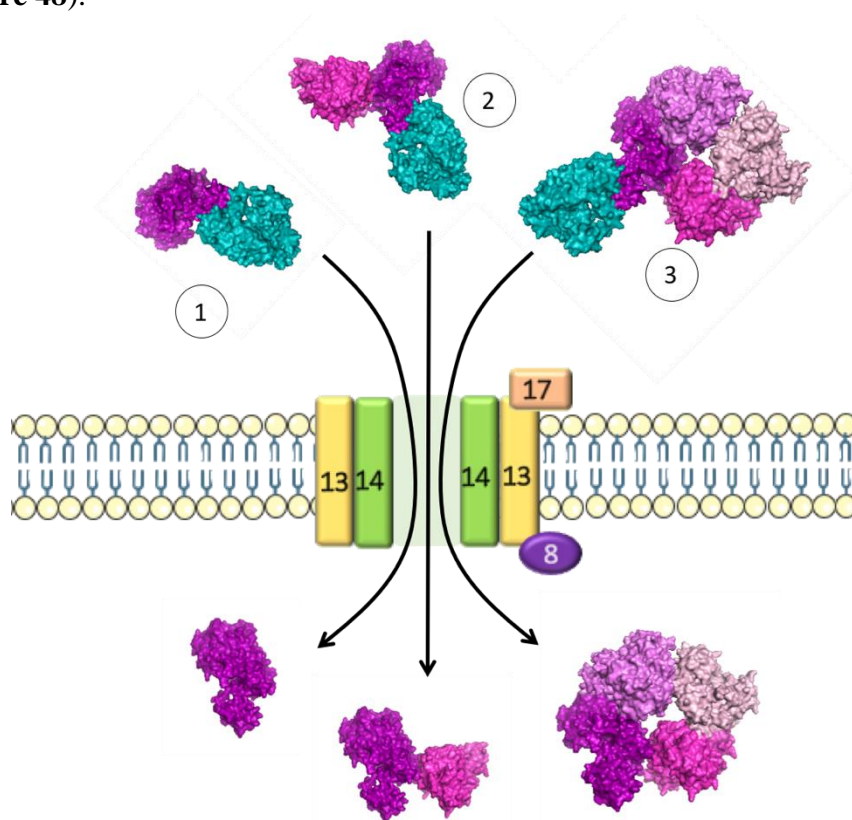


Figure 48: Possible Pcs60 import mechanisms. Pex5p (teal) could transport (1) a monomer, (2) a dimer, or (3) a tetramer of Pcs60 (pink) through the PTS1 pore. Pcs60 is coloured in pink, Pex5p in teal and represented as surfaces.

4.2.3. Characterisation of peroxisomal cargo protein MIF1 from plants

AtMIF1, similar to *Pcs60*, is a peroxisomal cargo protein. Unlike *Pcs60*, *AtMIF1* requires the formation of interfaces by its homo-trimerisation for its function, as these interfaces contain the ligand binding cavities with the Pro-2 (or Pro-1 after post-translational cleavage of Met-1) being the catalytic residue. This construct of *AtMIF1* contained a non-cleavable N-terminal purification tag, which prevented binding of the L-dopachrome substrate, as the additional residues concealed the catalytic Pro.

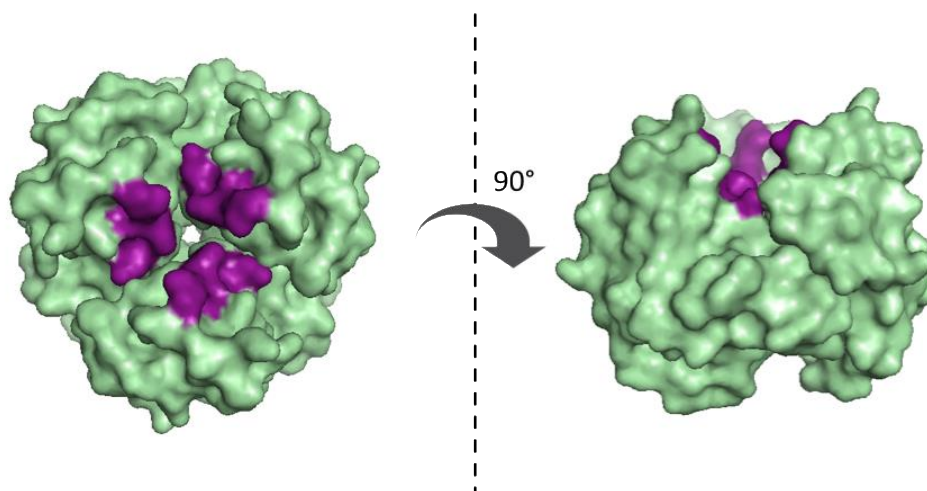


Figure 49: *AtMIF1* trimer PTS1 motif sites. The *AtMIF1* trimer is represented as a green surface, while the *AtPEX5* recognition sites, which are located at the top centre of the molecule, are coloured purple.

Although the organism of origin is different, the *Arabidopsis thaliana* mechanism of PTS1 import shares a lot of similarities with the one in yeast. To facilitate *AtPEX5* recognition and binding, *AtMIF1* contains an SKL motif at its C-terminus, which, even in the case of the trimer, is exposed and easily accessible by *AtPEX5* (**Figure 49**). Taking into account that *AtMIF1* functions as a trimer, we can perhaps speculate that one *AtPex5* receptor molecule would bind the trimer and import it through the PTS1 pore, which could be further investigated in future research.

5. Perspectives

Although a significant amount of structural information has been presented in this thesis, at this stage complementary experiments are still required in order to elucidate and describe the exact molecular mechanisms of PMP import and the potential role of the PEX19-PEX14 interaction. Moreover, additional open questions arise from our current investigation that could guide future research:

- How does PEX19 bind (and release) PMPs? How does the disorder of the N-terminus assist this process?
- What is the role of the PEX19-PEX14 interaction and what is the meaning behind the formation of its oligomeric assembly?
- Is there a competition between PEX3 and PEX14 for PEX19 binding? Could this potentially mediate the involvement of PEX19 in peroxisome biogenesis versus other roles?
- Is a PEX3-PEX19-PEX14 ternary complex possible to isolate and what is its potential function *in vivo*?

On the second half of this PhD thesis, the novel aspects of Pex5-mediated cargo import into peroxisomes was investigated. Two peroxisomal matrix proteins, which contain motifs for interaction with Pex5, namely Pcs60 and MIF1 were characterised by X-ray crystallography, and their structures were presented. Despite remaining questions regarding the mechanistic details of the import process, determination of a Pex5-cargo protein complex structure would provide considerable insight. At this point, the most important questions that need to be addressed include:

- Are there additional binding sites beyond the TPR domain of Pex5 that mediate binding with cargoes?
- Do Pcs60 and MIF1 become imported into peroxisomes in a monomeric or multimeric state?

To conclude, the research presented in this thesis will hopefully contribute to drawing a more elaborate model that describes the function of these two essential peroxisomal receptors and will further enhance our understanding of the molecular mechanisms that they are involved in.

6. Summary

Peroxisomes are dynamic eukaryotic organelles that require import of their membrane and matrix proteins by soluble receptors in the cytosol. Their biogenesis, protein import, and proliferation are regulated by a distinct set of proteins, collectively called peroxins.

Two peroxins, Pex3 and Pex19, are involved in the insertion of various Peroxisomal Membrane Proteins (PMPs) in the peroxisomal membrane and are thus crucial for its formation. Pex3 is a PMP that acts as a docking receptor for Pex19. The role of Pex19, in turn, is to bind, stabilise, and guide PMPs to the membrane-docked Pex3, where they are inserted into the peroxisomal membrane by an unknown mechanism. Pex14 is a PMP protein that is a key component of the peroxisomal import pore. Its interaction with Pex19 is established, yet little structural information is available about their full-length complex. In this thesis we investigated the conformation of the full-length human PEX19 in complex with the cytosolic domain of PEX3. A hybrid structural and biochemical approach was employed in order to characterize this interaction. Furthermore, the role of PEX19-PEX14 binding was addressed and the full-length complex was structurally characterised, providing insight into the stoichiometry, binding and shape of the never-before-described full-length assembly, using a variety of biochemical, biophysical and structural methods.

The second part of this thesis aims to shed light on the peroxisomal protein import mechanism process. Matrix proteins (cargoes) can be imported into the peroxisomal lumen using peroxisomal targeting signal 1 (PTS1), or peroxisomal targeting signal 2 (PTS2) import pathways. The PTS1 pathway is the most common, utilising peroxin Pex5p. Pex5p is a soluble receptor, cycling between a free cytoplasmic state -where it recognises and binds peroxisomal matrix proteins- and a membrane bound state, as part of the transient PTS1 pore. The interaction between Pex5p and cargo proteins occurs via the TPR domains (tetra-tricopeptide repeats) of Pex5p and the PTS1 sequence at the C-terminus of the cargo proteins. Two peroxisomal cargoes, Pcs60 and MIF1, were studied in this part of the PhD project. Pcs60 is a yeast peroxisomal oxalyl-CoA synthetase, while MIF1 is a plant peroxisomal cargo associated with stress response, both of which contain a PTS1 recognition signal peptide. These proteins were crystallised and structurally characterised by means of X-ray crystallography. Further characterisation of Pcs60 and its complex with Pex5p was performed by biophysical methods, small angle X-ray scattering (SAXS), and negative stain electron microscopy. Elucidation of the structure of the Pex5p-Pcs60 complex will lead to enhancing our understanding of peroxisomal protein import.

7. Zusammenfassung

Peroxisomen sind dynamische eukaryotische Organellen, welche ihre Membran- und Matrixproteine über lösliche Rezeptoren im Cytosol importieren. Ihre Biogenese, Proliferation und der Import von Proteinen werden durch einen eigenen Satz von Proteinen reguliert, die als Peroxine bezeichnet werden.

Zwei dieser Peroxine, Pex3 und Pex19, spielen eine Rolle in der Integration verschiedener peroxisomaler Membranproteine (PMPs) in die peroxisomale Membran und sind daher entscheidend für ihre Bildung. Pex3 ist ein PMP, welches als Kopplungsrezeptor für Pex19 dient. Die Aufgabe von Pex19 hingegen ist, PMPs zu binden, zu stabilisieren und zu dem an die Membran gekoppelten Pex3 zu leiten. Dort werden sie durch einen unbekanntem Mechanismus in die peroxisomale Membran eingefügt. Pex14 ist ein PMP, welches eine Leitkomponente der peroxisomalen Importpore darstellt. Seine Interaktion mit Pex19 ist bekannt, jedoch gibt es nur wenig strukturelle Information über den Vollängen-Komplex. In dieser Doktorarbeit untersuchten wir die Konformation des Vollängen-PEX19 in Komplex mit der cytosolischen Domäne von PEX3. Ein zweigleisiger, sowohl struktureller als auch biochemischer Ansatz kam zum Einsatz, um diese Interaktion zu charakterisieren. Zudem wurde sich mit der Funktion der PEX19-PEX14 Bindung und der strukturellen Charakterisierung des Vollängen-Komplexes befasst. Dabei wurden durch die Nutzung vielfältiger biochemischer, biophysikalischer und struktureller Methoden Einblicke in die Stöchiometrie, Bindung und Formgebung des nie zuvor beschriebenen Vollängen-Aufbaus erbracht.

Der zweite Teil dieser Doktorarbeit hat zum Ziel, Aufschluss über den Mechanismus des peroxisomalen Proteinimports zu geben. Matrixproteine (Kargos) können über den peroxisomalen Targeting-Signal 1 (PTS1) oder den peroxisomalen Targeting-Signal 2 (PTS2) Importweg eingebracht werden. Der PTS1 Importweg, der Peroxin Pex5p nutzt, ist der gebräuchlichste. Pex5p ist ein löslicher Rezeptor, der zwischen einem freien cytoplasmatischen Zustand – wo er peroxisomale Matrixproteine erkennt und bindet – und einem membrangebundenen Zustand, als Teil der kurzzeitig auftretenden PTS1 Pore, wechselt. Die Wechselwirkung zwischen Pex5p und Kargo Proteinen entsteht über die TPR Domänen (Tetratricopeptid-Wiederholungen) von Pex5p und der PTS1 Sequenz am C-Terminus der Kargo Proteine. Zwei peroxisomale Kargos, Pcs60 und MIF1, wurde in diesem Teil des PhD Projektes untersucht. Pcs60 ist ein peroxisomale Oxalyl-CoA Synthetase in Hefe, während MIF1 ein pflanzlicher peroxisomaler Kargo ist, der mit Stressantwort in Verbindung gebracht wird. Beide enthalten ein PTS1 Erkennungssignalpeptid. Diese Proteine wurden kristallisiert und mittels Röntgenkristallographie strukturell charakterisiert. Weiterhin wurde Pcs60 und sein Komplex mit Pex5p durch biophysikalische Methoden, Röntgen-Kleinwinkelstreuung (SAXS) und Negativ-Kontrast-Elektronenmikroskopie charakterisiert. Die Aufklärung der Struktur des Pex5p-Pcs60 Komplexes öffnet die Tür zu einem verbesserten Verständnis über den peroxisomalen Proteinimport.

8. Abbreviations

<i>A. thaliana</i>	<i>Arabidopsis thaliana</i>
APS	ammonium persulfate
ASU	asymmetric unit
AMP	adenosine monophosphate
ATP	adenosine triphosphate
BSA	bovine serum albumin
BS ³	bis-tris 2-(bis(2-hydroxyethyl)amino)-2-(hydroxymethyl)propane-1,3-diol
CC	coiled coil
CD	circular dichroism
CoA	coenzyme A
CTD	C-terminal domain
DNA	deoxyribonucleic acid
<i>E. coli</i>	<i>Escherichia coli</i>
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride
EDTA	ethylenediaminetetraacetate
EM	electron microscopy
ER	endoplasmic reticulum
ESI-MS	electrospray ionization mass spectrometry
ESRF	European Synchrotron Radiation Facility
FA	fluorescence anisotropy
FITC	fluorescein isothiocyanate
Fis1	fission protein 1
Get	guided entry of tail-anchored proteins
GST	Glutathione S-transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	high-performance liquid chromatography
IPTG	isopropyl- β -thiogalactopyranoside
LB	lysogeny broth
LC-MS	liquid chromatography mass spectrometry
MALDI-TOF	matrix-assisted laser desorption/ionization time of flight
MS	mass spectrometry
MALS	multiple angle light scattering
MES	2-(N-morpholino) ethanesulfonic acid
MS	mass spectrometry
mPTS	membrane peroxisomal targeting sequence
MW	molecular weight
MWCO	molecular weight cutoff
NHS	N-hydroxysuccinimide
NMR	nuclear magnetic resonance
NTA	nitrilotriacetic acid
PAGE	polyacrylamide gel electrophoresis
PBD	peroxisomal biogenesis disorder

PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol
PMP	peroxisomal membrane protein
PTS	peroxisome targeting signal
RALS	right angle light scattering
RCDP	rhizomelic chondrodysplasia punctata
Rmsd	root mean square deviation
ROS	reactive oxygen species
Rpm	revolutions per minute
SAXS	small angle x-ray scattering
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SDS	sodium dodecyl sulfate
SEC	size exclusion chromatography
SH3	Src homology 3
TA	tail-anchored
TCEP	tris-(2-carboxyethyl)phosphine
TEMED	tetramethylethylenediamine
TEV	tobacco etch virus
TMD	transmembrane domain
Tris	tris-(hydroxymethyl)aminomethane
TPR	tetra-trico-peptide repeat
UV	ultraviolet
VLCFA	very long chain fatty acid
XL-MS	crosslinking -mass spectrometry
ZSS	Zellweger syndrome spectrum

9. Bibliography

1. Adams PD, Afonine P V., Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH & IUCr (2010) *PHENIX*: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**: 213–221
2. Afonine P V, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH & Adams PD (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D. Biol. Crystallogr.* **68**: 352–67
3. Agrawal G & Subramani S (2013) Emerging role of the endoplasmic reticulum in peroxisome biogenesis. *Front. Physiol.* **4**: 286
4. Agrawal G & Subramani S (2016) De novo peroxisome biogenesis: Evolving concepts and conundrums. *Biochim. Biophys. Acta - Mol. Cell Res.* **1863**: 892–901
5. Albertini M, Rehling P, Erdmann R, Girzalsky W, Kiel JA, Veenhuis M & Kunau WH (1997) Pex14p, a peroxisomal membrane protein binding both receptors of the two PTS-dependent import pathways. *Cell* **89**: 83–92
6. Arakawa T & Timasheff SN (1982) Preferential interactions of proteins with salts in concentrated solutions. *Biochemistry* **21**: 6545–52
7. Aranovich A, Hua R, Rutenberg AD & Kim PK (2014) PEX16 contributes to peroxisome maintenance by constantly trafficking PEX3 via the ER. *J. Cell Sci.* **127**: 3675–86
8. Arnold K, Bordoli L, Kopp J & Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**: 195–201
9. Ashcroft NW & Mermin ND (1976) Solid state physics Holt, Rinehart and Winston
10. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T & Ben-Tal N (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**: W344–W350
11. Banerjee SK, Kessler PS, Saveria T & Parsons M (2005) Identification of trypanosomatid PEX19: functional characterization reveals impact on cell growth and glycosome size and number. *Mol. Biochem. Parasitol.* **142**: 47–55
12. Bharti P, Schliebs W, Schievelbusch T, Neuhaus A, David C, Kock K, Herrmann C, Meyer HE, Wiese S, Warscheid B, Theiss C & Erdmann R (2011a) PEX14 is required for microtubule-based peroxisome motility in human cells. *J. Cell Sci.* **124**:
13. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L & Schwede T (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**: W252–W258
14. Blanchet CE, Spilotros A, Schwemmer F, Graewert MA, Kikhney A, Jeffries CM, Franke D, Mark D, Zengerle R, Cipriani F, Fiedler S, Roessle M & Svergun DI (2015) Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J. Appl. Crystallogr.* **48**: 431–443
15. Blobel F & Erdmann R (1996) Identification of a yeast peroxisomal member of the family of AMP-binding proteins. *Eur. J. Biochem.* **240**: 468–76
16. Bordoli L, Kiefer F, Arnold K, Benkert J, Battey J & Schwede T (2008) Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc.* **4**: 1–13
17. van den Bosch H, Schutgens RBH, Wanders RJA & Tager JM (1992) Biochemistry of Peroxisomes. *Annu. Rev. Biochem.* **61**: 157–197
18. Bottger G, Barnett P, Klein AT, Kragt A, Tabak HF & Distel B (2000) *Saccharomyces cerevisiae* PTS1 receptor Pex5p interacts with the SH3 domain of the peroxisomal membrane protein Pex13p in an unconventional, non-PXXP-related manner. *Mol. Biol. Cell* **11**: 3963–76
19. Braverman N, Steel G, Obie C, Moser A, Moser H, Gould SJ & Valle D (1997) Human PEX7 encodes the

- peroxisomal PTS2 receptor and is responsible for rhizomelic chondrodysplasia punctata. *Nat. Genet.* **15**: 369–76
20. Brocard C & Hartig A (2006) Peroxisome targeting signal 1: Is it really a simple tripeptide? *Biochim. Biophys. Acta - Mol. Cell Res.* **1763**: 1565–1573
 21. Brocard C & Hartig A (2014) Molecular machines involved in peroxisome biogenesis and maintenance
 22. Brocard C, Kragler F, Simon MM, 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. *Biochem. Biophys. Res. Commun.* **204**: 1016–22
 23. Brosius U, Dehmel T & Gärtner J (2002) Two different targeting signals direct human peroxisomal membrane protein 22 to peroxisomes. *J. Biol. Chem.* **277**: 774–84
 24. Brul S, Westerveld A, Strijland A, Wanders RJ, Schram AW, Heymans HS, Schutgens RB, van den Bosch H & Tager JM (1988a) Genetic heterogeneity in the cerebrohepatorenal (Zellweger) syndrome and other inherited disorders with a generalized impairment of peroxisomal functions. A study using complementation analysis. *J. Clin. Invest.* **81**: 1710–5
 25. Brul S, Wiemer EA, Westerveld A, Strijland A, Wanders RJ, Schram AW, Heymans HS, Schutgens RB, Van den Bosch H & Tager JM (1988b) Kinetics of the assembly of peroxisomes after fusion of complementary cell lines from patients with the cerebro-hepato-renal (Zellweger) syndrome and related disorders. *Biochem. Biophys. Res. Commun.* **152**: 1083–9
 26. Burnett SF, Farré J-C, Nazarko TY & Subramani S (2015) Peroxisomal Pex3 activates selective autophagy of peroxisomes via interaction with the pexophagy receptor Atg30. *J. Biol. Chem.* **290**: 8623–31
 27. Caffrey M & Cherezov V (2009) Crystallizing membrane proteins using lipidic mesophases. *Nat. Protoc.* **4**: 706–731
 28. Calandra T & Roger T (2003) Macrophage migration inhibitory factor: a regulator of innate immunity. *Nat. Rev. Immunol.* **3**: 791–800
 29. Carvalho AF, Costa-Rodrigues J, Correia I, Costa Pessoa J, Faria TQ, Martins CL, Fransen M, Sá-Miranda C & Azevedo JE (2006) The N-terminal half of the peroxisomal cycling receptor Pex5p is a natively unfolded domain. *J. Mol. Biol.* **356**: 864–75
 30. Celniker G, Nimrod G, Ashkenazy H, Glaser F, Martz E, Mayrose I, Pupko T & Ben-Tal N (2013) ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. *Isr. J. Chem.* **53**: 199–206
 31. Chen Y, Pieuchot L, Loh RA, Yang J, Kari TMA, Wong JY & Jedd G (2014) Hydrophobic handoff for direct delivery of peroxisome tail-anchored proteins. *Nat. Commun.* **5**: 5790
 32. Choe J, Moyersoen J, Roach C, Carter TL, Fan E, Michels PAM & Hol WGJ (2003) Analysis of the Sequence Motifs Responsible for the Interactions of Peroxins 14 and 5, Which Are Involved in Glycosome Biogenesis in *Trypanosoma brucei* †. *Biochemistry* **42**: 10915–10922
 33. COLLINS K (2004) Ions from the Hofmeister series and osmolytes: effects on proteins in solution and in the crystallization process. *Methods* **34**: 300–311
 34. Costello JL, Castro IG, Camões F, Schrader TA, McNeill D, Yang J, Giannopoulou E-A, Gomes S, Pogenberg V, Bonekamp NA, Ribeiro D, Wilmanns M, Jedd G, Islinger M & Schrader M (2017) Predicting the targeting of tail-anchored proteins to subcellular compartments in mammalian cells. *J. Cell Sci.*: jcs.200204
 35. Delille HK, Agricola B, Guimaraes SC, Borta H, Lüers GH, Fransen M & Schrader M (2010) Pex11p β -mediated growth and division of mammalian peroxisomes follows a maturation pathway. *J. Cell Sci.* **123**:
 36. Deosaran E, Larsen KB, Hua R, Sargent G, Wang Y, Kim S, Lamark T, Jauregui M, Law K, Lippincott-Schwartz J, Brech A, Johansen T & Kim PK (2013) NBR1 acts as an autophagy receptor for peroxisomes. *J. Cell Sci.* **126**: 939–52
 37. van Dijken JP, Veenhuis M, Kreger-van Rij NJ & Harder W (1975) Microbodies in methanol-assimilating yeasts. *Arch. Microbiol.* **102**: 41–4
 38. Dimitrov L, Lam SK & Schekman R (2013) The role of the endoplasmic reticulum in peroxisome biogenesis. *Cold Spring Harb. Perspect. Biol.* **5**: a013243

39. Dodt G, Braverman N, Wong C, Moser A, Moser HW, Watkins P, Valle D & Gould SJ (1995) Mutations in the PTS1 receptor gene, PXR1, define complementation group 2 of the peroxisome biogenesis disorders. *Nat. Genet.* **9**: 115–125
40. Dodt G & Gould SJ (1996) Multiple PEX genes are required for proper subcellular distribution and stability of Pex5p, the PTS1 receptor: evidence that PTS1 protein import is mediated by a cycling receptor. *J. Cell Biol.* **135**: 1763–74
41. De Duve C & Baudhuin P (1966) Peroxisomes (microbodies and related particles). *Physiol. Rev.* **46**: 323–57
42. Dyer JM, McNew JA & Goodman JM (1996) The sorting sequence of the peroxisomal integral membrane protein PMP47 is contained within a short hydrophilic loop. *J. Cell Biol.* **133**: 269–80
43. EATON S, BARTLETT KB & POURFARZAM M (1996) Mammalian mitochondrial β -oxidation. *Biochem. J.* **320**:
44. Effelsberg D, Cruz-Zaragoza LD, Tonillo J, Schliebs W & Erdmann R (2015) Role of Pex21p for Piggyback Import of Gpd1p and Pnc1p into Peroxisomes of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **290**: 25333–42
45. Elferink RO, Marleen de V, Charles F & Groen A (2000) The role of phospholipids in bile formation: what can we learn from animals and human disease? *J. Hepatol.* **32**: 3–4
46. Emmanouilidis L, Schütz U, Tripsianes K, Madl T, Radke J, Rucktäschel R, Wilmanns M, Schliebs W, Erdmann R & Sattler M (2017) Allosteric modulation of peroxisomal membrane protein recognition by farnesylation of the peroxisomal import receptor PEX19. *Nat. Commun.* **8**: 14635
47. Emsley P, Lohkamp B, Scott WG & Cowtan K (2010) Features and development of *Coot*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**: 486–501
48. Erdmann R & Blobel G (1995) Giant peroxisomes in oleic acid-induced *Saccharomyces cerevisiae* lacking the peroxisomal membrane protein Pmp27p. *J. Cell Biol.* **128**: 509–23
49. Erdmann R & Schliebs W (2005a) Peroxisomal matrix protein import: the transient pore model. *Nat. Rev. Mol. Cell Biol.* **6**: 738–742
50. Erdmann R & Schliebs W (2005b) Peroxisomal matrix protein import: the transient pore model. *Nat. Rev. Mol. Cell Biol.* **6**: 738–42
51. Evans PR (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**: 282–292
52. Evans PR, Murshudov GN, K. C, K. D, A. KP, P. E, R. EP, C. FG, C. HK, S. F, K. W, R. G, F. LRM, P. BG, S. M, N. PA, R. H, W. K, A. KP, et al (2013) How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**: 1204–1214
53. Fagarasanu A, Fagarasanu M & Rachubinski RA (2007) Maintaining Peroxisome Populations: A Story of Division and Inheritance. *Annu. Rev. Cell Dev. Biol.* **23**: 321–344
54. Fagarasanu A, Mast FD, Knoblach B & Rachubinski RA (2010) Molecular mechanisms of organelle inheritance: lessons from peroxisomes in yeast. *Nat. Rev. Mol. Cell Biol.* **11**: 644–54
55. Fakieh MH, Drake PJM, Lacey J, Munck JM, Motley AM & Hettema EH (2013) Intra-ER sorting of the peroxisomal membrane protein Pex3 relies on its luminal domain. *Biol. Open* **2**: 829–37
56. Fang Y, Morrell JC, Jones JM & Gould SJ (2004) PEX3 functions as a PEX19 docking factor in the import of class I peroxisomal membrane proteins. *J. Cell Biol.* **164**: 863–75
57. Felber J-P & Golay A (1995) Regulation of nutrient metabolism and energy expenditure. *Metabolism* **44**: 4–9
58. Fodor K, Wolf J, Erdmann R, Schliebs W & Wilmanns M (2012) Molecular Requirements for Peroxisomal Targeting of Alanine-Glyoxylate Aminotransferase as an Essential Determinant in Primary Hyperoxaluria Type 1. *PLoS Biol.* **10**: e1001309
59. Foster J & Nakata PA (2014) An oxalyl-CoA synthetase is important for oxalate metabolism in *Saccharomyces cerevisiae*
60. Franke D, Kikhney AG & Svergun DI (2012) Automated acquisition and analysis of small angle X-ray scattering data. *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.* **689**: 52–59

61. Franke D, Svergun DI, IUCr, B. SH, T. DA, E. ME, A. RI, R. KA, H. R & I. SD (2009) *DAMMIF*, a program for rapid *ab-initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**: 342–346
62. Franssen M, Brees C, Ghys K, Amery L, Mannaerts GP, Ladant D & Van Veldhoven PP (2002) Analysis of mammalian peroxin interactions using a non-transcription-based bacterial two-hybrid assay. *Mol. Cell. Proteomics* **1**: 243–52
63. Franssen M, Vastiau I, Brees C, Brys V, Mannaerts GP & Van Veldhoven PP (2004) Potential role for Pex19p in assembly of PTS-receptor docking complexes. *J. Biol. Chem.* **279**: 12615–24
64. Franssen M, Vastiau I, Brees C, Brys V, Mannaerts GP & Van Veldhoven PP (2005) Analysis of human Pex19p's domain structure by pentapeptide scanning mutagenesis. *J. Mol. Biol.* **346**: 1275–86
65. Franssen M, Wylin T, Brees C, Mannaerts GP & Van Veldhoven PP (2001) Human pex19p binds peroxisomal integral membrane proteins at regions distinct from their sorting sequences. *Mol. Cell. Biol.* **21**: 4413–24
66. Freitas MO, Francisco T, Rodrigues TA, Alencastre IS, Pinto MP, Grou CP, Carvalho AF, Franssen M, Sa-Miranda C & Azevedo JE (2011) PEX5 Protein Binds Monomeric Catalase Blocking Its Tetramerization and Releases It upon Binding the N-terminal Domain of PEX14. *J. Biol. Chem.* **286**: 40509–40519
67. Freitas MO, Francisco T, Rodrigues TA, Lismont C, Domingues P, Pinto MP, Grou CP, Franssen M & Azevedo JE (2015) The peroxisomal protein import machinery displays a preference for monomeric substrates. *Open Biol.* **5**: 140236
68. Fujiki Y, Matsuzono Y, Matsuzaki T & Franssen M (2006) Import of peroxisomal membrane proteins: The interplay of Pex3p- and Pex19p-mediated interactions. *Biochim. Biophys. Acta - Mol. Cell Res.* **1763**: 1639–1646
69. Gandre-Babbe S & van der Blik AM (2008) The novel tail-anchored membrane protein Mff controls mitochondrial and peroxisomal fission in mammalian cells. *Mol. Biol. Cell* **19**: 2402–12
70. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD & Bairoch A (2005) Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook* pp 571–607. Totowa, NJ: Humana Press
71. Gatto GJ, Geisbrecht B V, Gould SJ & Berg JM (2000) Peroxisomal targeting signal-1 recognition by the TPR domains of human PEX5. *Nat. Struct. Biol.* **7**: 1091–5
72. Ghaedi K, Tamura S, Okumoto K, Matsuzono Y & Fujiki Y (2000) The peroxin pex3p initiates membrane assembly in peroxisome biogenesis. *Mol. Biol. Cell* **11**: 2085–102
73. Giannopoulou E-A, Emmanouilidis L, Sattler M, Dodt G & Wilmanns M (2016) Towards the molecular mechanism of the integration of peroxisomal membrane proteins. *Biochim. Biophys. Acta - Mol. Cell Res.* **1863**: 863–869
74. Girzalsky W, Hoffmann LS, Schemenewitz A, Nolte A, Kunau W-H & Erdmann R (2006) Pex19p-dependent targeting of Pex17p, a peripheral component of the peroxisomal protein import machinery. *J. Biol. Chem.* **281**: 19417–25
75. Girzalsky W, Rehling P, Stein K, Kipper J, Blank L, Kunau WH & Erdmann R (1999) Involvement of Pex13p in Pex14p localization and peroxisomal targeting signal 2-dependent protein import into peroxisomes. *J. Cell Biol.* **144**: 1151–62
76. Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E & Ben-Tal N (2003) ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* **19**: 163–4
77. Gloeckner CJ, Mayerhofer PU, Landgraf P, Muntau AC, Holzinger A, Gerber JK, Kammerer S, Adamski J & Roscher AA (2000) Human adrenoleukodystrophy protein and related peroxisomal ABC transporters interact with the peroxisomal assembly protein PEX19p. *Biochem. Biophys. Res. Commun.* **271**: 144–50
78. Glover JR, Andrews DW & Rachubinski RA (1994) *Saccharomyces cerevisiae* peroxisomal thiolase is imported as a dimer. *Proc. Natl. Acad. Sci. U. S. A.* **91**: 10541–5
79. Golisz A, Sugano M, Hiradate S & Fujii Y (2011) Microarray analysis of Arabidopsis plants in response to allelochemical l-DOPA. *Planta* **233**: 231–240
80. Götte K, Girzalsky W, Linkert M, Baumgart E, Kammerer S, Kunau WH & Erdmann R (1998) Pex19p, a farnesylated protein essential for peroxisome biogenesis. *Mol. Cell. Biol.* **18**: 616–28

81. Gould SG, Keller GA & Subramani S (1987) Identification of a peroxisomal targeting signal at the carboxy terminus of firefly luciferase. *J. Cell Biol.* **105**: 2923–31
82. Gould SJ, Keller GA, Hosken N, Wilkinson J & Subramani S (1989) A conserved tripeptide sorts proteins to peroxisomes. *J. Cell Biol.* **108**: 1657–64
83. Haan GJ, Faber KN, Baerends RJS, Koek A, Krikken A, Kiel JAKW, van der Klei IJ & Veenhuis M (2002) Hansenula polymorpha Pex3p Is a Peripheral Component of the Peroxisomal Membrane. *J. Biol. Chem.* **277**: 26609–26617
84. Hadden DA, Phillipson BA, Johnston KA, Brown L-A, Manfield IW, El-Shami M, Sparkes IA & Baker A (2006) Arabidopsis PEX19 is a dimeric protein that binds the peroxin PEX10. *Mol. Membr. Biol.* **23**: 325–36
85. Hagen S, Drepper F, Fischer S, Fodor K, Passon D, Platta HW, Zenn M, Schliebs W, Girzalsky W, Wilmanns M, Warscheid B & Erdmann R (2015) Structural insights into cargo recognition by the yeast PTS1 receptor. *J. Biol. Chem.* **290**: 26610–26
86. Halbach A, Landgraf C, Lorenzen S, Rosenkranz K, Volkmer-Engert R, Erdmann R & Rottensteiner H (2006) Targeting of the tail-anchored peroxisomal membrane proteins PEX26 and PEX15 occurs through C-terminal PEX19-binding sites. *J. Cell Sci.* **119**: 2508–2517
87. Halbach A, Lorenzen S, Landgraf C, Volkmer-Engert R, Erdmann R & Rottensteiner H (2005) Function of the PEX19-binding site of human adrenoleukodystrophy protein as targeting motif in man and yeast. PMP targeting is evolutionarily conserved. *J. Biol. Chem.* **280**: 21176–82
88. Halbach A, Rucktäschel R, Rottensteiner H & Erdmann R (2009) The N-domain of Pex22p can functionally replace the Pex3p N-domain in targeting and peroxisome formation. *J. Biol. Chem.* **284**: 3906–16
89. Hashimoto K, Kato Z, Nagase T, Shimozawa N, Kuwata K, Omoya K, Li A, Matsukuma E, Yamamoto Y, Ohnishi H, Tochio H, Shirakawa M, Suzuki Y, Wanders RJA & Kondo N (2005) Molecular Mechanism of a Temperature-Sensitive Phenotype in Peroxisomal Biogenesis Disorder. *Pediatr. Res.* **58**: 263–269
90. Hashimoto T (1999) Peroxisomal β -Oxidation Enzymes. *Neurochem. Res.* **24**: 551–563
91. Hattula K, Hirschberg D, Kalkkinen N, Butcher SJ & Ora A (2014) Association between the intrinsically disordered protein PEX19 and PEX3. *PLoS One* **9**: e103101
92. Heiland I & Erdmann R (2005) Biogenesis of peroxisomes. Topogenesis of the peroxisomal membrane and matrix proteins. *FEBS J.* **272**: 2362–72
93. Hettema EH, Erdmann R, van der Klei I & Veenhuis M (2014) Evolving models for peroxisome biogenesis. *Curr. Opin. Cell Biol.* **29**: 25–30
94. Hettema EH, Girzalsky W, van Den Berg M, Erdmann R & Distel B (2000) Saccharomyces cerevisiae pex3p and pex19p are required for proper localization and stability of peroxisomal membrane proteins. *EMBO J.* **19**: 223–33
95. Hoepfner D, Schildknecht D, Braakman I, Philippsen P & Tabak HF (2005) Contribution of the endoplasmic reticulum to peroxisome formation. *Cell* **122**: 85–95
96. Höhfeld J, Veenhuis M & Kunau WH (1991) PAS3, a Saccharomyces cerevisiae gene encoding a peroxisomal integral membrane protein essential for peroxisome biogenesis. *J. Cell Biol.* **114**:
97. Honsho M & Fujiki Y (2000) Topogenesis of Peroxisomal Membrane Protein Requires a Short, Positively Charged Intervening-loop Sequence and Flanking Hydrophobic Segments: STUDY USING HUMAN MEMBRANE PROTEIN PMP34. *J. Biol. Chem.* **276**: 9375–9382
98. HRUBAN Z & SWIFT H (1964) URICASE: LOCALIZATION IN HEPATIC MICROBODIES. *Science* **146**: 1316–8
99. Huhse B, Rehling P, Albertini M, Blank L, Meller K & Kunau WH (1998) Pex17p of Saccharomyces cerevisiae is a novel peroxin and component of the peroxisomal protein translocation machinery. *J. Cell Biol.* **140**: 49–60
100. Hunt JE & Trelease RN (2004) Sorting pathway and molecular targeting signals for the Arabidopsis peroxin 3. *Biochem. Biophys. Res. Commun.* **314**: 586–96
101. Islinger M, Li KW, Seitz J, Völkl A & Lüers GH (2009) Hitchhiking of Cu/Zn superoxide dismutase to

- peroxisomes--evidence for a natural piggyback import mechanism in mammals. *Traffic* **10**: 1711–21
102. Itoh R & Fujiki Y (2006) Functional domains and dynamic assembly of the peroxin Pex14p, the entry site of matrix proteins. *J. Biol. Chem.* **281**: 10196–205
103. Jones JM, Morrell JC & Gould SJ (2001) Multiple distinct targeting signals in integral peroxisomal membrane proteins. *J. Cell Biol.* **153**: 1141–50
104. Jones JM, Morrell JC & Gould SJ (2004) PEX19 is a predominantly cytosolic chaperone and import receptor for class I peroxisomal membrane proteins. *J. Cell Biol.* **164**: 57–67
105. Kabsch W (2010) XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66**: 125–32
106. Kammerer S, Arnold N, Gutensohn W, Mewes H-W, Kunau W-H, Höfler G, Roscher AA & Braun A (1997) Genomic Organization and Molecular Characterization of a Gene Encoding HsPXF, a Human Peroxisomal Farnesylated Protein. *Genomics* **45**: 200–210
107. Kammerer S, Holzinger A, Welsch U & Roscher AA (1998) Cloning and characterization of the gene encoding the human peroxisomal assembly protein Pex3p. *FEBS Lett.* **429**: 53–60
108. Kannenberg F, Ellinghaus P, Assmann G & Seedorf U (1999) Aberrant oxidation of the cholesterol side chain in bile acid synthesis of sterol carrier protein-2/sterol carrier protein-x knockout mice. *J. Biol. Chem.* **274**: 35455–60
109. Karplus PA & Diederichs K (2012) Linking Crystallographic Model and Data Quality. *Science (80-.)*. **336**: 1030–1033
110. Kashiwayama Y, Asahina K, Shibata H, Morita M, Muntau AC, Roscher AA, Wanders RJA, Shimozawa N, Sakaguchi M, Kato H & Imanaka T (2005) Role of Pex19p in the targeting of PMP70 to peroxisome. *Biochim. Biophys. Acta - Mol. Cell Res.* **1746**: 116–128
111. Keller GA, Gould S, Deluca M & Subramani S (1987) Firefly luciferase is targeted to peroxisomes in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **84**: 3264–8
112. Kemp S, Valianpour F, Mooyer PAW, Kulik W & Wanders RJA (2004) Method for Measurement of Peroxisomal Very-Long-Chain Fatty Acid α -Oxidation in Human Skin Fibroblasts Using Stable-Isotope-Labeled Tetracosanoic Acid. *Clin. Chem.* **50**: 1824–1826
113. Kiel JAKW, Otzen M, Veenhuis M & van der Klei IJ (2005) Obstruction of polyubiquitination affects PTS1 peroxisomal matrix protein import. *Biochim. Biophys. Acta* **1745**: 176–86
114. Kikhney AG & Svergun DI (2015a) A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **589**: 2570–2577
115. Kikhney AG & Svergun DI (2015b) A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **589**: 2570–2577
116. Kim PK, Hailey DW, Mullen RT & Lippincott-Schwartz J (2008) Ubiquitin signals autophagic degradation of cytosolic proteins and peroxisomes. *Proc. Natl. Acad. Sci. U. S. A.* **105**: 20567–74
117. Kim PK & Hettema EH (2015) Multiple Pathways for Protein Transport to Peroxisomes. *J. Mol. Biol.* **427**: 1176–1190
118. Kim PK & Mullen RT (2013) PEX16: a multifaceted regulator of peroxisome biogenesis. *Front. Physiol.* **4**: 241
119. Kim PK, Mullen RT, Schumann U & Lippincott-Schwartz J (2006) The origin and maintenance of mammalian peroxisomes involves a de novo PEX16-dependent pathway from the ER. *J. Cell Biol.* **173**: 521–32
120. van der Klei IJ, Yurimoto H, Sakai Y & Veenhuis M (2006) The significance of peroxisomes in methanol metabolism in methylotrophic yeast. *Biochim. Biophys. Acta - Mol. Cell Res.* **1763**: 1453–1462
121. Knoblach B, Sun X, Coquelle N, Fagarasanu A, Poirier RL & Rachubinski RA (2013) An ER-peroxisome tether exerts peroxisome population control in yeast. *EMBO J.* **32**: 2439–53
122. Knoops K, Manivannan S, Cepinska MN, Krikken AM, Kram AM, Veenhuis M & van der Klei IJ (2014) Preperoxisomal vesicles can form in the absence of Pex3. *J. Cell Biol.* **204**: 659–68
123. Koch A, Thiemann M, Grabenbauer M, Yoon Y, McNiven MA & Schrader M (2003) Dynammin-like protein 1 is involved in peroxisomal fission. *J. Biol. Chem.* **278**: 8597–605

-
124. Koch A, Yoon Y, Bonekamp NA, McNiven MA & Schrader M (2005) A Role for Fis1 in Both Mitochondrial and Peroxisomal Fission in Mammalian Cells. *Mol. Biol. Cell* **16**: 5077–5086
 125. Koch J & Brocard C (2012) PEX11 proteins attract Mff and human Fis1 to coordinate peroxisomal fission. *J. Cell Sci.* **125**: 3813–3826
 126. Koch J, Pranjic K, Huber A, Ellinger A, Hartig A, Kragler F & Brocard C (2010) PEX11 family members are membrane elongation factors that coordinate peroxisome proliferation and maintenance. *J. Cell Sci.* **123**:
 127. Konarev P V., Volkov V V., Sokolova A V., Koch MHJ, Svergun DI, IUCr, C. B, R. K, J. KMH, M. MS, J. D, D. G, R. W, H. GG, C. R, A. G, Y. H, Y. S, T. M, U. K, et al (2003) *PRIMUS*: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**: 1277–1282
 128. Kovacs WJ, Olivier LM & Krisans SK (2002) Central role of peroxisomes in isoprenoid biosynthesis. *Prog. Lipid Res.* **41**: 369–391
 129. Kragt A, Voorn-Brouwer T, van den Berg M & Distel B (2005a) The *Saccharomyces cerevisiae* Peroxisomal Import Receptor Pex5p Is Monoubiquitinated in Wild Type Cells. *J. Biol. Chem.* **280**: 7867–7874
 130. Kragt A, Voorn-Brouwer T, van den Berg M & Distel B (2005b) Endoplasmic reticulum-directed Pex3p routes to peroxisomes and restores peroxisome formation in a *Saccharomyces cerevisiae* pex3Delta strain. *J. Biol. Chem.* **280**: 34350–7
 131. Krissinel E & Henrick K (2007) Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372**: 774–797
 132. Kunau W-H, Dommès V & Schulz H (1995) β -Oxidation of fatty acids in mitochondria, peroxisomes, and bacteria: A century of continued progress. *Prog. Lipid Res.* **34**: 267–342
 133. Lam SK, Yoda N & Schekman R (2011) A vesicle carrier that mediates peroxisome protein traffic from the endoplasmic reticulum. *Proc. Natl. Acad. Sci. U. S. A.* **108**: E51–2
 134. Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T & Ben-Tal N (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**: W299–W302
 135. Langer G, Cohen SX, Lamzin VS & Perrakis A (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**: 1171–9
 136. Lanyon-Hogg T, Warriner SL & Baker A (2010) Getting a camel through the eye of a needle: the import of folded proteins by peroxisomes. *Biol. cell* **102**: 245–63
 137. Lazarow PB & De Duve C (1976) A fatty acyl-CoA oxidizing system in rat liver peroxisomes; enhancement by clofibrate, a hypolipidemic drug. *Proc. Natl. Acad. Sci. U. S. A.* **73**: 2043–6
 138. Lazarow PB & Fujiki Y (1985) Biogenesis of peroxisomes. *Annu. Rev. Cell Biol.* **1**: 489–530
 139. Van der Leij I, Van den Berg M, Boot R, Franse M, Distel B & Tabak HF (1992) Isolation of peroxisome assembly mutants from *Saccharomyces cerevisiae* with different morphologies using a novel positive selection procedure. *J. Cell Biol.* **119**: 153–62
 140. Li Z & Nair SK (2015) Structural Basis for Specificity and Flexibility in a Plant 4-Coumarate:CoA Ligase. *Structure* **23**: 2032–2042
 141. Liu X, Ma C & Subramani S (2012) Recent advances in peroxisomal matrix protein import. *Curr. Opin. Cell Biol.* **24**: 484–489
 142. Lodhi IJ & Semenkovich CF (2014) Peroxisomes: A Nexus for Lipid Metabolism and Cellular Signaling. *Cell Metab.* **19**: 380–392
 143. Long F, Vagin AA, Young P & Murshudov GN (2008) BALBES: a molecular-replacement pipeline. *Acta Crystallogr. D. Biol. Crystallogr.* **64**: 125–32
 144. Mannaerts GP & Van Veldhoven PP (1993) [Peroxisomal beta-oxidation]. *Verh. K. Acad. Geneesk. Belg.* **55**: 45–78
 145. Marshall PA, Krimkevich YI, Lark RH, Dyer JM, Veenhuis M & Goodman JM (1995) Pmp27 promotes peroxisomal proliferation. *J. Cell Biol.* **129**: 345–55
 146. Mast FD, Fagarasanu A & Rachubinski R (2010) The peroxisomal protein importomer: a bunch of transients with expanding waistlines. *Nat. Cell Biol.* **12**: 203–5
-

-
147. Matsuzaki T & Fujiki Y (2008) The peroxisomal membrane protein import receptor Pex3p is directly transported to peroxisomes by a novel Pex19p- and Pex16p-dependent pathway. *J. Cell Biol.* **183**: 1275–86
148. Matsuzono Y & Fujiki Y (2006) In vitro transport of membrane proteins to peroxisomes by shuttling receptor Pex19p. *J. Biol. Chem.* **281**: 36–42
149. Matsuzono Y, Kinoshita N, Tamura S, Shimozawa N, Hamasaki M, Ghaedi K, Wanders RJA, Suzuki Y, Kondo N & Fujiki Y (1999) Human PEX19: cDNA cloning by functional complementation, mutation analysis in a patient with Zellweger syndrome, and potential role in peroxisomal membrane assembly. *Proc. Natl. Acad. Sci.* **96**: 2116–2121
150. Matsuzono Y, Matsuzaki T & Fujiki Y (2006) Functional domain mapping of peroxin Pex19p: interaction with Pex3p is essential for function and translocation. *J. Cell Sci.* **119**: 3539–50
151. Matthews BW (1968a) Solvent content of protein crystals. *J. Mol. Biol.* **33**: 491–7
152. Matthews BW (1968b) Solvent content of protein crystals. *J. Mol. Biol.* **33**: 491–7
153. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**: 658–674
154. Meinecke M, Cizmowski C, Schliebs W, Krüger V, Beck S, Wagner R & Erdmann R (2010) The peroxisomal importomer constitutes a large and highly dynamic pore. *Nat. Cell Biol.* **12**: 273–7
155. Merk M, Mitchell RA, Endres S & Bucala R (2012) D-dopachrome tautomerase (D-DT or MIF-2): doubling the MIF cytokine family. *Cytokine* **59**: 10–17
156. Morgner N & Robinson C V. (2012) *Mass ign*: An Assignment Strategy for Maximizing Information from the Mass Spectra of Heterogeneous Protein Assemblies. *Anal. Chem.* **84**: 2939–2948
157. Motley AM & Hettema EH (2007) Yeast peroxisomes multiply by growth and division. *J. Cell Biol.* **178**: 399–410
158. Motley AM, Nuttall JM & Hettema EH (2012) Pex3-anchored Atg36 tags peroxisomes for degradation in *Saccharomyces cerevisiae*. *EMBO J.* **31**: 2852–68
159. Mukai S & Fujiki Y (2006) Molecular Mechanisms of Import of Peroxisome-targeting Signal Type 2 (PTS2) Proteins by PTS2 Receptor Pex7p and PTS1 Receptor Pex5pL. *J. Biol. Chem.* **281**: 37311–37320
160. Mullis K, Faloona F, Scharf S, Saiki R, Horn G & Erlich H (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* **51 Pt 1**: 263–73
161. Munck JM, Motley AM, Nuttall JM & Hettema EH (2009) A dual function for Pex3p in peroxisome formation and inheritance. *J. Cell Biol.* **187**: 463–471
162. Muntau AC, Roscher AA, Kunau W-H & Dodt G (2003) The interaction between human PEX3 and PEX19 characterized by fluorescence resonance energy transfer (FRET) analysis. *Eur. J. Cell Biol.* **82**: 333–42
163. Nair DM, Purdue PE & Lazarow PB (2004) Pex7p translocates in and out of peroxisomes in *Saccharomyces cerevisiae*. *J. Cell Biol.* **167**: 599–604
164. Neufeld C, Filipp F V, Simon B, Neuhaus A, Schüller N, David C, Kooshapur H, Madl T, Erdmann R, Schliebs W, Wilmanns M & Sattler M (2009) Structural basis for competitive interactions of Pex14 with the import receptors Pex5 and Pex19. *EMBO J.* **28**: 745–54
165. Neuhaus A, Kooshapur H, Wolf J, Meyer NH, Madl T, Saidowsky J, Hambruch E, Lazam A, Jung M, Sattler M, Schliebs W & Erdmann R (2014) A novel Pex14 protein-interacting site of human Pex5 is critical for matrix protein import into peroxisomes. *J. Biol. Chem.* **289**: 437–48
166. Niederhoff K, Meindl-Beinker NM, Kerksen D, Perband U, Schäfer A, Schliebs W & Kunau W-H (2005) Yeast Pex14p possesses two functionally distinct Pex5p and one Pex7p binding sites. *J. Biol. Chem.* **280**: 35571–8
167. Nordgren M & Fransen M (2014) Peroxisomal metabolism and oxidative stress. *Biochimie* **98**: 56–62
168. Nuttall JM, Motley A & Hettema EH (2011) Peroxisome biogenesis: recent advances. *Curr. Opin. Cell Biol.* **23**: 421–6
169. Otera H, Harano T, Honsho M, Ghaedi K, Mukai S, Tanaka A, Kawai A, Shimizu N & Fujiki Y (2000) The mammalian peroxin Pex5pL, the longer isoform of the mobile peroxisome targeting signal (PTS) type 1 transporter, translocates the Pex7p.PTS2 protein complex into peroxisomes via its initial docking site, Pex14p.
-

- J. Biol. Chem.* **275**: 21703–14
170. Otera H, Setoguchi K, Hamasaki M, Kumashiro T, Shimizu N & Fujiki Y (2002) Peroxisomal targeting signal receptor Pex5p interacts with cargoes and import machinery components in a spatiotemporally differentiated manner: conserved Pex5p WXXXXF/Y motifs are critical for matrix protein import. *Mol. Cell. Biol.* **22**: 1639–55
171. Otzen M, Rucktäschel R, Thoms S, Emmrich K, Krikken AM, Erdmann R & van der Klei IJ (2012) Pex19p contributes to peroxisome inheritance in the association of peroxisomes to Myo2p. *Traffic* **13**: 947–59
172. Panjkovich A & Svergun DI (2016) SASpy: a PyMOL plugin for manipulation and refinement of hybrid models against small angle X-ray scattering data. *Bioinformatics* **32**: 2062–2064
173. Petoukhov M V, Franke D, Shkumatov A V, Tria G, Kikhney AG, Gajda M, Gorba C, Mertens HDT, Konarev P V & Svergun DI (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**: 342–350
174. Petoukhov M V & Svergun DI (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* **89**: 1237–50
175. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC & Ferrin TE (2004) UCSF Chimera?A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**: 1605–1612
176. Platta HW, Hagen S & Erdmann R (2013) The exportomer: the peroxisomal receptor export machinery. *Cell. Mol. Life Sci.* **70**: 1393–1411
177. Platta HW, Hagen S & Erdmann R (2014) The Peroxisomal Exportomer. In *Molecular Machines Involved in Peroxisome Biogenesis and Maintenance* pp 347–370. Vienna: Springer Vienna
178. Platta HW, Magraoui F El, Schlee D, Grunau S, Girzalsky W & Erdmann R (2007) Ubiquitination of the peroxisomal import receptor Pex5p is required for its recycling. *J. Cell Biol.* **177**: 197–204
179. Purdue PE & Lazarow PB (2001) Peroxisome biogenesis. *Annu. Rev. Cell Dev. Biol.* **17**: 701–52
180. Purdue PE, Yang X & Lazarow PB (1998) Pex18p and Pex21p, a novel pair of related peroxins essential for peroxisomal targeting by the PTS2 pathway. *J. Cell Biol.* **143**: 1859–69
181. Pusey ML, Snyder RS & Naumann R (1986) Protein crystal growth. Growth kinetics for tetragonal lysozyme crystals. *J. Biol. Chem.* **261**: 6524–9
182. Reddy JK & Mannaerts GP (1994) Peroxisomal Lipid Metabolism. *Annu. Rev. Nutr.* **14**: 343–370
183. Reggiori F (2005) The Actin Cytoskeleton Is Required for Selective Types of Autophagy, but Not Nonspecific Autophagy, in the Yeast *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **16**: 5843–5856
184. Rehling P, Marzioch M, Niesen F, Wittke E, Veenhuis M & Kunau WH (1996) The import receptor for the peroxisomal targeting signal 2 (PTS2) in *Saccharomyces cerevisiae* is encoded by the *PAS7* gene. *EMBO J.* **15**: 2901–13
185. Rehling P, Skaletz-Rorowski A, Girzalsky W, Voorn-Brouwer T, Franse MM, Distel B, Veenhuis M, Kunau WH & Erdmann R (2000) Pex8p, an intraperoxisomal peroxin of *Saccharomyces cerevisiae* required for protein transport into peroxisomes binds the PTS1 receptor pex5p. *J. Biol. Chem.* **275**: 3593–602
186. Rinner O, Seebacher J, Walzthoeni T, Mueller L, Beck M, Schmidt A, Mueller M & Aebersold R (2008) Identification of cross-linked peptides from large sequence databases. *Nat. Methods* **5**: 315
187. Rottensteiner H, Kramer A, Lorenzen S, Stein K, Landgraf C, Volkmer-Engert R & Erdmann R (2004) Peroxisomal membrane proteins contain common Pex19p-binding sites that are an integral part of their targeting signals. *Mol. Biol. Cell* **15**: 3406–3417
188. Rucktäschel R, Girzalsky W & Erdmann R (2011) Protein import machineries of peroxisomes. *Biochim. Biophys. Acta - Biomembr.* **1808**: 892–900
189. Rucktäschel R, Thoms S, Sidorovitch V, Halbach A, Pechlivanis M, Volkmer R, Alexandrov K, Kuhlmann J, Rottensteiner H & Erdmann R (2009) Farnesylation of Pex19p is required for its structural integrity and function in peroxisome biogenesis. *J. Biol. Chem.* **284**: 20885–20896
190. Sacksteder KA (2000) PEX19 Binds Multiple Peroxisomal Membrane Proteins, Is Predominantly Cytoplasmic, and Is Required for Peroxisome Membrane Synthesis. *J. Cell Biol.* **148**: 931–944
191. Saidowsky J, Dodt G, Kirchberg K, Wegner A, Nastainczyk W, Kunau WH & Schliebs W (2001) The di-

- aromatic pentapeptide repeats of the human peroxisome import receptor PEX5 are separate high affinity binding sites for the peroxisomal membrane protein PEX14. *J. Biol. Chem.* **276**: 34524–9
- 192.Sato Y, Shibata H, Nakatsu T, Nakano H, Kashiwayama Y, Imanaka T & Kato H (2010) Structural basis for docking of peroxisomal membrane protein carrier Pex19p onto its receptor Pex3p. *EMBO J.* **29**: 4083–4093
- 193.Schliebs W & Kunau W-H (2004) Peroxisome membrane biogenesis: the stage is set. *Curr. Biol.* **14**: R397–9
- 194.Schliebs W & Kunau W-H (2006) PTS2 co-receptors: diverse proteins with common features. *Biochim. Biophys. Acta* **1763**: 1605–12
- 195.Schliebs W, Saidowsky J, Agianian B, Dodt G, Herberg FW & Kunau WH (1999) Recombinant human peroxisomal targeting signal receptor PEX5. Structural basis for interaction of PEX5 with PEX14. *J. Biol. Chem.* **274**: 5666–73
- 196.Schmidt F, Dietrich D, Eyllenstein R, Groemping Y, Stehle T & Dodt G (2012) The role of conserved PEX3 regions in PEX19-binding and peroxisome biogenesis. *Traffic* **13**: 1244–60
- 197.Schmidt F, Treiber N, Zocher G, Bjelic S, Steinmetz MO, Kalbacher H, Stehle T & Dodt G (2010) Insights into peroxisome function from the structure of PEX3 in complex with a soluble fragment of PEX19. *J. Biol. Chem.* **285**: 25410–25417
- 198.Schrader M, Bonekamp NA & Islinger M (2012) Fission and proliferation of peroxisomes. *Biochim. Biophys. Acta* **1822**: 1343–57 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22240198> [Accessed June 27, 2015]
- 199.Schrader M, Costello JL, Godinho LF, Azadi AS & Islinger M (2016) Proliferation and fission of peroxisomes — An update. *Biochim. Biophys. Acta - Mol. Cell Res.* **1863**: 971–983
- 200.Schrader M, Grille S, Fahimi HD & Islinger M (2013) Peroxisome interactions and cross-talk with other subcellular compartments in animal cells. *Subcell. Biochem.* **69**: 1–22
- 201.Schrodinger LLC (2010) The PyMOL Molecular Graphics System, Version 1.3r1.
- 202.Schueller N, Holton SJ, Fodor K, Milewski M, Konarev P, Stanley WA, Wolf J, Erdmann R, Schliebs W, Song Y-H & Wilmanns M (2010) The peroxisomal receptor Pex19p forms a helical mPTS recognition domain. *EMBO J.* **29**: 2491–2500
- 203.Shibata H, Kashiwayama Y, Imanaka T & Kato H (2004) Domain architecture and activity of human Pex19p, a chaperone-like protein for intracellular trafficking of peroxisomal membrane proteins. *J. Biol. Chem.* **279**: 38486–38494
- 204.Shimizu N, Itoh R, Hirono Y, Otera H, Ghaedi K, Tateishi K, Tamura S, Okumoto K, Harano T, Mukai S & Fujiki Y (1999) The peroxin Pex14p. cDNA cloning by functional complementation on a Chinese hamster ovary cell mutant, characterization, and functional analysis. *J. Biol. Chem.* **274**: 12593–604
- 205.Shimozawa N, Tsukamoto T, Nagase T, Takemoto Y, Koyama N, Suzuki Y, Komori M, Osumi T, Jeannette G, Wanders RJA & Kondo N (2004) Identification of a new complementation group of the peroxisome biogenesis disorders and PEX14 as the mutated gene. *Hum. Mutat.* **23**: 552–558
- 206.Shiozawa K, Konarev P V, Neufeld C, Wilmanns M & Svergun DI (2009) Solution structure of human Pex5.Pex14.PTS1 protein complexes obtained by small angle X-ray scattering. *J. Biol. Chem.* **284**: 25334–42
- 207.da Silva TF, Sousa VF, Malheiro AR & Brites P (2012) The importance of ether-phospholipids: A view from the perspective of mouse models. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1822**: 1501–1508
- 208.Singh I, Moser AE, Moser HW & Kishimoto Y (1984) Adrenoleukodystrophy: Impaired Oxidation of Very Long Chain Fatty Acids in White Blood Cells, Cultured Skin Fibroblasts, and Amniocytes. *Pediatr. Res.* **18**: 286–290
- 209.Smith JJ & Aitchison JD (2013) Peroxisomes take shape. *Nat. Rev. Mol. Cell Biol.* **14**: 803–17
- 210.Snyder WB, Koller A, Choy AJ & Subramani S (2000) The Peroxin Pex19p Interacts with Multiple, Integral Membrane Proteins at the Peroxisomal Membrane. *J. Cell Biol.* **149**:
- 211.Soukupova M, Sprenger C, Gorgas K, Kunau WH & Dodt G (1999) Identification and characterization of the human peroxin PEX3. *Eur. J. Cell Biol.* **78**: 357–74
- 212.Stanley WA, Filipp F V, Kursula P, Schüller N, Erdmann R, Schliebs W, Sattler M & Wilmanns M (2006)

- Recognition of a functional peroxisome type 1 target by the dynamic import receptor pex5p. *Mol. Cell* **24**: 653–63
213. Stanley WA & Wilmanns M (2006) Dynamic architecture of the peroxisomal import receptor Pex5p. *Biochim. Biophys. Acta* **1763**: 1592–8
214. Stein K, Schell-Steven A, Erdmann R & Rottensteiner H (2002) Interactions of Pex7p and Pex18p/Pex21p with the peroxisomal docking machinery: implications for the first steps in PTS2 protein import. *Mol. Cell. Biol.* **22**: 6056–69
215. Studier FW (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**: 207–34
216. Su J-R, Takeda K, Tamura S, Fujiki Y & Miki K (2009) Crystal structure of the conserved N-terminal domain of the peroxisomal matrix protein import receptor, Pex14p. *Proc. Natl. Acad. Sci. U. S. A.* **106**: 417–21
217. Subramani S (1998) Components involved in peroxisome import, biogenesis, proliferation, turnover, and movement. *Physiol. Rev.* **78**: 171–88
218. Swinkels BW, Gould SJ, Bodnar AG, Rachubinski RA & Subramani S (1991) A novel, cleavable peroxisomal targeting signal at the amino-terminus of the rat 3-ketoacyl-CoA thiolase. *EMBO J.* **10**: 3255–62
219. Szilard RK, Titorenko VI, Veenhuis M & Rachubinski RA (1995) Pay32p of the yeast *Yarrowia lipolytica* is an intraperoxisomal component of the matrix protein translocation machinery. *J. Cell Biol.* **131**: 1453–69
220. Tabak HF, Braakman I & van der Zand A (2013) Peroxisome formation and maintenance are dependent on the endoplasmic reticulum. *Annu. Rev. Biochem.* **82**: 723–44
221. Tam YYC, Fagarasanu A, Fagarasanu M & Rachubinski RA (2005) Pex3p Initiates the Formation of a Preperoxisomal Compartment from a Subdomain of the Endoplasmic Reticulum in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **280**: 34933–34939
222. Tanaka A, Yasuhara S, Kawamoto S, Fukui S & Osumi M (1976) Development of Microbodies in the yeast *Kloeckera* growing on methanol. *J. Bacteriol.* **126**: 919–27
223. Terlecky SR, Nuttley WM, McCollum D, Sock E & Subramani S (1995) The *Pichia pastoris* peroxisomal protein PAS8p is the receptor for the C-terminal tripeptide peroxisomal targeting signal. *EMBO J.* **14**: 3627–34
224. Theodoulou FL, Bernhardt K, Linka N & Baker A (2013) Peroxisome membrane proteins: multiple trafficking routes and multiple functions? *Biochem. J.* **451**: 345–52
225. Thompson JD, Higgins DG & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–80
226. Thoms S, Harms I, Kalies K-U & Gärtner J (2012) Peroxisome formation requires the endoplasmic reticulum channel protein Sec61. *Traffic* **13**: 599–609
227. Till A, Lakhani R, Burnett SF & Subramani S (2012a) Pexophagy: The Selective Degradation of Peroxisomes. *Int. J. Cell Biol.* **2012**: 1–18
228. Till A, Lakhani R, Burnett SF & Subramani S (2012b) Pexophagy: the selective degradation of peroxisomes. *Int. J. Cell Biol.* **2012**: 512721
229. Titorenko VI, Chan H & Rachubinski RA (2000) Fusion of small peroxisomal vesicles in vitro reconstructs an early step in the in vivo multistep peroxisome assembly pathway of *Yarrowia lipolytica*. *J. Cell Biol.* **148**: 29–44
230. Titorenko VI, Nicaud J-M, Wang H, Chan H & Rachubinski RA (2002) Acyl-CoA oxidase is imported as a heteropentameric, cofactor-containing complex into peroxisomes of *Yarrowia lipolytica*. *J. Cell Biol.* **156**: 481–94
231. Tria G, Mertens HDT, Kachala M & Svergun DI (2015) Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2**: 207–17
232. Tsukada H, Mochizuki Y & Fujiwara S (1966) The nucleoids of rat liver cell microbodies. Fine structure and enzymes. *J. Cell Biol.* **28**: 449–60

-
233. Tyanova S, Temu T, Carlson A, Sinitcyn P, Mann M & Cox J (2015) Visualization of LC-MS/MS proteomics data in MaxQuant. *Proteomics* **15**: 1453–1456
234. Veenhuis M & van der Klei IJ (2014) A critical reflection on the principles of peroxisome formation in yeast. *Front. Physiol.* **5** MAR:
235. Völkl A, Baumgart E & Fahimi HD (1988) Localization of urate oxidase in the crystalline cores of rat liver peroxisomes by immunocytochemistry and immunoblotting. *J. Histochem. Cytochem.* **36**: 329–36
236. Walton PA, Gould SJ, Feramisco JR & Subramani S (1992) Transport of microinjected proteins into peroxisomes of mammalian cells: inability of Zellweger cell lines to import proteins with the SKL tripeptide peroxisomal targeting signal. *Mol. Cell. Biol.* **12**: 531–41
237. Walton PA, Hill PE & Subramani S (1995) Import of stably folded proteins into peroxisomes. *Mol. Biol. Cell* **6**: 675–83
238. Wanders RJA (2004) Metabolic and molecular basis of peroxisomal disorders: A review. *Am. J. Med. Genet.* **126A**: 355–375
239. Wanders RJA & Waterham HR (2006) Biochemistry of mammalian peroxisomes revisited. *Annu. Rev. Biochem.* **75**: 295–332
240. Wanders RJ & Brites P (2010) Biosynthesis of ether-phospholipids including plasmalogens, peroxisomes and human disease: new insights into an old problem. *Clin. Lipidol.* **5**: 379–386
241. Wanders RJ, Jansen GA & Skjeldal OH (2001) Refsum disease, peroxisomes and phytanic acid oxidation: a review. *J. Neuropathol. Exp. Neurol.* **60**: 1021–31
242. Wang X, Unruh MJ & Goodman JM (2001) Discrete targeting signals direct Pmp47 to oleate-induced peroxisomes in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **276**: 10897–905
243. Wang Y, Yi H, Wang M, Yu O & Jez JM (2011) Structural and Kinetic Analysis of the Unnatural Fusion Protein 4-Coumaroyl-CoA Ligase::Stilbene Synthase. *J. Am. Chem. Soc.* **133**: 20684–20687
244. Waterham HR & Ebberink MS (2012) Genetics and molecular basis of human peroxisome biogenesis disorders. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1822**: 1430–1441
245. Weichenberger CX & Rupp B (2014) Ten years of probabilistic estimates of biocrystal solvent content: new insights via nonparametric kernel density estimate. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **70**: 1579–1588
246. Wickner W & Schekman R (2005) Protein Translocation Across Biological Membranes. *Science (80-.)*. **310**: 1452–1456
247. Will GK, Soukupova M, Hong X, Erdmann KS, Kiel JA, Dodt G, Kunau WH & Erdmann R (1999) Identification and characterization of the human orthologue of yeast Pex14p. *Mol. Cell. Biol.* **19**: 2265–77
248. Williams C, van den Berg M & Distel B (2005) *Saccharomyces cerevisiae* Pex14p contains two independent Pex5p binding sites, which are both essential for PTS1 protein import. *FEBS Lett.* **579**: 3416–3420
249. Williams C & van der Klei IJ (2013) Pexophagy-linked degradation of the peroxisomal membrane protein Pex3p involves the ubiquitin-proteasome system. *Biochem. Biophys. Res. Commun.* **438**: 395–401
250. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A & Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.* **67**: 235–42
251. Yagita Y, Hiromasa T & Fujiki Y (2013) Tail-anchored PEX26 targets peroxisomes via a PEX19-dependent and TRC40-independent class I pathway. *J. Cell Biol.* **200**: 651–66
252. Yamashita S, Abe K, Tatemichi Y & Fujiki Y (2014) The membrane peroxin PEX3 induces peroxisome-ubiquitination-linked pexophagy. *Autophagy* **10**: 1549–64
253. Yernaux C, Franssen M, Brees C, Lorenzen S & Michels PAM (2009) Trypanosoma brucei glycosomal ABC transporters: identification and membrane targeting. *Mol. Membr. Biol.* **23**: 157–72
254. van der Zand A, Braakman I & Tabak HF (2010) Peroxisomal Membrane Proteins Insert into the Endoplasmic Reticulum. *Mol. Biol. Cell* **21**: 2057–2065
255. van der Zand A, Gent J, Braakman I & Tabak HF (2012) Biochemically distinct vesicles from the
-

- endoplasmic reticulum fuse to form peroxisomes. *Cell* **149**: 397–409
256. Zhang X, Roe SM, Hou Y, Bartlam M, Rao Z, Pearl LH & Danpure CJ (2003) Crystal structure of alanine:glyoxylate aminotransferase and the relationship between genotype and enzymatic phenotype in primary hyperoxaluria type 1. *J. Mol. Biol.* **331**: 643–52

10. Appendix

10.1. DNA and protein sequences

Human PEX19

cDNA:

ATGGCCCGCCGCTGAGGAAGGCTGTAGTGTCGGGGCCGAAGCGGACAGGGAATTGGA
 GGAGCTTCTGGAAAGTGCTCTTGATGATTTTCGATAAAGCCAAACCCCTCCCCAGCACCCCC
 TTCTACCACCACGGCCCCCTGATGCTTCGGGGCCCCAGAAGAGATCGCCAGGAGACACTGC
 CAAAGATGCCCTCTTCGCTTCCCAAGAGAAGTTTTTCCAGGAACTATTTCGACAGTGA
 GGCTTCCCAAGCCACTGCGGAGTTCGAGAAGGCAATGAAGGAGTTGGCTGAGGAAGAACC
 CCACCTGGTGGAGCAGTTCACAAAAGCTCTCAGAGGCTGCAGGGAGAGTGGGCAGTGATAT
 GACCTCCCAACAAGAATTCACCTTCTTGCCTAAAGGAAACACTAAGTGGATTAGCCAAAA
 TGCCACTGACCTTCAGA
 ACTCCAGCATGTTCGGAAGAAGAGCTGACCAAGGCCATGGAGGG
 GCTAGGCATGGACGAAGGGGATGGGGAAGGGAACATCCTCCCCATCATGCAGAGTATTAT
 GCAGAACCTACTCTCCAAGGATGTGCTGTACCCATCACTGAAGGAGATCACAGAAAAGTA
 TCCAGAATGGTTGCAGAGTCATCGGGAATCTCTACCTCCAGAGCAGTTTGAAAAATATCA
 GGAGCAGCACAGCGTCATGTGCAAAAATATGTGAGCAGTTTGAGGCAGAGACCCCCACAGA
 CAGTCAAACCACTCAAAGGCTCGTTTTGAGATGGTGTGGATCTTATGCAGCAGCTACA
 AGATTTAGGCCATCTCCAAAAGAGCTGGCTGGAGAGATGCCTCCTGGCCTCAACTTTGA
 CCTGGATGCCCTCAATCTTTTCGGGCCACCAGGTGCCAGTGGTGAACAGTGTCTGATCAT
 GTGA

Protein:

>P40855

MAAAEEGCSVGAEADRELEELLESALDDFDKAKPSPAPPSTTTAPDASGPQKRSPGDTAK
 DALFASQEKFFQELFDSELASQATAEFEKAMKELAEELPHLVEQFQKLSEAAGRVSMT
 SQQEFTSCLKETLSGLAKNATDLQNSSMSEELTKAMEGLMDEGDGEGNILPIMQSIMQ
 NLLSKDVLYPSLKEITEKYPEWLQSHRESLPPEQFEKYQEQHSVMCKICEQFEAETPTDS
 ETTQKARFEMVLDLMQQQLQDLGHPPKELAGEMPPGLNFDLALNLSGPPGASGEQCLIM

Human PEX3

cDNA:

ATGCTGAGGTCTGTATGGAATTTTCTGAAACGCCACAAAAGAAATGCATCTTCTCTG
 GGCACGGTCCTTGGAGGAGTATATATTTCTGGGGAAATATGGACAGAAGAAAATCAGAGAA
 ATACAGGAAAGGGAGGCTGCAGAATACATTGCCCAAGCACGACGACAATATCATTTTGAA
 AGTAACCAGAGGACTTGCAATATGACAGTGCTGTCCATGCTTCCAACACTGAGAGAGGCC
 TTAATGCAGCAACTGAATTCCGAGAGCCTCACAGCTCTGCTAAAAAACAGGCCTTCAAAC
 AAGCTAGAAATATGGGAGGATCTGAAGATAATAAGTTTCACAAGAAGTACTGTGGCTGTA
 TACAGTACCTGTATGCTGGTTGTTCTTTTTCGGGGTCCAGTTAAACATAATTGGTGGATAT
 ATTTACCTGGATAATGCAGCAGTTGGCAAAAATGGCACTACAATCTTGCTCCCCAGAT
 GTCCAACAGCAGTATTTATCAAGTATTCAGCACCTACTTGGAGATGGCCTGACAGAATTG
 ATCACTGTCATTAACAAGCTGTGCAGAAGGTTTTAGGAAGTGTCTCTTAAACATTCT
 TTGTCCCTTTTGGACTTGGAGCAAAAACATAAAGAAATCAGAAATCTCGTTGAGCAGCAT
 AAGTCTTCTTCTTGGATTAATAAAGATGGATCCAAACCTTTATTATGCCATTATATGATG
 CCAGATGAAGAAACTCCATTAGCAGTGCAGGCCTGTGGACTTTCTCCTCGAGACATTACC

ACTATTAACTTCTCAATGAACTAGAGACATGTTGGAAAGCCCAGATTTTAGTACAGTT
 TTGAATACCTGTTTAAACCGAGGTTTTAGTAGACTTCTAGACAATATGGCTGAGTTCTTT
 CGACCTACTGAACAGGACCTGCAACATGGTAACTCTATGAATAGTCTTTCCAGTGTCAGC
 CTGCCTTTAGCTAAGATAATTCCAATAGTAAACGGACAGATCCATTTCAGTTTGCAGTGAA
 ACACCTAGTCATTTTGTTCAGGATCTGTTGACAATGGAGCAAGTGAAAGACTTTGCTGCT
 AATGTGTATGAAGCTTTTAGTACCCCTCAGCAACTGGAGAAATGA

Protein:**>P56589**

MLRSVWNFLKRHKKCI FLGTVLGGVYILGKYGQKKIREIQEREEAEYIAQARRQYHFES
 NQRTCNMVLSMLPTLREALMQQLNSESALTALLKNRPSNKLEIWEDLKIISFTRSTVAVY
 STCMLVLLRVQLNIIGGYIYLDNAAVGKNGTTILAPPDVQQQYLSSIQHLLGDGLTELI
 TVIKQAVQKVLGVSLSLKHSLSLDLEQKLKEIRNLVEQHKSSSWINKDGSKPLLCHYMMP
 DEETPLAVQACGLSPRDIITIKLLNETRDMLSPDFSTVLNNTCLNRGFSRLLDNMAEFFR
 PTEQDLQHGNSMNSLSSVSLPLAKIIPIVNGQIHSVCSETPSHFVQDLLTMEQVKDFAAN
 VYEAFSTPQQLEK

*Human PEX14***cDNA:**

ATGGCGTCTCGGAGCAGGCAGAGCAGCCGAGCCAGCCAAGCTCTACTCCAGGAA
 GTGAAAATGTGCTGCCTCGAGAGCCGCTGATTGCCACGGCAGTGAAGTTTCTACAGAATT
 CCCGGTCCGCCAGAGCCCACTTGCAACCAGGAGAGCATTCTAAAGAAGAAAGGGCTGA
 CAGATGAAGAGATTGATATGGCCTTCCAGCAGTCGGGCACTGCTGCCGATGAGCCTTCGT
 CCTTGGGCCCAGCCACACAGGTGGTTCCTGTCCAGCCCCCTCACCTCATATCTCAGCCAT
 ACAGTCCCGCAGGCTCCCGATGGCGAGATTACGGCGCCCTGGCCATCATCATGGCAGGCA
 TTGCATTTGGCTTTCACCAGCTCTACAAGAAATACCTGCTCCCCCTCATCCTGGGCGGCC
 GAGAGGACAGAAAGCAGCTGGAGAGGATGGAGGCCGGTCTCTCTGAGCTGAGTGGCAGCG
 TGGCCCAGACAGTGAAGTACAGACGACCCCTCGCTCCGTCAGGAGCTGCTGATTC
 AGCAGCAGCAGAAGATCCAGGAGCTTGCCACGAGCTGGCCGCTGCCAAGGCCACCACAT
 CCACCAACTGGATCCTGGAGTCCCAGAATATCAACGAAGTCAAGTCCGAAATTAACCTCT
 TGAAAGGGCTTCTTTAAATCGGAGGCAGTTCCTCCATCCCCATCAGCCCCGAAGATCC
 CCTCCTGGCAGATCCCAGTCAAGTCAACGTCACCTCCAGCCCTGCGGCCGTGAACCACC
 ACAGCAGCAGCGACATCTCACCTGTCAGCAACGAGTCCACGTCGTCCTCGCCTGGGAAGG
 AGGGCCACAGCCCCGAGGGCTCCACGGTCACTTACTTGGTGGGCCCCCAGGAGGAAG
 GCGAGGGGGTGGTGGACGTCAAGGGCCAGGTGCGGATGGAGGTGCAAGGCGAGGAGGAGA
 AGAGGGAGGACAAGGAGGACGAGGAGGATGAGGAGGATGATGATGTGAGCCATGTGGACG
 AGGAGGACTGCCTGGGGGTGCGAGGGGAGGACCGCGGGGCGGGGATGGGCAGATCAACG
 AGCAGGTGGAGAAGCTGCGGCGGCCGAGGGCGCCAGCAACGAGAGTGAAGCGGGACTAG

Protein:**>O75381**

MASSEQAEQPSQPSSTPGSENVLPREPLIATAVKFLQNSRVRQSPLATRRRAFLKKKGLTD
 EEIDMAFQQSGTAADEPSSLGPATQVVPVQPPHLISQPYSYPAGSRWRDYGALAIMAGIA
 FGFHQLYKKYLLPLILGGREDRKQLERMEAGLSELSGSVAQTVTQLQTTLASVQELLIQQ
 QQKIQELAHELAAAKATTSTNWILESQNINELKSEINSLKGLLLNRRQFPSPSPAPKIPS
 WQIPVKSPSPSSPAAVNHSSSDISPVSNESTSSSPGKEGHSPEGSTVTYHLLGQPQEEGE
 GVVDVKGQVRMEVQGEEEKREDKEDDEDEEDDDVSHVDEEDCLGVQREDRRGGDGQINEQ
 VEKLRPEGASNESERD

Yeast Pex5p**cDNA:**

ATGGACGTAGGAAGTTGCTCAGTGGGAAATAATCCGCTTGCGCAGTTGCACAAACATACT
 CAGCAGAACAAATCGCTTCAGTTTAATCAGAAGAATAATGGGCGTCTTAATGAGTCACCT
 CTACAGGGTACCAACAAGCCAGGTATTAGTGAGGCTTTTATATCCAATGTTAATGCTATT
 TCACAAGAAAACATGGCGAATATGCAAAGTTTCATAAACGGAGAACCACTGATCGATGAT
 AAAAGAAGAATGGAAATAGGGCCATCCTCAGGCAGGCTTCCACCTTTTTCAAACGTACAT
 TCTCTACAGACTTCAGCAAACCCAACCCAAATTAAGGGAGTGAACGATATATCTCATTGG
 TCACAGGAATTTCAAGGTAGTAATAGTATTCAAAATAGAAACGCGGATACAGGAAATTCA
 GAAAAGGCATGGCAGCGTGGCTCAACAACCGCATCAAGCCGGTTTCAGTACCCTAATACC
 ATGATGAATAACTATGCTTATGCTTCTATGAACAGTCTTAGTGGATCAAGGCTCCAATCG
 CCTGCTTTCATGAATCAACAACAGTCTGGTTCGTTCTAAAGAAGGAGTCAATGAGCAAGAA
 CAACAACCCTGGACAGATCAGTTTGAAAAGCTGGAAAAAGAAGTCTCAGAAAACCTGGAC
 ATAAATGATGAAATAGAGAAGGAGGAAAATGTGAGTGAAGTAGAACAAAACAAACCAGAA
 ACTGTTGAGAAGGAAGAAGGAGTATATGGAGATCAGTATCAATCTGATTTCCAAGAAGTG
 TGGGATAGCATAACAAGGACGCTGAAGAAGTCTTGCCATCCGAATTAGTTAATGATGAC
 CTCAATCTAGGAGAAGACTACTTGAAATATCTCGGCGGTAGAGTAAATGGGAACATCGAG
 TATGCTTTTCAATCTAACAACGAATATTTAATAATCCTAATGCTTATAAAATTTGGCTGC
 CTACTGATGGAAAACGGAGCCAAATTGAGCGAGGCAGCGCTAGCATTGGAAGCTGCTGTT
 AAAGAAAACCCGACCATGTGGATGCATGGCTAAGATTGGGTCTAGTACAAAACCCAGAAT
 GAAAAAGAGTTGAACGGCATAAGCGCCCTCGAAGAATGTTTAAAGTTAGACCCAAAGAAT
 CTGGAGGCAATGAAAACCTTAGCGATAAGTTATATAAACGAAGGTTATGATATGAGCGCC
 TTCACAATGCTGGATAAATGGGCAGAACTAAGTACCCGAAATTTGGTCAAGGATCAAG
 CAACAAGATGACAAAATTTCAAAAAGAGAAAGGGTTACCCATATTGATATGAACGCTCAT
 ATCACAAGCAATTTTGGCAACTAGCAAACAATTTAAGCACAATAGATCCTGAAATACAA
 CTATGCTTGGGTCTCTTATTTTACACGAAAGATGATTTTGACAAAACCATAGATTGCTTT
 GAAAGTGCCTTGAGGGTGAATCCTAATGACGAACTCATGTGGAATAGATTAGGGGCTTCA
 TTGGCCAATTTCCAATAGATCAGAGGAAGCAATCCAAGCCTATCATAGGGCACTACAATA
 AAACCTTCTTTTGTAGAGCTCGCTATAATCTGGCGGTATCATCCATGAATATAGGCTGT
 TTCAAAGAAGCAGCAGGCTACTTATTAAGTGTCTAAGTATGCATGAAGTGAACACTAAT
 AATAAAAAAGGAGACGTTGGATCTCTCTTGAATACGTACAATGATACTGTTATAGAGACT
 TTGAAGAGAGTTTTTATAGCGATGAATAGAGATGATTTACTTCAAGAAGTGAAGCCAGGC
 ATGGACCTGAAAAGATTTAAAGGAGAATTTTCGTTTTGA

Protein:**>P35056**

MDVGSCSVGNPLAQLHKHTQQNKSLQFNQKNNGRNLNESPLQGTNKPGESEAFISNVNAI
SQENMANMQRFINGEPLIDDKRRMEIGPSSGRLPPFSNVHSLQTSANPTQIKGVNDISHW
SQEFQGSNSIQNRNADTGNSEKAWQRGSTTASSRFQYPNTMMNNYAYASMNSLSGSRLOS
PAFMNQQQSGRSKEGVNEQEQQPWTQDFEKLKLEKVENLDINDEIEKEENVSEVEQNKPE
TVEKEEGVYGDQYQSDQFEVWDSIHKDAEEVLPSELVNDDLNLGEDYLKYLGGRVNGNIE
YAFQSNNEYFNPNAYKIGCLLMENGAKLSEAALAFEAAVKEKPDHVDLRLGLVQTQN
EKELNGISALEECLKLDPKNLEAMKTLAISYINEGYDMSAFTMLDKWAETKYPIWSRIK
QQDDKFQKEKGFTHIDMNAHITKQFLQLANLSTIDPEIQQLCLGLLFYTKDDFDKTI DCF
ESALRVNPNDELMWNRLGASLANRSEEAIQAYHRALQLKPSFVRARYNLAVSSMNI GC
FKEAAGYLLSVLSMHEVNTNKKGDVGSLLNTYNDTVIETLKRVIAMNRDDLQEVKPG
MDLKRFKGEFSF

*Yeast Pcs60***cDNA:**

ATGACAAGTGCCGCTACTGTTACTGCTTCGTTCAACGATACTTTTAGCGTATCCGATAAT
GTCGCCGTTATTGTTCCCTGAAACGGACACTCAGGTGACCTACAGGGATCTATCCCACATG
GTGGGTCACTTCCAGACCATGTTACAAATCCTAATTCTCCATTGTACGGAGCTGTTTTTC
AGACAAGATACAGTGGCGATATCCATGCGTAATGGGCTGGAATTTATCGTCGCTTTCCTC
GGTGCTACTATGGACGCTAAAATTGGCGCGCCCTTGAATCCCAATTATAAGGAAAAGGAG
TTCAATTTTTATTTGAATGACCTGAAATCTAAGGCGATTTGCGTCCCAAAGGGTACCACA
AAGTTACAGAGTTCTGAAATCTAAAATCTGCCTCCACGTTTGGATGTTTTATCGTAGAG
CTGGCCTTCGATGCGACCAGGTTTAGGGTAGAGTATGATATATACTCTCCAGAGGACAAC
TACAAAAGGGTTATTTACCGGTCTTTGAACAACGCCAAATTTGTCAACACAAATCCCCTT
AAATTCCTGGGTTTGGCCGTTCCAGTGACGTTGCCCTGATTTTGCATACCAGTGGTACC
ACCTCCACTCCAAAACGGTGCCTTTGTTACATTTGAACATTGTGAGAAGCACGTTGAAC
ATTGCTAACACTTACAAGCTAACGCCCTTGGACAGATCTTATGTCGTGATGCCTCTTTTC
CACGTCCATGGGTTAATTGGTGTTTTACTTTCCACTTTTAGAACTCAGGGTTCTGTGTG
GTTCCCGATGGATTCCATCCAAAGTTATTCTGGGACCAATTTGTTAAGTACAACGTGTAAT
TGGTTCAGTTGCGTTCCACAATAAGCATGATTATGCTGAACATGCCCAAACCAAACCCT
TTCCACACATTAGATTCATCAGATCGTGTTCTTCTGCTTTGGCTCCAGCAACGTTCAT
AAGCTGGAGAAGGAATTCAATGCACCTGTCTTGGAGGCCTATGCGATGACCGAAGCATCA
CATCAAATGACCTCAAACAATCTGCCTCCAGGAAAAGAGAAAGCCTGGTACTGTGGGCCAG
CCACAAGGAGTCAACGTCGTATTCTAGATGACAATGACAATGTCTTGCCCCCGGGCAA
GTCGGCGAAGTTTCCATCAGAGGCGAAAACGTCACTTTGGGGTATGCTAATAATCCAAA
GCTAACAAGGAGAACTTACCAAGAGAGAGAACTATTTCAGAACCGGTGACCAAGGTTAT
TTCGACCCTGAGGGGTTTTTGGTCTTACAGGCAGAATCAAAGAGCTTATCAACAGGGGT
GGTGAAAAGATTTACCCATTGAGCTCGACGGCATTATGCTATCGCATCCAAAGATCGAT
GAAGCCGTTGCATTTGGTGTTCCTCCGACGATATGTACGGCCAAGTAGTTCAAGCCGCCATT
GTTTTGAAGAAGGGAGAAAAATGACCTACGAAGAACTGGTGAACCTTCTTAAAGAAGCAC
CTAGCCTCTTTCAAATTTCCAACCAAGGTGTACTTTGTTGATAAGCTACCAAAAACCGCT
ACAGGTAAAATCCAGAGAAGAGTTATCGCAGAACTTTTGCTAAGAGCAGCAGAAATAAG
AGTAAGTTGTAG

Protein:

>P38137

MTSAATVTASFNDTFSVSDNVAVIVPETDTQVTYRDLSHMVGHFQTMFTNPNSPLYGAVF
RQDTVAISMRNGLEFIVAFLGATMDAKIGAPLNPYKEKEFNFYLN DLKSKAICV PKGTT
KLQSSEILKSASTFGCFIVELAFDATRFRVEYDIYSPEDNYKRVIYRSLNNAKFVNTNPV
KFPGFARSSDVALILHTSGTTSTPKTVPLLHLNIVRSTLNIANTYKLTPLDRSYVVMPLF
HVHGLIGVLLSTFRTQGSVVVPDGFHPKLFWDQFVKYNCNWFSCVPTISMIMLNMPKPNP
FPHIRFIRSCSSALAPATFHKLEKEFNAPVLEAYAMTEASHQMTSNNLPPGKRKPGTVGQ
PQGVTVVILDDNDNVLP PGKVGEVSIRGENVT LGYANNPKANKENFTKRENYFRTGDQGY
FDPEGFLVLTGRIKELINRGGEKISPIELDGIMLSHPKIDEAVAFGVPDDMYGQVVQAAI
VLKKGEKMTYEELVNFLKKHLASFKIPTKVYFVDKLPKTATGKIQRRIAETFAKSSRNK
SKL

AtMIF1

cDNA:

ATGCCTTGTCTTTACATTACAACAAACGTCAATTTTGACGGCGTTAACACCGATCCGTTT
TACTCCGAAGTCACCAAAGCCGTCGCTTCCATCGTCGGACGACCTCAA AACTTAGTGATG
GTGGTGTGAAGGGATCAGTAGAGATAGTATTTGGAGGGAACAAAGAAGCAGCTGCATAT
GCAGAGATTGTGTCAATGGGAGGCATCACCAAACAAGTTAAGAGAGAACTCATAGCGACC
GTTGGTTCTATTCTTCACACTCATTTTTCTATTTCATCCCACTCGTTTTATCTTTAAAGTT
TTTGATATTAATTCTTTGCCTCTTCCTTCTAAACTTTAG

Protein:

>AEE78824.1

MPCLYITTNVNF DGVNTDPFYSEVTKAVASIVGRPQNLVMVVLKGSVEIVFGGNKEAAAY
AEIVSMGGITKQVKRELIATVGSILHTHFSIHPTRFIFKVFDINSLPLPSKL

11. *Curriculum vitae*

Personal information

Name	Evdokia Anastasia Giannopoulou
Date of birth	10 April 1988
Place of birth	Patras, Greece
Nationality	Greek

Education

PhD studies	European Molecular Biology Laboratory (EMBL) Universitätsklinikum Hamburg-Eppendorf (UKE)
Research internship	EMBL Hamburg 2012-2013
M.Sc. studies	University of Patras, Department of Biology (2010-2012)
B.Sc. studies	University of Patras, Department of Biology (2006-2010)

Scientific Publications

1. J. Costello, I. Castro, F. Camões, T. Schrader, D. McNeall, J. Yang, E.A. Giannopoulou, S. Gomes, V. Pogenberg, N. Bonekamp, D. Ribeiro, M. Wilmanns, G. Jedd, M. Islinger, M. Schrader, Predicting the targeting of tail-anchored proteins to subcellular compartments in mammalian cells, *J Cell Sci.*, March 2017.
2. E.A. Giannopoulou, L. Emmanouilidis, M. Sattler, G. Dodt, M. Wilmanns, Towards the molecular mechanism of the integration of peroxisomal membrane proteins, *Biochim Biophys Acta.* 1863(5):863-9, May 2016.
3. S. Fili, A. Valmas, M. Norrman, G. Schluckebier, D. Beckers, T. Degen, J. Wright, A. Fitch, F. Gozzo, A.E. Giannopoulou, F. Karavassili, I. Margiolaki, Human insulin polymorphism upon ligand binding and pH variation: the case of 4-ethylresorcinol, *IUCrJ.* 2(Pt 5):534-44, August 2015.
4. A. Valmas, K. Magiouf, S. Fili, M. Norrman, G. Schluckebier, D. Beckers, T. Degen, J. Wright, A. Fitch, F. Gozzo, A.E. Giannopoulou, F. Karavassili, I. Margiolaki, Novel crystalline phase and first-order phase transitions of human insulin complexed with two distinct phenol derivatives, *Acta Crystallogr D Biol Crystallogr.*, 71(Pt 4):819-28, April 2015.
5. I. Margiolaki, A.E. Giannopoulou, J. Wright, L. Knight, M. Norrman, G. Schluckebier, A. Fitch, R.B. Von Dreele, High-resolution powder X-ray data reveal the T(6) hexameric form of bovine insulin. *Acta Crystallogr D Biol Crystallogr.*, 69(Pt 6):978-90, June 2013.
6. F. Karavassili, A.E. Giannopoulou, E. Kotsiliti, L. Knight, M. Norrman, G. Schluckebier, L. Drube, A. Fitch, J. Wright, I. Margiolaki, Structural studies of human insulin cocrystallized with phenol or resorcinol via powder diffraction. *Acta Crystallogr D Biol Crystallogr.*, 68(Pt 12):1632-41, December 2012.

Funding

1. Marie Curie Initial Training Networks (ITN) - PERFUME Early Stage Researcher (ESR) Scholarship, funded by European Union Seventh Framework Programme under Grant Agreement No. 316723, funded 2013- 2016
2. Deutsche Forschungsgemeinschaft (DFG) programme PerTrans (Structure And Function Of The Peroxisomal Translocon), funded 2016- 2017

12. Acknowledgements

Firstly, I would like to thank my supervisor at EMBL, Dr. Matthias Wilmanns, for giving me the opportunity to work in his group and for trusting me with such a challenging project. I would also like to thank Professor Georg Lüers for agreeing to be my supervisor at the Universitätsklinikum Hamburg-Eppendorf, as well as Dr. Christian Löw, both of who supported me throughout my PhD as members of my Thesis Advisory Committee (TAC).

I would like to express my gratitude to Dr. Daniel Passon, for guiding me during the first years of my PhD and giving me the chance to continue his work on Pcs60. His contribution of protocols and plasmids was invaluable. Additionally, I would like to thank Dr. Nabil Hanna, for his great contribution in the Pcs60-Pex5p complex. His optimised crystals were the grounds for all research performed on this complex since. I would also like to express my gratitude to Dr. Spyros Chatziefthimiou, for being my supervisor during my trainee days at EMBL Hamburg. He taught me how to express and purify proteins, how to solve protein structures and how to multitask. Furthermore, I would like to point out his contribution in determining the challenging Pcs60 structure.

I am thankful to all members of the PEX group at EMBL Hamburg, Dr. Daniel Passon, Dr. Nabil Hanna, Dr. Vadim Rimsa, Thomas Seine, Dr. Anne-Sophie Huart, Dr. Jérôme Bürgi, and Morlin Milewski for the useful conversations, Dr. Vivian Pogenberg for help with biophysical methods and fluorescence anisotropy, Dr. Annabel Parret for her support, as well as the whole Wilmanns group for the nice working environment.

I would like to acknowledge our collaborators from the EMBL-Hamburg BioSAXS group, Dr. Haydyn Mertens (for help with the Pex19-related complexes), and Dr. Cy Jeffries (for help with the Pex5p-Pcs60 complexes), as well as the beamline scientists at the P12 beamline at Petra III. I would also like to acknowledge the scientists at the Sample Preparation and Characterisation facility (SPC) at EMBL- Hamburg for their support and the beamline scientists at the ID30A-1 / MASSIF-1 beamline for assistance with data collection. I would also like to acknowledge Dr. Christos Gatsogiannis, Dr. Bjoern Klink and Pascal Lill from the Max Planck Institute at

Dortmund for the electron microscopy analysis, as well as Dr. Friedel Drepper, and Julian Bender from the University of Freiburg for the mass spectrometry of crosslinked samples and native mass spectrometry analysis. Lastly, my thanks go to our collaborators from the University of Stavanger and University of Hamburg, Dr. Delphine Crape and Professor Sigrun Reumann for the MIF1 project.

Moreover, I would like to thank my amazing friends, Anne-Sophie Huart, Emma-Ruoqi Xu, Diana Mendes-Freire and Kate Beckham for putting up with me every day, despite the mood swings and frustrations. I will really miss you all!

Last but not least, I would not have managed anything without the support of my wonderful family, my parents, Dimitrios Giannopoulos and Calliope Rigopoulos and my little sister Konstantina Giannopoulou. Thank you for always believing in my crazy ideas, even if it means going all the way to the other side of the world.

13. **Eidesstattliche Versicherung**

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe.

Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.

Ich erkläre mich einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Unterschrift: