Re-Design of a Receptor-Targeting Signal Interaction to Create a New Peroxisomal Trafficking Pathway

Laura Louise Cross

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds

Astbury Centre for Structural Molecular Biology

August 2016

The candidate confirms that the submitted work is her own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms the appropriate credit has been given within the thesis where reference has been made to the work of others.

Sections 1.2 and 1.3 of Chapter 1 (Introduction) of the thesis are adapted from a jointly authored publication: Cross LL, Ebeed HT, Baker A. Peroxisome biogenesis, protein targeting mechanisms and PEX gene functions in plants. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 2016. 1863(5):850-62. L.L. Cross (the candidate) wrote section 4 (Peroxisomal matrix protein import), produced figures 1 and 2, and contributed towards section 1 (Introduction) and section 6 (Conclusions and future perspectives) (all authors contributed towards these sections). Other authors wrote the other sections of the review article, and HT Ebeed produced table 1.

Acknowledgements

Firstly, thanks to my supervisors: Stuart Warriner, Alison Baker, and Alan Berry. Stuart, thank you for your support, your excitement about the project, for teaching me how to code, and for all your help with the screen processing; Alison, thank you for your honesty, and for giving me the chance to publish with you; Alan, thank you for the opportunity to carry out this research, and for introducing me to the exciting and infuriating nature of directed evolution. Thank you to Nicola Skoulding and Tom Lanyon-Hogg, who built a good platform from which this research could be built (and Tom for your advice at the beginning of the project). I would like to thank the following people, who have helped in this research. Their contributions were as follows:

i) **Sarah Gunn:** Cloning and sequencing of the *At*PEX5(340-728) construct.
ii) **Sam Liver:** Site-directed mutagenesis of *At*PEX5(340-728) for the creation of variant D507K.
iii) **Hersey Underwood:** Site-directed mutagenesis of *At*PEX5(340-728) for the creation of variants D505H, D505K-D507K, V533W, T536W, N537Q and N537T.
iv) **Heba Ebeed:** The finding of *Physcomitrella patens* homologues corresponding to PTS1 proteins in *Arabidopsis thaliana*.
v) **Rupesh Paudyal:** All *in vivo* work: cloning of PTS1 and candidate PTS1* sequences downstream of fluorescent proteins, cloning of *At*PEX5(340-728) (and my favourite variant) for transient expression in *Physcomitrella patens*, and the many hours spent behind the microscope.

I would also like to thank Chiara Airoldi and Yasuko Kamisugi for performing initial *in vivo* experiments.

Thank you to the biology and chemistry lab groups that I have had the pleasure of being split between. To the Berry lab: thank you for keeping things lively (and colourful with all the chilli plants!). To lab 1.49: thank you for being so welcoming and creating such a friendly atmosphere to work in. I would especially like to thank Diana (your steaks can't be beaten) and Claire for the fun trip to Japan, Kat for all of your advice and lending me such a comfy chair to write up from, Chadamas for being such a smiley person to work with, Chris for the useful biology chats, James W for the useful chemistry chats, Dan and Sílvia for chemistry help and fun, Ivona, Valeria and Phil, Zoe, Matt, Marion, Andrea and Tatiana. Thank you to Chi Trinh and Nasir Khan in

Astbury, and Chris Empson and Martin Huscroft in Chemistry for your invaluable help with equipment.

Katie, thank you for the relaxing breaks, de-stressing chats and pub trips. Rong, thank you for all the trips away and the badminton. To my oldest friends Emma, Fiona, Christine and Sam, thank you for always being there for a laugh (even in a flood emergency situation).

Alun (revenge at last)- thank you for being my rock: *"always there, an odd shape and cold to the touch!"*. Thank you for being so supportive and putting up with me, no matter how stressed and annoying I must have been over the past year(s).

To my family, thank you for your support and for everything really. With dad remembering the word 'receptor' and mum remembering the word 'peroxisome', together you've got this!

Abstract

Peroxisomes are crucial for cell survival but do not possess a genome, so protein import is essential for correct functioning of the organelle. The PTS1 (peroxisomal targeting signal 1) pathway is the major, and in some organisms the only, transport pathway for proteins to enter the peroxisome. PTS1 is a recognition sequence at the C-terminus of peroxisomal cargo proteins, which allows their binding to the receptor protein PEX5 (peroxin 5). PEX5 then acts as the vehicle for transporting PTS1-cargo proteins into peroxisomes.

This work concerned the disruption of the natural PEX5:PTS1 interaction and the generation of an interaction between a mutated form of PEX5 (PEX5*) and a peptide representing a non-natural PTS1 (PTS1*). An *in vitro* protein-peptide binding screen was developed to test *Arabidopsis thaliana* PEX5-C (*At*PEX5-C), and variants, with a library of peptides to identify corresponding binding peptides. For wild-type *At*PEX5-C, peptides identified were consistent with bioinformatics predictions for plant PTS1 sequences, confirming the validity of the screen. Fluorescence anisotropy was then used to validate binding peptides, revealing that two variants of *At*PEX5-C, D505H and N601A, exhibited reduced affinity to a representative native PTS1 (YQSKL-CO_2H), yet increased affinity to a non-PTS1 sequence (YQSYY-CO_2H). When these two mutations were combined, the affinity for YQSKL was further reduced and the affinity for YQSYY was further enhanced. This effect was amplified when an additional mutation, D507T, was incorporated into the *At*PEX5-C double variant. The resulting triple variant was termed PEX5*. Various 9-amino acid sequences were added upstream of YQSYY, and the optimal resulting 14-amino acid sequence was termed PTS1*.

The *in vitro*-validated binding of PEX5* to PTS1* was tested *in vivo* (in the moss *Physcomitrella patens*; by R. Paudyal) to investigate whether expression of PEX5* would allow the import of a fluorescent protein with a C-terminal PTS1* tag into peroxisomes.

Table of Contents

Abbreviations

Amino acids

Amino acid	Three-letter code	One-letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Organisms

At	*Arabidopsis thaliana*
Hs	*Homo sapiens*
Mm	*Mus musculus*
Phypa	*Physcomitrella patens*
Sc	*Saccharomyces cerevisiae*
Tb	*Trypanosoma brucei*

Units

Å	angstroms
bp	base pairs
°C	degrees Celsius
cm	centimetres
Da	daltons
g	gram
x g	times gravity (relative centrifugal force)
h	hour(s)
kDa	kilodaltons
L	litre
M	moles per L (molar)
mg	milligram
min	minute(s)
mL	millilitre
mM	millimolar
mol	mole
µg	microgram
µL	microlitre
µM	micromolar
µm	micrometres
ng	nanogram
nM	nanomolar
nm	nanometres
s	seconds
v/v	volume per volume
w/v	weight per volume

General

aa	Amino acid(s)
AAA	ATPases associated with various cellular activities
AI	Autoinduction
ATP	Adenosine triphosphate
Boc	*tert*-Butoxycarbamate
BSA	Bovine serum albumin

C18	Octadecyl carbon chain
CD	Circular dichroism
CFP	Cyan fluorescent protein
DCM	Dichloromethane
DIPEA	*N,N*-Diisopropylethylamine
DMF	*N,N*-Dimethylformamide
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic Acid
dNTPs	Deoxynucleotides
EDC	1-(3-Dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride
EDTA	Ethylenediaminetetraacetic acid
EIC	Extracted ion chromatogram
ER	Endoplasmic reticulum
ESI(+)	Positive electrospray ionisation
Et_2O	Diethyl ether
FA	Fluorescence anisotropy
Fmoc	Fluorenylmethoxycarbonyl
GFP	Green fluorescent protein
HCTU	2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate
HPLC	High-performance liquid chromatography
HRP	Horseradish peroxidase
IC_{50}	Half-maximal inhibitory concentration
IPTG	Isopropyl-β-D-thiogalactoside
K_d	Dissociation constant
K_i	Inhibition constant
LB	Lysogeny broth
LC	Liquid chromatography
LC-MS	Liquid chromatography-mass spectrometry
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
LoxP	Locus of X(cross)-over in P1
M + H	Mass + a proton
m/z	Mass-to-charge ratio
OD	Optical density
Pbf	2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction

PDB	Protein Data Bank
PEG	Polyethylene glycol
PEX	Peroxin
pK_a	Logarithmic acid dissociation constant
PMP	Peroxisomal membrane protein
PTS	Peroxisomal targeting signal
PWM	Positional weight matrix
Q-TOF	Quadrupole time-of-flight
ReT	Retention time
RFP	Red fluorescent protein
RING	Really interesting new gene
RT	Room temperature
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SH3	Src homology 3
SPPS	Solid phase peptide synthesis
SPR	Surface plasmon resonance
tBu	Tertiary butyl
TCEP	Tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid
TPR	Tetratricopeptide repeat
Trt	Trityl
VBA	Visual Basic for Applications
WT	Wild-type
YFP	Yellow fluorescent protein

Chapter 1

Introduction: A Way to Re-Purpose Peroxisomes?

The overall aim of this research is to manipulate protein import into peroxisomes, 're-engineering' the peroxisome as a synthetic organelle. The peroxisome is ideally suited to this purpose, as it acquires all constituent proteins from the cytosol through conserved protein import pathways (Smith and Aitchison, 2013). This means that the function of the peroxisome is determined by the proteins imported, so it could be possible to manipulate the peroxisome's function by targeting these conserved import pathways. This method could allow non-peroxisomal proteins to be directed to the peroxisome, whilst discriminating against natural peroxisomal cargo.

1.1 Plant peroxisomes; adaptable cellular organelles

Peroxisomes are cellular membrane-bound organelles involved in pathways critical to the function of the cell. They are present in almost all eukaryotic cells and, among other functions, play a role in β-oxidation (for the breakdown of fatty acids to produce energy, and the production of plant hormones for effective growth, development, and function of the plant), and the metabolism of reactive oxygen species (ROS) (reviewed in Hu et al., 2012). Based on the requirements of the cell, peroxisomes can have different additional functions so adaptability is an important feature of these organelles (Mast et al., 2015; Goto-Yamada et al., 2015).

Peroxisomes were first observed as "Microbodies" in 1954 by electron microscopy (Rhodin, 1954; Bernhard and Rouiller, 1956), within mouse kidney cells, and the successful isolation of microbodies from rat liver cells was first published in 1965 (Baudhuin et al., 1965). These microbodies were subsequently named peroxisomes because of their involvement in the metabolism of the reactive oxygen species hydrogen peroxide (de Duve and Baudhuin, 1966). The first observation of peroxisomes in plant cells, termed "phragmosomes" at the time, was published in 1958 (Porter and Caulfield, 1958; Porter and Machado, 1960; Manton, 1961). A review was

published in 1966, remarking on the morphological similarity between these phragmosomes in plants (then re-named “plant microbodies”) and the animal microbodies being reported at the time (Mollenhauer et al., 1966).

There are different types of peroxisomes in plants, including glyoxysomes, leaf peroxisomes, and unspecialised peroxisomes (Olsen, 1998; Hayashi et al., 2000a; Kamada et al., 2003). These have different functions and so each requires a different overall protein content; peroxisomes are able to handle a high protein concentration in order to carry out their range of functions (Heupel et al., 1991). The different specialised peroxisomes are still collectively referred to as peroxisomes, as they share many of the same enzymes involved in β-oxidation (Pracharoenwattana and Smith, 2008). In **Figure 1**, the right-hand panel (insert) shows a section of a plant cell treated with the stain DAB (3,3’-diaminobenzidine). This is a cytochemical stain for catalase, a protein that is found abundantly within the peroxisome.

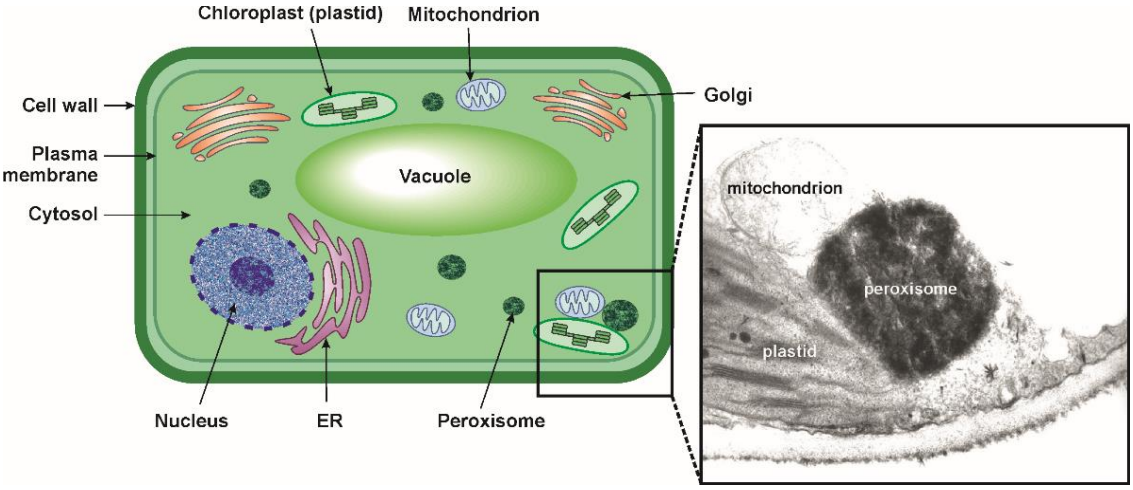


Figure 1| The peroxisome in the context of a cartoon plant cell. Right-hand panel (insert) image courtesy of Alison Baker. A DAB (3,3’-diaminobenzidine) stain for catalase has been used for this section of the plant cell, highlighting the peroxisome. ER, endoplasmic reticulum.

The peroxisome does not possess its own genome and instead acquires proteins from the cytosol, through an import pathway which allows fully folded proteins to enter the matrix of the peroxisome (reviewed in Lanyon-Hogg et al., 2010). Here, proteins can function within the peroxisome as needed for the requirements of the cell. Peroxisomes have evolved to change their protein content quickly in response to their environment, through protein import and degradation pathways within the organelle. Peroxisome turnover and synthesis also contribute to the change in peroxisomal function (Young and Bartel, 2016).

1.1.1 Adaptability of peroxisomal function

Various pathways are in place within the peroxisome for the turnover of proteins, in order to facilitate substantial changes in their protein content for a change in overall function. For example, in plants, during the early stages of post-germinative growth (before seedlings begin to photosynthesise) cells contain specialised peroxisomes known as glyoxysomes which contain enzymes involved in the glyoxylate cycle. After β-oxidation for the metabolism of storage oils (an energy source within the plant), the glyoxylate cycle is responsible for the subsequent production of carbohydrate to gain energy in order to form a shoot during germination. When the plant greens and begins to photosynthesise, it requires peroxisomes to take part in a new process within leaf cells: photorespiration. Photorespiration takes place to carry out the recycling of 2-phosphoglycolate, a by-product formed when the enzyme rubisco catalyses the addition of O_2 to ribulose-1,5-biphopshate, rather than the addition of CO_2, as is the case during photosynthesis (Zhu et al., 2010). In the peroxisomes of a photosynthesising plant, the glyoxylate pathway is redundant so plants are able to convert glyoxysomes to leaf peroxisomes able to take part in photorespiration. This process is known as the 'functional transition' of plant peroxisomes (Goto-Yamada et al., 2015).

There were two hypotheses in place for how peroxisomes undergo the functional transition: the 'one-population' hypothesis, that peroxisomes retain their structure and membrane barrier to the cytosol but their entire protein content is replaced; and the 'two-population' hypothesis, that peroxisomes containing newly redundant proteins are gradually turned over to allow the population of the cell by new peroxisomes with the required function (Beevers, 1979). Recent research suggests that the one-population hypothesis is the predominant mode of functional transition: the way in which peroxisomes are able to adapt to the change is through dedicated protein import pathways, introducing new proteins from the cytosol, and the degradation of existing redundant proteins within the peroxisome (reviewed in Goto-Yamada et al., 2015). Redundant proteins in the peroxisome are degraded by proteases (Goto-Yamada et al., 2014).

It appears that any remaining peroxisomes that have not undergone the functional transition are degraded in an autophagy process specific to peroxisomes. This process is known as pexophagy (Farmer et al., 2013; Kim et al., 2013; Goto-Yamada et al.,

2014; Yoshimoto et al., 2014). After a number of peroxisomes have undergone pexophagy (if damaged or redundant), new peroxisomes are needed.

1.2 Formation and maintenance of peroxisomes

The latest research suggests that new peroxisomes arise from a combination of *de novo* formation from the endoplasmic reticulum (ER), and growth and division of, and protein import to, pre-existing peroxisomes (reviewed in Agrawal and Subramani, 2016). Proteins that are required to function within or at the membrane of the peroxisome reach their destination through either direct peroxisomal targeting or trafficking in vesicles from the ER (reviewed in Kim and Hettema, 2015, and Mayerhofer, 2016). In plant peroxisomes, some peroxisomal membrane proteins have been found to traffic to peroxisomes *via* the ER prior to their differentiation into specialised peroxisomes (reviewed in Baker and Paudyal, 2014) (**Figure 2**).

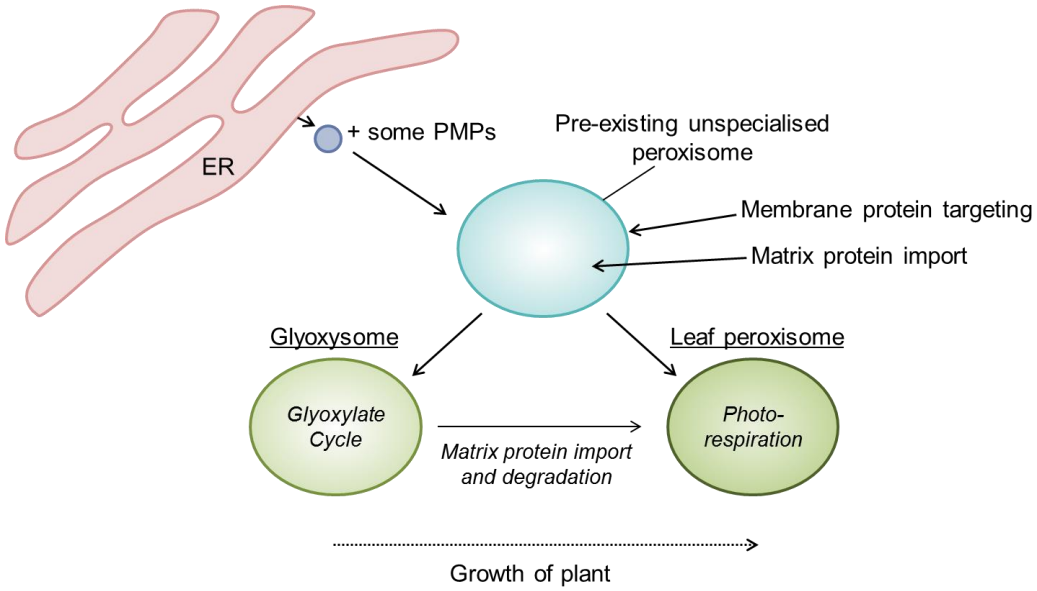


Figure 2| A schematic of the formation and differentiation of plant peroxisomes. Vesicles containing some peroxisomal membrane proteins (PMPs) bud from the ER, joining pre-existing peroxisomes. Following establishment of PMPs in the membrane of the peroxisome, targeting of peroxisomal matrix proteins can begin. Import of specific peroxisomal matrix proteins allows efficient differentiation of the peroxisome, and differentiation of one type of peroxisome to another occurs through a combination of matrix protein import and degradation (Cross et al., 2016). Copyright clearance license obtained from Elsevier; license number 3881361447160. © Elsevier.

Proteins that function in the formation and/or maintenance of peroxisomes are called peroxins. There are a range of peroxins involved in the formation and maintenance of

peroxisomes in the plant cell, and these can be categorised by their functions (**Table 1**).

Function	Name	First cloned from plants
Membrane protein targeting	PEX3	Hunt and Trelease, 2004
	PEX16	Lin et al., 1999
	PEX19	Hadden et al., 2006
Division and proliferation of peroxisomes	PEX11	Lingard and Trelease, 2006
Matrix protein import: PTS1-protein receptor	PEX5	Brickner et al., 1998; Kragler et al., 1998; Wimmer et al., 1998
Matrix protein import: PTS2-protein receptor	PEX7	Li et al., 2003
Matrix protein import: PTS1-/PTS2-protein receptor docking at the peroxisomal membrane	PEX13	Mano et al., 2006
	PEX14	Lopez-Huertas et al., 1999; Hayashi et al., 2000b
Matrix protein import: RING-finger (E3 ligase) complex	PEX2	Hu et al., 2002
	PEX10	Baker et al., 2000; Schumann et al., 2003
	PEX12	Fan et al., 2005
Matrix protein import: E2 ligase	PEX4	Zolman et al., 2005
Matrix protein import: Membrane anchoring of E2 ligase	PEX22	Zolman et al., 2005
Matrix protein import: AAA ATPase complex	PEX1	Lopez-Huertas et al., 2000
	PEX6	Kaplan et al., 2001
Matrix protein import: Membrane anchoring of AAA ATPase complex	APEM9	Goto et al., 2011

Table 1| Roles of peroxins in plants. Peroxins are required for both the formation and maintenance of peroxisomes. PTS, peroxisomal targeting signal; RING, really interesting new gene; AAA, ATPases associated with various cellular activities; PEX, peroxin; APEM, aberrant peroxisome morphology.

Peroxin (PEX) proteins are involved in various stages of peroxisome formation and maintenance. These stages are shown in **Figure 2**. One of these stages, protein import into the matrix of the peroxisome, is controlled by PEX5 and PEX7 and this stage is the focus of this study. Targeting of peroxisomal membrane proteins to the

peroxisome membrane is a crucial process as this establishes the composition and, therefore, role of the peroxisome membrane. When peroxisomal membrane proteins are in place, the peroxisome can carry out the import of matrix proteins (the peroxisomal matrix protein import pathway depends on peroxisomal membrane proteins for its function).

The following sections will cover peroxisomal matrix protein import, as this is the pathway which controls the protein composition and, therefore, function of the peroxisome. In order to work towards producing a synthetic organelle from the peroxisome, matrix protein import seems the obvious pathway to target and manipulate.

1.3 The peroxisomal matrix protein import pathway

Import of cargo proteins to the matrix of the peroxisome from the cytosol is dependent on two pathways, the peroxisomal targeting signal 1 (PTS1)-mediated pathway and the PTS2-mediated pathway. Although these two pathways begin with distinct receptor-targeting signal binding events, there is co-dependence between the pathways and the separation between the pathways becomes more ambiguous at all processes downstream from, and including, docking at the peroxisomal membrane (Nito et al., 2002; Woodward and Bartel, 2005; Ramón and Bartel, 2010). PTSs are recognition sequences at the C- (PTS1) or N- (PTS2) terminus of cargo proteins, and these sequences are recognised by specific receptor proteins (**Figure 3**) in the cytosol.

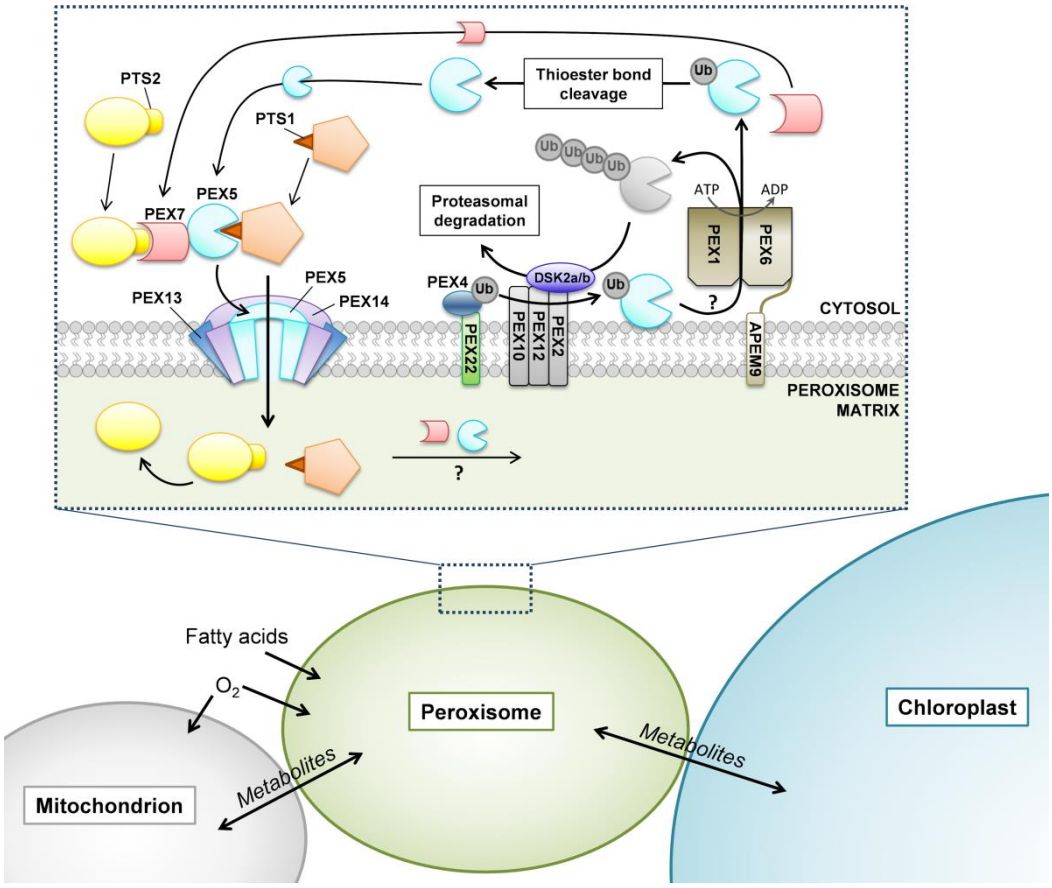


Figure 3| Scheme for the import of proteins into the matrix of the plant peroxisome. Peroxisomal cargo proteins contain either a PTS1 (C-terminal) or a PTS2 (N-terminal) sequence. These are bound by PEX5 or PEX7, respectively, which allows the targeting of cargo proteins to the peroxisomal membrane for import. PTS2 is cleaved from PTS2-cargo proteins within the peroxisomal matrix. After cargo protein import, PEX5 and possibly PEX7 are recycled to the cytosol for another round of import. Lysine polyubiquitination of PEX5 occurs when there is accumulation of PEX5 at the peroxisome membrane. This results in the proteasomal degradation of PEX5 in a pathway that could involve DSK2a and DSK2b (PEX2/PEX12-binding proteins). PTS, peroxisomal targeting signal; PEX, peroxin; Ub, ubiquitin; APEM, aberrant peroxisome morphology; DSK, ubiquitin domain-containing protein. (Cross et al., 2016). Copyright clearance license obtained from Elsevier; license number 3881361447160. © Elsevier.

The peroxisomal matrix protein import cycle begins with the PTS1/PTS2-cargo protein binding its cognate receptor, followed by docking of this complex at peroxisomal matrix proteins (PMPs). Translocation of folded cargo is achieved though interaction of peroxin 5 (PEX5)/peroxin 7 (PEX7) with peroxisomal membrane proteins. Cargo is released once in the peroxisomal matrix and PEX5/PEX7 is recycled to the cytosol for another round of import.

1.3.1 Cargo recognition by the PTS receptors

PEX5 and PEX7 are responsible for the recognition of PTSs on proteins destined for the peroxisome (**Figure 4**). PEX5 binds to the PTS1 sequence on PTS1-cargo proteins, and PEX7 binds to the PTS2 sequence on PTS2-cargo proteins. In the plant and mammalian systems, the co-receptor for PEX7 is PEX5 (Nito et al., 2002; Woodward and Bartel, 2005; Braverman et al., 1998); however, in *S. cerevisiae* the co-receptors are PEX18 and PEX21 (Purdue et al., 1998), and in other yeasts the co-receptor is PEX20 (Titorenko et al., 1998).

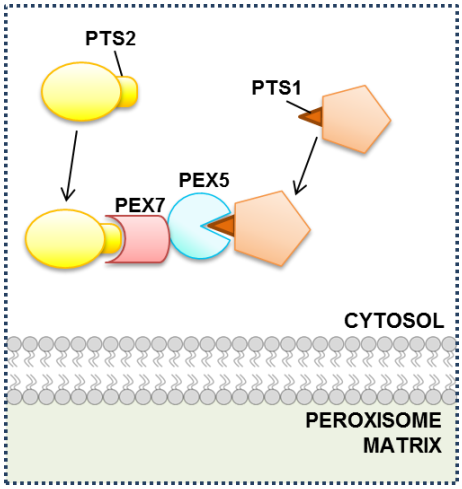


Figure 4| Cargo recognition in the plant peroxisomal matrix protein import pathway. (Cross et al., 2016). Copyright clearance license obtained from Elsevier; license number 3881361447160. © Elsevier.

The PTS1-mediated import pathway is the most common pathway for cargo proteins to enter the plant peroxisome. PEX5 is also a co-receptor for PEX7 in the PTS2 import pathway: this has been found through genetic studies, showing that mutation of a site within PEX5 of Chinese hamster ovary cells (Ser214Phe) disrupted PTS2 import but had no apparent effect on PTS1 import (Matsumura et al., 2000). Mutation of the equivalent residue in *Arabidopsis* PEX5 (Ser318Leu) also had the same effect of disrupting only PTS2 import (Zolman et al., 2000; Woodward and Bartel, 2005). PTS2 import with this *A. thaliana* protein variant was recovered by the simultaneous expression of a truncated PEX5 variant comprising the N-terminal domain, $PEX5_{454}$, which implies that PEX7 relies on the N-terminal domain of PEX5 for PTS2 import. In rice and humans, there are two splice variants of PEX5 which result in two possible isoforms of the protein: PEX5S and PEX5L (Lee et al., 2006; Braverman et al., 1998). PEX5L contains the PEX7 binding site, so can act as a co-receptor in the PTS2 pathway, whereas PEX5S does not have the ability to bind to PEX7 so can only play a

role in the PTS1 import pathway. In *Arabidopsis thaliana*, only the equivalent of the longer splice variant of PEX5 is expressed (Hayashi et al., 2005; Woodward and Bartel, 2005).

1.3.1.1 PTS1-cargo recognition

The most abundant PTS1 sequences display the properties [small]-[basic]-[hydrophobic]-C terminus. These have been generally accepted as the requirements for PTS1-mediated peroxisomal targeting; however, it is becoming clear that this C-terminal sequence can be quite varied, and can rely heavily on its upstream sequence to change the targeting abilities of the PTS1-cargo protein (Mullen et al., 1997; Maynard et al., 2004; Brocard and Hartig, 2006; Reumann, 2011; Chowdhary et al., 2012). A PTS1 prediction algorithm (PTS1 Predictor) was created by Neuberger and colleagues for the prediction of PTS1-containing proteins in largely metazoan and fungal species (Neuberger et al., 2003a; Neuberger et al., 2003b). The bioinformatics tool PredPlantPTS1 allows prediction of the probability of specifically a plant C-terminal sequence allowing PTS1-mediated targeting to peroxisomes (Lingner et al., 2011; Reumann et al., 2012). This has been validated by *in vitro* binding studies and *in vivo* import experiments (Skoulding et al., 2015).

The C-terminal domain of PEX5 is responsible for PTS1 binding. This has been shown by the expression of only the C-terminal domain of *Arabidopsis* PEX5. An *in vitro* binding study with this N-terminally truncated variant of PEX5 has demonstrated that binding of a PTS1 peptide is effectively unchanged when comparing the full-length protein to the TPR domain alone (Skoulding, 2011).

1.3.1.2 PTS2-cargo recognition

The consensus for a PTS2 sequence is [R/K]-[L/V/I/Q]-X_2-[L/V/I/H/Q]-[L/S/G/A/K]-X-[H/Q]-[L/A/F] and this sequence is found near the N-terminus of PTS2-cargo proteins (Petriv et al., 2004). To date, no PTS2 prediction tool has been published.

Genetic studies of PEX7 have shown that reduced PTS2 import is observed when expression of PEX7 is knocked down (Hayashi et al., 2005; Woodward and Bartel, 2005; Ramón and Bartel, 2010). Interestingly, a missense mutation in PEX7 (Thr124Ile) results in reduced PTS1 and PTS2 import, which suggests that PEX5 can

also rely on PEX7 for import, possibly suggesting some co-dependence (Ramón and Bartel, 2010).

1.3.1.3 PTS-mediated import of folded proteins

Peroxisomes are unusual in that they facilitate the import of fully folded proteins (Walton et al., 1995). The PTS1 pathway in yeast has been shown to accommodate large complexes through the membrane by forming a pore that can reach up to 9.25 nm in diameter (Meinecke et al., 2010). A phenomenon known as 'piggyback import' also occurs in the PTS1 and PTS2 import pathways. This is where a protein without a PTS can associate with a PTS-cargo protein in order for both proteins to gain access to the peroxisomal matrix (McNew and Goodman, 1994; Glover et al., 1994; Kato et al., 1999). Some of the subunits of the enzyme protein phosphatase 2A (PP2A), which seems to be involved in β-oxidation in peroxisomes, have been found to target to the peroxisome by piggyback import. In this case, only one of the subunits of PP2A possesses a PTS1, [S]-[S]-[L]-C terminus, and this sequence targets all associated subunits of PP2A to the peroxisome (Kataya et al., 2015).

1.3.2 The peroxisomal membrane docking complex

Cargo-loaded peroxisomal targeting receptors dock to proteins at the membrane of the peroxisome for delivery of the cargo to the matrix of the peroxisome. PEX14 and PEX13 are the major known docking proteins (**Figure 5**).

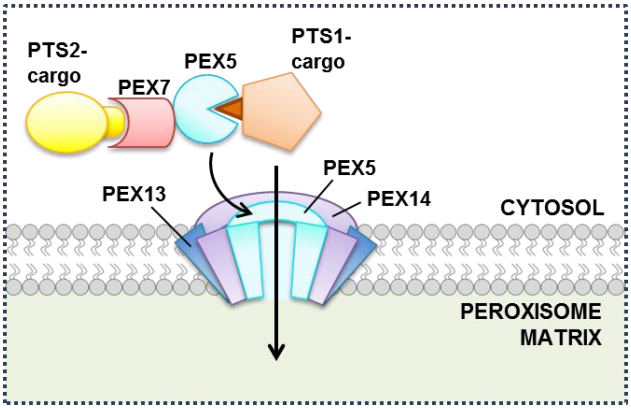


Figure 5| Peroxisome membrane docking in the plant peroxisomal matrix protein import pathway. (Cross et al., 2016). Copyright clearance license obtained from Elsevier; license number 3881361447160. © Elsevier.

It has come to light that PEX14, although playing a major role in facilitating peroxisomal protein import in some organisms, is not essential for the process in

Arabidopsis (Hayashi et al., 2000b; Monroe-Augustus et al., 2011; Burkhart et al., 2013) or the yeast *Hansenula polymorpha* (Salomons et al., 2000). *Arabidopsis* PEX14 mutants, which reduced PEX14 protein expression levels, resulted in reduced PTS1 and PTS2 import (Monroe-Augustus et al., 2011; Burkhart et al., 2013). This suggests that PEX14 is still important for *Arabidopsis* matrix protein import into the peroxisome, even if it is not essential for this process. A dynamic transient pore between PEX5 and PEX14 has been demonstrated to form in the *Saccharomyces cerevisiae* system, which is presumed to allow cargo to enter the matrix of the peroxisome (Meinecke et al., 2010). This was initially suspected when size exclusion chromatography and immunoblot analysis revealed a complex the size of PEX5:PEX14 but lacking cargo protein. When this complex was reconstituted into liposomes and incubated with purified PEX5-cargo, a dynamic pore could be measured by monitoring ion channel activity (Meinecke et al., 2010). Recent work by Meinecke and colleagues suggests that this is a water-filled pore, which expands depending on the size and oligomeric state of cargo (Meinecke et al., 2016). In the absence of PEX14, it could be postulated that PEX5 can form PEX5-only pores by self-oligomerisation as low levels of PTS1-cargo import are still observed in the absence of PEX14 in both *Arabidopsis* and yeast (Monroe-Augustus et al., 2011; Salomons et al., 2000). Also, Kerssen and colleagues showed that *Saccharomyces cerevisiae* PEX5 containing two point mutations, which prevented its association with PEX13 and PEX14, still associated with the peroxisome membrane (Kerssen et al., 2006).

PEX13 is also part of the docking complex for cargo-loaded receptors, and an interaction between PEX13 and PEX14 has been reported in yeast and mammals (Albertini et al., 1997; Fransen et al., 1998). In *Arabidopsis*, a point mutation within PEX13 (Glu243Lys) results in reduced PTS1 and PTS2 import (Woodward et al., 2014), and the site of this mutation sits within the putative Src homology 3 (SH3) domain of PEX13, which has been determined as the binding site for PEX14 in yeast and mammals (Albertini et al., 1997; Fransen et al., 1998). Other genetic studies in *Arabidopsis*, in which PEX13 is truncated or its expression is knocked down, have also resulted in reduced PTS1 and PTS2 import (Mano et al., 2006; Nito et al., 2007). Studies of PEX13 have also given insight into the overall import process. For example, knockdown of PEX13 expression levels exacerbates the effects of mutation in ‘early acting’ peroxins PEX5 and PEX14, resulting in further impaired peroxisomal protein import, yet restores the detrimental effects of mutation in ‘late acting’ peroxins PEX1 and PEX6 (involved in receptor recycling), helping to restore peroxisomal protein

import (Ratzel et al., 2011). This suggests that the 'early acting' peroxins and the 'late acting' peroxins are linked in overall function but not in mechanism (Ratzel et al., 2011).

Interactions between the *Arabidopsis* import receptors and the docking peroxins have been studied *in vitro* using yeast-two-hybrid studies, filter binding and pull-down analysis (Nito et al., 2002; Mano et al., 2006; Lanyon-Hogg et al., 2014). In an *Arabidopsis* study by Nito and colleagues, no binding of PEX14 to PEX7 was observed, unlike the mammalian and yeast import systems (Nito et al., 2002). Yeast-two-hybrid was also used to investigate binding of *Arabidopsis* PEX13 to PEX5 and PEX7, and it was shown that PEX13 interacts with PEX7 but not PEX5 (Mano et al., 2006). In the mammalian import system, binding of PEX5 to PEX14 caused the release of the PTS1 cargo catalase (Freitas et al., 2011); however, recent work in *Arabidopsis* has shown that the interaction between PEX5 and a PTS1-representative peptide is unchanged when increasing concentrations of PEX14 are titrated into the mixture (Lanyon-Hogg et al., 2014). It has also been demonstrated by Lanyon-Hogg and colleagues that PTS1-PEX5-PEX7-PTS2 interact by pull-down from the cytosolic fraction of Arabidopsis, but with *At*PEX14 only PTS1-PEX5-PEX7 are pulled down, which suggests a potential role for PEX14 in PTS2 cargo unloading (Lanyon-Hogg et al., 2014).

The PTS receptor(s) must be recycled to the cytosol for subsequent rounds of cargo import. Ubiquitin is an important part of the recycling process of the PTS receptor(s), as ubiquitin must be transferred onto PEX5, and possibly PEX7, to allow for receptor recycling (Platta et al., 2016).

1.3.3 The mechanism for ubiquitination of the PTS receptors

In order for multiple rounds of cargo import into peroxisomes to be carried out, the PEX5 receptor must be recycled to the cytosol. This process is governed by monoubiquitination of a cysteine residue (in yeast (Williams et al., 2007) and mammals (Carvalho et al., 2007a)) which also appears to be conserved in plant PEX5. There is also the option for PEX5 degradation by polyubiquitination of one or more lysine residues of PEX5, which directs PEX5 through the RADAR (Receptor Accumulation and Degradation in the Absence of Recycling) pathway if there is inefficient export of PEX5, and therefore a build-up of PEX5 at the peroxisome membrane (Léon et al., 2006) (**Figure 6**). Ubiquitination requires an E1 ubiquitin-activating enzyme, which

transfers ubiquitin to an E2 ubiquitin-conjugating enzyme. This allows the transfer of ubiquitin to the target protein in the presence of an E3 ligase (Deshaies and Joazeiro, 2009).

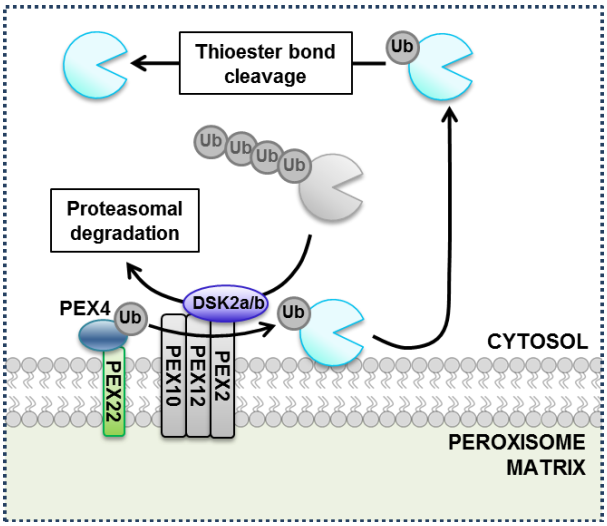


Figure 6| Ubiquitination of the PTS1 receptor in the plant peroxisomal matrix protein import pathway. (Cross et al., 2016). Copyright clearance license obtained from Elsevier; license number 3881361447160. © Elsevier.

PEX4 has been identified in yeast as the E2 ubiquitin-conjugating enzyme which comes into close proximity to PEX12 of the PEX2/10/12 RING-finger (E3 ligase) complex to allow the catalysis of ubiquitin transfer from PEX4 to PEX5 (Platta et al., 2007; Platta et al., 2009) (**Figure 6**). PEX4 is anchored to the membrane by PEX22 (Koller et al., 1999; Zolman et al., 2005). Mammals do not possess PEX4, so rely on cytosolic E2 enzymes (reviewed in Francisco et al., 2014). The ubiquitination machinery for receptor recycling in plants has not been extensively studied; however, it has been found that the *Arabidopsis* PEX4-PEX22 complex can restore function of yeast peroxisomes in which PEX4 and PEX22 are deficient (Zolman et al., 2005). In yeasts, the function of PEX4 is enhanced by the cytosolic domain of PEX22 (Platta et al., 2007; Williams et al., 2007; El Magraoui et al., 2014). Knockdown of PEX22 expression in *Arabidopsis* by T-DNA insertion enhances effects caused by PEX4 knockdown (Zolman et al., 2005), suggesting that the two proteins are connected in mechanism. As *Arabidopsis* and yeast PEX4-PEX22 seem to be interchangeable, and *Arabidopsis* PEX22 has also been found to be an enhancer of PEX4 (Zolman et al., 2005), it seems likely that that the plant ubiquitin machinery for recycling or degradation of PEX5 works in a similar way to the yeast system.

PEX2, PEX10 and PEX12 form the RING complex, so named because each of these peroxins contains at least a partial RING-finger domain. PEX2 and PEX10 contain full RING-finger domains, whereas PEX12 contains a partial RING-finger domain. The RING-finger domains of the three proteins from *Arabidopsis* have been isolated and studied to elucidate function, and it has been found that all three possess E3 ligase activity (Kaur et al., 2013). The RING complex is vital for peroxisome function and therefore cell survival, which has been shown through genetic studies. Mutants of the three proteins (*pex2-T-DNA*; *pex10-1*; *pex12-T-DNA*), in which T-DNA insertion results in frameshift, are all lethal to the plant (Hu et al., 2002; Schumann et al., 2003; Sparkes et al., 2003; Fan et al., 2005; Nito et al., 2007). Truncation of PEX10 (*pex10-W313**) also results in a lethal phenotype (Prestele et al., 2010). Knockdown of expression of the three RING peroxins by RNA interference (*pex2i*, *pex10i* and *pex12i*) all result in reduced PTS1 and PTS2 import (Nito et al., 2007), potentially due to the reduced presence of PEX5 in the cytosol as less PEX5 is ubiquitinated and available for export to the cytosol.

In yeast, PEX12 is the E3 ligase involved in PEX5 monoubiquitination and PEX2 is the E3 ligase involved in PEX5 polyubiquitination (Platta et al., 2009), although it remains to be seen if this is the case in plants. A quality control mechanism also appears to be in place for PEX7, as the GTPase rabE1c has been found to be involved in PEX7 degradation in *Arabidopsis* (Cui et al., 2013), and PEX7 degradation that was reliant on PEX20 and PEX5 was reported in yeast (Hagstrom et al., 2014). To meet either fate of recycling or degradation, PEX5 must be exported from the peroxisome membrane. For receptor recycling, a membrane-anchored AAA ATPase (ATPases associated with various cellular activities) complex is required (reviewed in Grimm et al., 2016).

1.3.4 The receptor recycling complex

Three peroxins, PEX1, PEX6 and APEM9 (aberrant peroxisome morphology-9) are required for the PTS receptor(s) to be recycled back to the cytosol for subsequent rounds of peroxisomal import (**Figure 7**). The AAA ATPase complex is comprised of PEX1 and PEX6. PEX1 and PEX6 are anchored to the cytosolic side of the membrane by APEM9 (through the association of APEM9 with PEX6) (Goto et al., 2011), and this complex seems to be important for both the PTS1 and PTS2 import cycles in *Arabidopsis*. This has been found through genetic studies of the three peroxins. When expression of PEX1, PEX6 or APEM9 is knocked down separately (*pex1i*, *pex6i*,

apem9i), the result is reduced PTS1 and PTS2 import (Nito et al., 2007; Goto et al., 2011). This suggests a link between receptor recycling and cargo protein import.

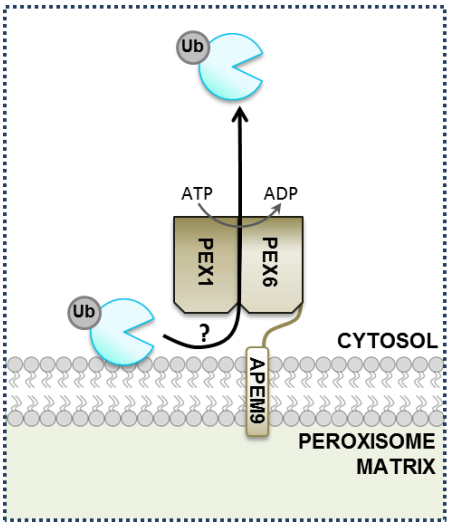


Figure 7| Receptor recycling in the plant peroxisomal matrix protein import pathway. (Cross et al., 2016). Copyright clearance license obtained from Elsevier; license number 3881361447160. © Elsevier.

Disruption of APEM9, by T-DNA insertion at either the N- or C-terminal portion of the protein causing frameshift, results in a lethal phenotype (Goto et al., 2011). This highlights the importance of receptor recycling for the peroxisomal matrix protein import process. In APEM9, a missense mutation (*apem9-1*; Gly278Glu) results in reduced PTS1 and PTS2 import (Goto et al., 2011; Li et al., 2014). In PEX6, a mutation in the C-terminal portion of the protein (*pex6-1*; Arg766Gln) reduces PEX5 levels, whereas a mutation in the N-terminal portion (*pex6-2*; Leu328Phe) does not seem to affect PEX5 levels (Zolman and Bartel, 2004; Burkhart et al., 2013). This could be due to the location of the mutation, or it could be that the change in chemical properties upon the Arg→Gln mutation in *pex6-1* results in a more defective phenotype than the more conservative Leu→Phe mutation.

The role of PEX1 and PEX6 is to export PEX5 from the peroxisome membrane back into the cytosol, and this has been demonstrated in mammalian and yeast systems (Miyata et al., 2012; Platta et al., 2005). It is thought that conformational changes occur in the AAA ATPases in response to ATP consumption, which could allow PEX5 to be pulled out of the peroxisome membrane (Platta and Erdmann, 2007). A co-factor of PEX6, AWP1 (associated with PRK1), is an important protein for mammalian PEX5 export. AWP1 has been found to interact with mammalian PEX5, with a higher affinity

to monoubiquitinated PEX5 (Miyata et al., 2012). This suggests that AWP1 could be involved in pulling PEX5 from the peroxisome membrane and into the cytosol. A more recent study in *Saccharomyces cerevisiae* shows that the position of the monoubiquitination site, near the N-terminus of Pex5p, is important for a functional PEX5 receptor, and the nature of the cysteine residue itself is important for efficient PEX5 recycling (Schwartzkopff et al., 2015). In this study, Schwartzkopff and colleagues showed that mutation of the cysteine at the monoubiquitination site of PEX5 to alanine ($PEX5_{C6A}$) resulted in a non-functional PTS1-import receptor; mutation of the cysteine to lysine ($PEX5_{C6K}$) resulted in polyubiquitination of $PEX5_{C6K}$, so this receptor was still functional but much less efficient at recycling, as proteasomal degradation took place much more readily (Schwartzkopff et al., 2015). Receptor recycling is an important part of the import pathway as this regenerates PEX5 to the cytosol where it can bind another PTS1-containing cargo protein.

Recent electron micrographs of the PEX1/6 AAA ATPase complex from yeast have revealed that this complex is a hexamer comprising a trimer of dimers: each dimer consisting of a monomer of PEX1 and a monomer of PEX6 (Ciniawsky et al., 2015). This work has shown that rotational movement of the complex in response to ATP appears to be responsible for the export of PEX5 from the membrane, and that this export may require partial or complete unfolding of PEX5 (Ciniawsky et al., 2015). The receptor recycling stage of plant peroxisomal matrix protein import still requires unravelling and it will be exciting to see how this process works in plants, and how this differs from mammals and fungi.

1.3.5 Export-driven import

Recent publications are suggesting that there is a link between import of PTS1-cargo and export of the PEX5 protein back to the cytosol. A link between these two processes was first suggested when parallels were drawn between peroxisomal protein import and the endoplasmic reticulum-associated degradation (ERAD) pathway, with similarities between the exit of misfolded proteins from the ER membrane and the exit of PEX5 from the peroxisome membrane being remarked upon (Schliebs et al., 2010). Models of an export-driven import process have been produced for human peroxisomal protein import (Brown et al., 2014), and PEX5 has also recently been found halted at the peroxisomal membrane in a monoubiquitinated state when a bulky tag (representing PTS1-cargo that cannot be released) was added to the C-terminus of mammalian PEX5 (Nordgren et al., 2015). A covalent label-transfer

system, for studying *in vitro* interactions important for peroxisomal import, was set up recently by the Baker laboratory (Bhogal et al., 2016), which showed that the interaction between PEX5 (the PTS1 receptor) and PEX14 (part of the membrane docking complex for PTS receptors) in *A. thaliana* was independent of cargo, yet dependent on ATP. This could mean that the export of PEX5 from the peroxisomal membrane (an ATP-dependent process) is dependent on the actions of the docking and translocation machinery (DTM).

It has been suggested, in the mammalian import system, that the RING complex (section 1.3.3) acts as part of the importomer and that export of the PEX5 protein from the membrane could be linked to the cargo unloading process (Brown et al., 2014). In yeast, two stable complexes were isolated, one containing Pex14p, Pex17p (a component of the translocation import machinery in yeast) and a small amount of Pex13p, and the other containing the three RING finger peroxins Pex2p, Pex10p and Pex12p (Agne et al., 2003). Interestingly, these two complexes were associated into a larger complex in the presence of Pex8p, a crucial peroxin for peroxisomal protein import in yeast (Agne et al., 2003). Possible links between the cargo import components and the receptor export components have also been found in plants, as it was found that levels of PEX12 are enhanced when PEX13 is truncated, suggesting a compensatory mechanism (Mano et al., 2006), and also APEM9 has been found to interact with PEX13 (Li et al., 2014).

Three models for the dynamics of the PEX5 protein in cargo delivery and export into the cytosol have been proposed for the mammalian PTS1 import system (Brown et al., 2014): 1) PEX5 could be pulled from the membrane by PEX1/6 in a process twinned with cargo unloading; 2) the two processes, PEX5 being pulled from the membrane and cargo unloading, could be separate; 3) the two processes are cooperatively coupled – as a monoubiquitinated PEX5 is extracted from the membrane, this allows a cargo-loaded PEX5 protein to release its cargo into the peroxisomal matrix. PEX5:PEX14 have been found in a 1:5 ratio under natural conditions (Gouveia et al., 2000), and a 1:1 ratio when export of PEX5 is blocked (Meinecke et al., 2010). This suggests a build-up of PEX5 at the DTM when cargo-unloaded PEX5 cannot be exported and supports the cooperatively coupled model (Brown et al., 2014).

1.3.6 A focus on the PTS1-mediated import pathway

Subsequent sections of this introduction will focus on the PTS1 protein import pathway, for the following reasons: 80% of matrix proteins contain a C-terminal PTS1 and are imported into the peroxisome *via* the PTS1-mediated import pathway; in some organisms, for example the diatom *Phaeodactylum tricornutum*, the PTS2 pathway seems to be completely lost and the PTS1 pathway is responsible for all protein import into peroxisomes (Gonzalez et al., 2011); and the PTS1-mediated import pathway has been studied more extensively than the PTS2-mediated import pathway. More available data addressing the PTS1-mediated import pathway will allow for an informed attempt at the manipulation of this pathway in this study.

1.4 PEX5 structure and function

PEX5 is a modular protein, with the C-terminal domain crucial for PTS1 binding and the N-terminal domain involved in interaction with PEX7, possibly translocation, and recycling of PEX5 (**Figure 8**). It was thought that the N-terminus could inhibit peroxisome membrane docking and translocation prior to cargo binding, with Hsp70 needed to produce a PTS1-binding competent state of PEX5 (Harano et al., 2001). Recent research suggests that this is not the case, as fluorescence anisotropy studies have shown that full-length PEX5 and just the C-terminal domain (PEX5-C) both bind to a representative PTS1 sequence YQSKL-CO_2H with very similar affinity (Harper et al., 2003; Skoulding et al., 2015). Also, co-immunoprecipitation of PEX5 with PEX14 has been seen in the absence of PTS1 cargo protein (Lanyon-Hogg, 2012).

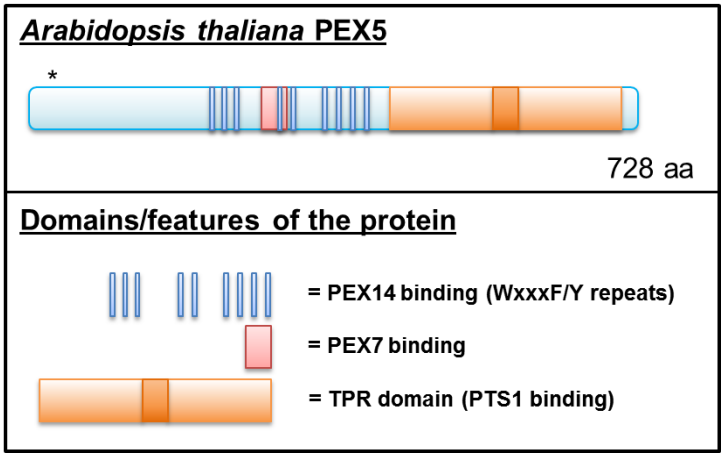


Figure 8| A simplified representation of *At*PEX5. Monoubiquitination (conserved cysteine residue) and polyubiquitination (conserved lysine residues) sites are located near the N-terminus of *At*PEX5 and this region is highlighted by an asterisk. (Cross et al., 2016). Copyright clearance license obtained from Elsevier; license number 3881361447160. © Elsevier.

PEX5 contains nine WxxxF/Y (Trp-x-x-x-Phe/Tyr) pentapeptide repeats (**Figure 8**) which were found to be crucial in the interaction of PEX5 with PEX14 (Schliebs et al., 1999). These repeats bind to the N-terminal portion of PEX14, as demonstrated by: size exclusion chromatography of *H. sapiens* PEX5 with and without an N-terminal fragment of PEX14 (Schliebs et al., 1999); competition analysis of *H. sapiens* PEX5 versus the peptide WAQEF for binding to the 78 N-terminal residues of PEX14 (Saidowsky et al., 2001); pull-down of PEX5 with progressive N-terminal deletions of PEX14 (Madrid and Jardim, 2005); *in vivo* analysis of PTS1 cargo import with site-directed mutants of PEX14 (Su et al., 2009); and structural determination of the PEX14 interacting with peptides from the PEX5 sequence, by nuclear magnetic resonance (Neufeld et al., 2009 (PDB file 2W84); Neuhaus et al., 2014 (PDB file 4BXU)).

The PEX7 binding site within PEX5 contains a conserved serine residue (S213 in *Homo sapiens* PEX5 (*Hs*PEX5), S318 in *Arabidopsis thaliana* (*At*PEX5)) and is between sets of the WxxxF/Y repeats that associate PEX5 with PEX14 (Schliebs et al., 1999; Woodward and Bartel, 2005). Binding of PEX5 to the PTS1 sequence occurs by interaction of PTS1 with the TPR (tetratricopeptide repeat) domain.

1.4.1 The N-terminal domain of PEX5

The N-terminal domain of PEX5 is largely unfolded, as suggested by Carvalho and colleagues and confirmed by small angle X-ray scattering of PEX5 which revealed an

extended N-terminal domain (Carvalho et al., 2006; Shiozawa et al., 2009). Insertion of PEX5 into the membrane *via* the N-terminus, as hypothesised in the transient pore model, was thought to be cargo-protein dependent (Gouveia et al., 2003). It has, however, been found that a mutation in PEX5 (N526K in *Hs*PEX5L), which diminishes PTS1 binding by PEX5, results in a PEX5 variant able to insert into the peroxisome membrane in an *in vitro* import system (Carvalho et al., 2007b). This shows that PEX5 can have the ability to insert into the membrane in the absence of bound cargo protein. This was also supported by the finding that PEX5 could bind to PEX14 without prior binding to PTS1 cargo (Lanyon-Hogg et al., 2014).

1.4.2 The C-terminal domain of PEX5, and PTS1 binding

As stated, the C-terminal domain of PEX5 binds cargo proteins *via* their PTS1 sequence. The disordered nature of the N-terminal domain of PEX5 means that it has only been possible to crystallise the C-terminal domain (PEX5-C). The structure of PEX5-C:PTS1 reveals that PEX5-C is composed mainly of two sets of three tetratricopeptide repeats (TPRs), comprising TPRs 1-3 and TPRs 5-7. These two binding faces make up the PTS1-binding site (with TPR4 acting as a linker), which contains conserved asparagine residues that can make polar interactions with the backbone of the PTS1 sequence. Many other conserved amino acids in the PEX5 binding site are responsible for side chain interactions of the PTS1 sequence, either directly or indirectly *via* water molecules (Gatto et al., 2000; Stanley et al., 2006; Sampathkumar et al., 2008; Fodor et al., 2012; Fodor et al., 2015).

Each TPR usually consists of two antiparallel α-helices (D'Andrea and Regan, 2003) in a helix-turn-helix motif, but TPR3 was found in an X-ray crystal structure to be present as one elongated helix (Kumar et al., 2001 (PDB file 1HXI)). Each TPR is a 34-amino acid sequence which can act as an 'interaction scaffold' (Allan and Ratajczak, 2011). When many copies of consensus TPRs are synthesised in one chain, they appear to group together in sets of 3 TPRs to form binding sites with the overall structure resembling a spiral (D'Andrea and Regan, 2003). The TPRs of PEX5 are unusual in that TPRs 1-3 and TPRs 5-7 group together to form an overall binding site for interacting proteins. This has been confirmed through mutation of specific residues in both sets of the 3 TPRs of PEX5, showing that both sets of TPRs play a role in PTS1-binding. **Figure 9**, an I-TASSER-predicted structural model of the C-terminal domain of *A. thaliana* PEX5 (*At*PEX5-C) (without the flexible WxxxF/Y repeat region – *At*PEX5(444-728)) with YQSKL in its binding site, demonstrates the way in which

YQSKL is bound to the TPR domain of PEX5. The I-TASSER-predicted model was based on five PDB 'hits' with well-aligning primary sequences (in order of sequence similarity (most → least): PDB files 1FCH (chain A) (Gatto et al., 2000), 3CVN (chain A) (Sampathkumar et al., 2008), 3CV0 (chain A) (Sampathkumar et al., 2008), 4EQF (chain A) (Bankston et al., 2012), and 1W3B (chain A) (Jinek et al., 2004)).

The "7C-loop", which is found downstream from TPR7 (highlighted in **Figure 9**), could be important in cargo binding as a notable change in its position has been observed upon PTS1-binding (Stanley et al., 2006). This is expected to convert PEX5 from a relatively open conformation to a closed, more ring-like conformation, though the function of this movement in PTS1-cargo import is not yet clear (Stanley et al., 2006).

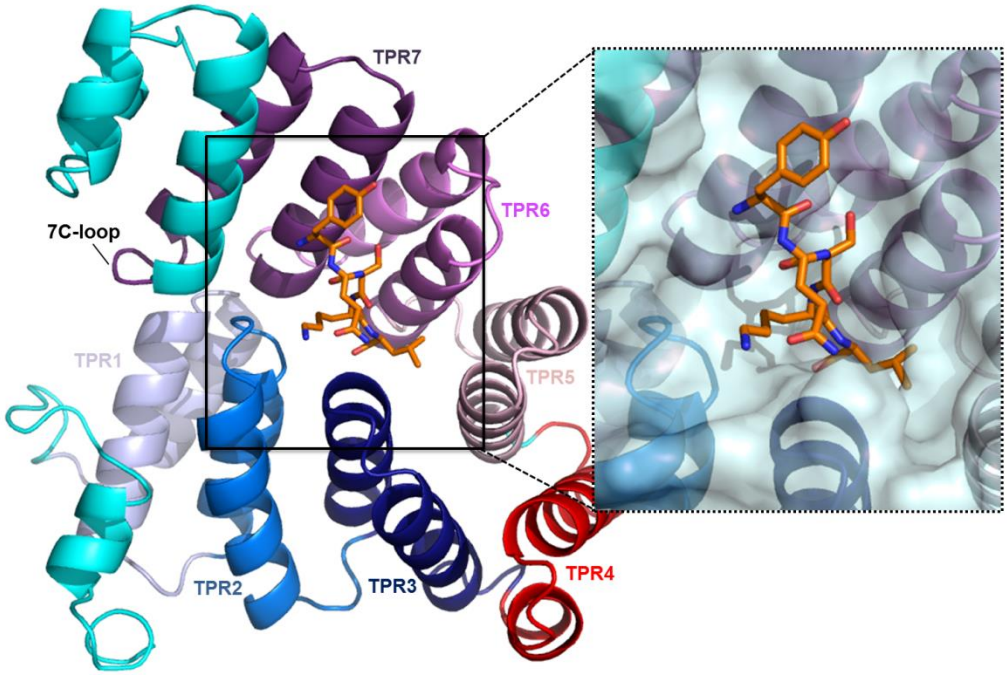


Figure 9| The predicted structural model of the *At*PEX5 TPR domain with YQSKL (from the structure of *Hs*PEX5-C:PTS1 (1FCH)) (consensus PTS1) bound. The PTS1 sequence is recognised between two arch-like groups of three TPRs, TPR1-3 (coloured blue) and TPR5-7 (coloured purple); TPR4 (coloured red) seems to act as a hinge region between the two groups. The structural model of *At*PEX5-C was produced using I-TASSER (Zhang, 2008; Roy et al., 2010; Yang et al., 2015; Yang and Zhang, 2015) (peptide: (Gatto et al., 2000), PDB file 1FCH). TPR, tetratricopeptide repeat.

It has been suggested that TPRs move in a rigid nature in response to PTS1 binding, which seems to transform the TPR domain to a closed, almost circular, conformation (Stanley et al., 2006). The rigid movement of the TPR domain in response to PTS1 binding was first thought to be mediated by TPR4 (Gatto et al., 2000); however,

another research group have found that TPR5 and 6 could mediate the change in structure upon cargo binding (Stanley et al., 2007).

Important TPRs for the binding of PTS1 sequences to *Hs*PEX5 are TPR2, 3, 6 and 7. These TPRs contain residues that have been found to be significant for PTS1 binding. The four conserved asparagine residues of *Hs*PEX5L, N415, N526, N534 and N561 (N537, N628, N636 and N663 in *At*PEX5, respectively), are found in close proximity to PTS1 in all current crystal structures of the cargo-bound *Hs*PEX5 TPR domain (Gatto et al., 2000; Stanley et al., 2006; Fodor et al., 2012). The PTS1 'pocket' of PEX5 contains three key regions: the hydrophobic pocket that commonly associates with leucine at the C-terminal position, another pocket that associates with a small amino acid such as serine, and negative residues that usually associate with lysine or arginine *via* a water molecule. The overall interactions within the PTS1 binding site can be seen in the interaction plot within **Figure 10**.

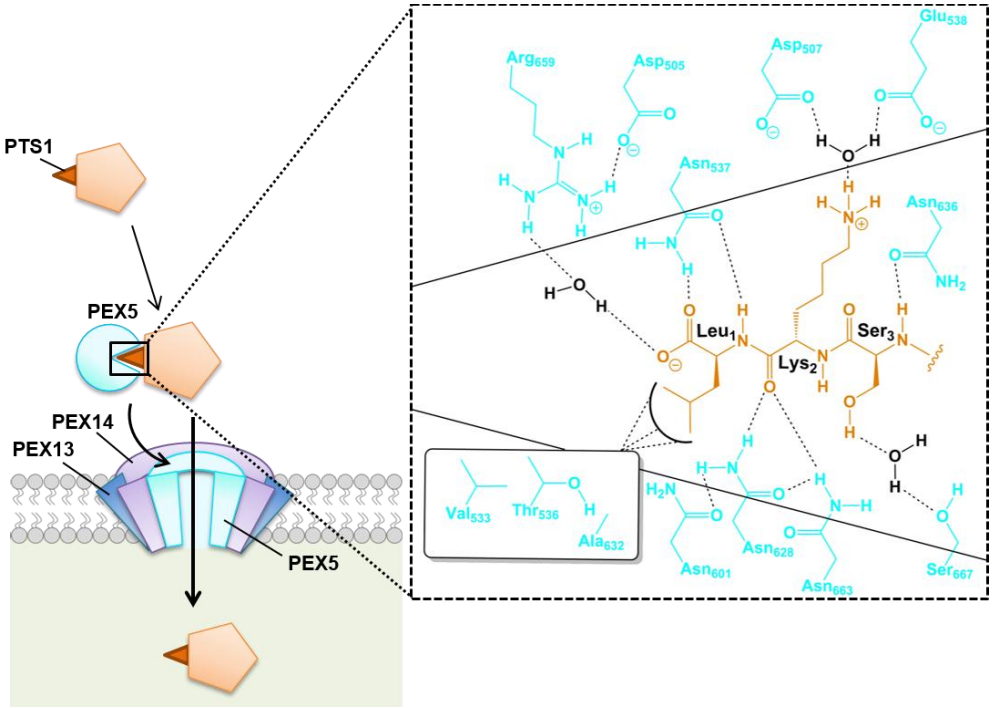


Figure 10| PEX5 recognition of the three C-terminal residues of a PTS1. An interaction plot is shown, using *At*PEX5 numbering, based on existing PEX5-C:PTS1 crystal structures.

Figure 9 and **Figure 10** show that there appear to be specified regions in the PTS1-binding site of PEX5 for interaction with the side chains present in the PTS1. The sequence used to represent a PTS1 in **Figure 10**, SKL-CO_2H, was originally found in the peroxisomal protein luciferase (Gould et al., 1987). This has since been found to

be one of the most common PTS1 sequences, and the general PTS1 consensus was thereafter defined as having the properties: [small]-[basic]-[hydrophobic]-CO_2H (Lametschwandtner et al., 1998).

A range of natural PTS1 sequences are now known, and numerous PEX5:PTS1 crystal structures reveal differences in the PTS1 sequence. *Hs*PEX5S was crystallised with the model PTS1 peptide YQSKL (Gatto et al., 2000). *Hs*PEX5L has been crystallised with sterol carrier protein 2 (mSCP2) as the PTS1 protein, of which the PTS1 is AKL (Stanley et al., 2006). The unusual PTS1 of alanine-glyoxylate aminotransferase (AGT), crystallised with *Hs*PEX5L, is KKL (Fodor et al., 2012). Other PTS1 variations are seen in X-ray crystal structures of *Tb*PEX5 and *Mm*PEX5-like protein. Some examples of these, and *Hs*PEX5 PTS1 binding, are shown in **Figure 11**.

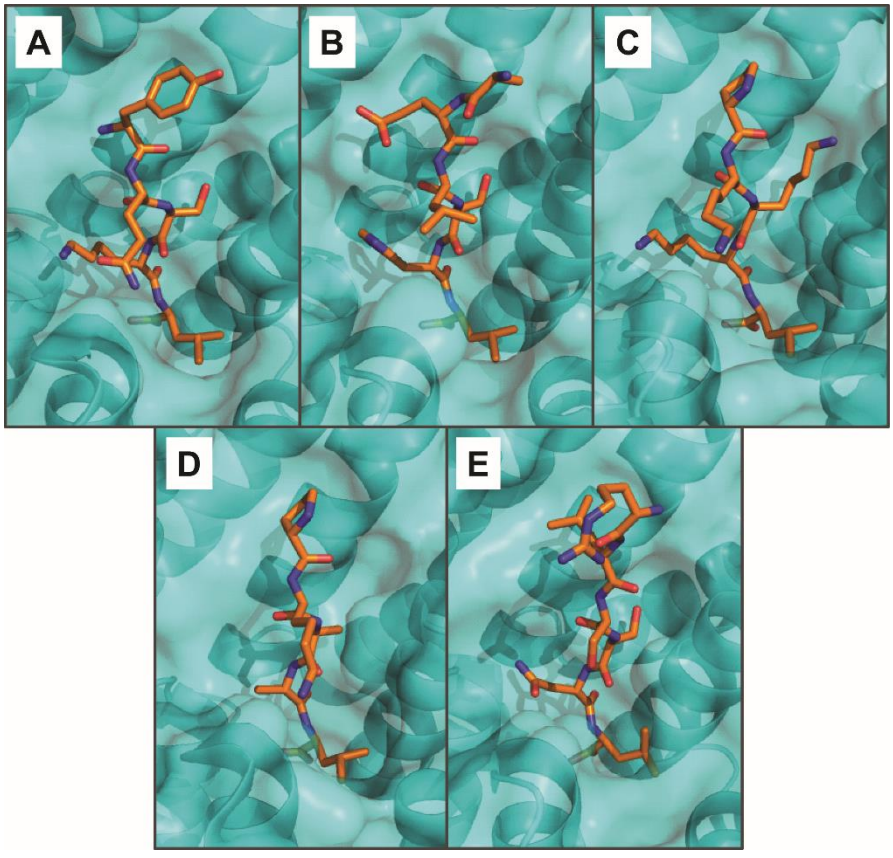


Figure 11| The overall shape of the PTS1-binding site of PEX5, bound to a range of PTS1 sequences. A = *Hs*PEX5 with YQSKL, B = *Tb*PEX5 with FNELSHL, C = *Hs*PEX5 with AGT (5 C-terminal aa = PKKKL), D = *Hs*PEX5 with mutated AGT (5 C-terminal aa = PKAAL), E = *Mm*PEX5-like protein with SRLSSNL (Gatto et al., 2000 (1FCH), Sampathkumar et al., 2008 (3CV0), Fodor et al., 2012 (3R9A), Fodor et al., 2015 (4KXK), Bankston et al., 2012 (4EQF)).

Through observation of the structures of the PEX5 TPR domain in complex with the two PTS1-containing proteins, mSCP2 and AGT, it can be seen that the shape of the PTS1-binding site is adaptable. For example, the pocket seems to expand in order to accommodate the PTS1 sequence –KKL (**Figure 11, C**). Recent crystal structures show the binding of PEX5 to two mutated PTS1 sequences (KKL→AKL and KKL→AAL) at the C-terminus of a PTS1 protein AGT (Fodor et al., 2015). The PTS1 mutation KKL→AAL (**Figure 11, D**) allowed AGT to bind with higher affinity to PEX5 than the PTS1 mutation resulting in a canonical PTS1 sequence (KKL→AKL). This suggests that the overall size reduction of the PTS1 resulted in compensatory compaction of the PTS1-binding site. The dynamic nature of the PTS1 binding site could be the reason that PEX5 can bind to so many variants of PTS1.

Point mutations that have been discovered or created in the TPR domain of PEX5 are shown in **Table 2**. These mutations all affected the functioning of either PTS1- or PTS2-mediated import.

Variant	***At*PEX5 equivalent**	**Location of mutation**	**Import functional?**	**Reference**
*Hs*PEX5L N415S	N537S	TPR3	PTS1: **x** PTS2: ✓	Ebberink et al., 2009
*Hs*PEX5S N489K	N628K	TPR6	PTS1: **x** PTS2: ✓	Dodt et al., 1995
*Pp*PEX5 N460K	N628K	TPR6	PTS1: **x** PTS2: ✓	Dodt et al., 1995
*Hs*PEX5L N526K	N628K	TPR6	PTS1: **x** PTS2: ✓	Carvalho et al., 2007b; Ebberink et al., 2009
*Hs*PEX5L R557W	R659W	TPR7	PTS1: **x** PTS2: **x**	Ebberink et al., 2009

Table 2| Point mutations in the PEX5 TPR domain and their peroxisomal import effects. Mutations of the equivalent residues to N537 and N628 in *Arabidopsis thaliana* PEX5 resulted in defective PTS1-mediated import but functional PTS2-mediated import. Mutation of the equivalent residue to R659 in *Arabidopsis thaliana* PEX5 resulted in defective PTS1- and PTS2-mediated import. Alignments were performed using BioEdit (sequence alignment editor). *Hs*, *Homo sapiens*; *Pp*, *Pichia pastoris*.

Binding experiments have also been performed with PEX5 variants, and the effects of the mutations on PTS1 binding measured. These PEX5 variants are summarised in **Table 3**.

Variant	*At*PEX5 equivalent	Location of mutation	PTS1 binding? (K_d <5,000 nM or Y2H binding seen)	Reference
*Hs*PEX5S N489K	N628K	TPR6	✓ (K_d = 840 ± 80 nM)	Maynard and Berg, 2007
*Sc*PEX5 N360A	N504A	TPR2	✓ (Y2H binding seen)	Klein et al., 2001
*Sc*PEX5 E361K	D505K	TPR2	✓ (Y2H binding seen)	Klein et al., 2001
*Sc*PEX5 E363A	D507A	TPR2	✓ (Y2H binding seen)	Klein et al., 2001
*Sc*PEX5 I389D	V533D	TPR3	**x** (no Y2H binding seen)	Klein et al., 2001
*Sc*PEX5 N393A	N537A	TPR3	**x** (no Y2H binding seen)	Klein et al., 2001
*Sc*PEX5 N393D	N537D	TPR3	**x** (no Y2H binding seen)	Klein et al., 2001
*Sc*PEX5 N393S	N537S	TPR3	**x** (no Y2H binding seen)	Klein et al., 2001
*Sc*PEX5 N393Y	N537Y	TPR3	**x** (no Y2H binding seen)	Klein et al., 2001
*Sc*PEX5 E394A	E538A	TPR3	✓ (Y2H binding seen)	Klein et al., 2001
*Sc*PEX5 L404P	L548P	TPR3	**x** (no Y2H binding seen)	Klein et al., 2001
*Sc*PEX5 L465P	V598P	TPR5	**x** (no Y2H binding seen)	Klein et al., 2001
*Sc*PEX5 N503A	N636A	TPR6	✓ (Y2H binding seen)	Klein et al., 2001
*Sc*PEX5 N503D	N636D	TPR6	✓ (Y2H binding seen)	Klein et al., 2001
*Sc*PEX5 N503Y	N636Y	TPR6	✓ (Y2H binding seen)	Klein et al., 2001
*Sc*PEX5 S504A	S637A	TPR6	✓ (Y2H binding seen)	Klein et al., 2001
*Sc*PEX5 N505A	V638A	TPR6	✓ (Y2H binding seen)	Klein et al., 2001
*Sc*PEX5 R526A	R659A	TPR7	✓ (Y2H binding seen)	Klein et al., 2001
*Sc*PEX5 S534L	S667L	TPR7	✓ (Y2H binding seen)	Klein et al., 2001
*Sc*PEX5 N537A	N670A	TPR7	✓ (Y2H binding seen)	Klein et al., 2001

Table 3| Point mutations in the PEX5 TPR domain and their PTS1-binding effects. The mutation which resulted in defective PTS1-mediated import *in vivo* (*Hs*PEX5S N489K) still appears to bind a representative PTS1 peptide with a relatively high affinity. Key residues, that result in loss of PTS1 binding when mutated, are those equivalent to N537 and V533 in *Arabidopsis thaliana* PEX5. Alignments were performed using BioEdit (sequence alignment editor). *Hs*, *Homo sapiens*; *Sc*, *Saccharomyces cerevisiae*; TPR, tetratricopeptide repeat; Y2H, yeast-two-hybrid.

In all of the studies reported in **Table 3**, [S]-[K]-[L]-CO_2H was used as the representative native PTS1 sequence. This signal sequence is, however, extremely variable and many sequences can function as PTS1s.

1.5 What makes a PTS1 sequence?

As previously stated, the first PTS1 sequence discovered was -SKL at the C-terminus of wild-type firefly luciferase. This protein is found in cells of the lantern organ of the firefly, in which it is localised to peroxisomes (Keller et al., 1987). The -SKL sequence of luciferase was found to have the ability to target luciferase to insect peroxisomes as, when the three C-terminal residues were removed, luciferase was no longer imported into peroxisomes and was found in the cytosol (Gould et al., 1987; Gould et al., 1988; Gould et al., 1989). Mutations were then carried out in the three C-terminal residues of luciferase in order to define what constitutes a PTS1. At this point, the original consensus sequence was defined as [S/A/C]-[K/R/H]-[L]-CO_2H (Gould et al., 1989). Peroxisomal protein import was found to be conserved across yeast, plants, insects and mammals when firefly luciferase was expressed in each of these cell types and was transported into peroxisomes in all cases (Gould et al., 1990). In one example, a short C-terminal sequence from peroxisomal protein glycolate oxidase (-[R]-[A]-[V]-[A]-[R]-[L]-CO_2H) was found to target β-glucuronidase, a non-peroxisomal protein, to plant peroxisomes (Volokita, 1991).

The PEX5:PTS1 interaction is used as a means by which peroxisomes can be visualised (Monosov et al., 1996). In order to do this, a PTS1 sequence is attached to the C-terminus of a fluorescent protein. In **Figure 12**, CFP-PTS1 has been imported into peroxisomes, which highlights the peroxisomes in the moss (*Physcomitrella patens*) cells imaged.

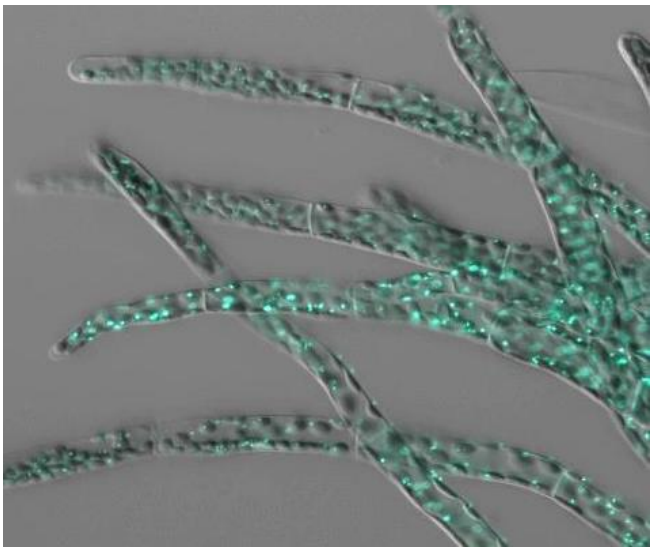


Figure 12| Peroxisomes in the moss *Physcomitrella patens*, visualised using CFP-PTS1. Image courtesy of Yasuko Kamisugi.

Binding of naturally-occurring PTS1 sequences by PEX5 has been measured quantitatively using fluorescence anisotropy. This assay involves the titration of an increasing concentration of protein against a fixed concentration of fluorescently labelled peptide (representing a PTS1 sequence). Anisotropy can then be used to calculate the dissociation constant (K_d) for each interaction tested. The affinity of PEX5 for strong PTS1 sequences is typically very high, with ~nM dissociation constants (Gatto et al., 2003; Harper et al., 2003; Skoulding et al., 2015).

A number of mutational studies have been performed to look at the binding of a wide range of PTS1 sequences to wild-type PEX5 (Klein et al., 2001; Gatto et al., 2003; Maynard et al., 2004; Maynard and Berg, 2007; Ghosh and Berg, 2010). One research group created *Saccharomyces cerevisiae* PEX5 mutants and investigated their binding to a PTS1 sequence, SKL (Klein et al., 2001). Several PEX5 mutations resulted in loss of wild-type PTS1 interaction (**Table 3**). A suppressor screen was also carried out by Klein and colleagues to find mutations that could allow the binding of PEX5 to another C-terminal sequence, SEL, which is a non-PTS1 when attached to the C-terminus of GFP (Distel et al., 1992). [S]-[E]-[L]-CO_2H has also been used as a negative control for PEX5 binding (Dodt et al., 1995; Gatto et al., 2000) and peroxisomal import (Gould et al., 1989). Mutations that allowed the interaction of PEX5 with SEL were E361K, N503D, N503Y and S534L (shown in **Table 3**, along with equivalent *At*PEX5 residues). **Figure 10** shows that these equivalent *At*PEX5 residues are all predicted to associate with the PTS1 sequence.

In plants, chimeric genes have been constructed by Hayashi and colleagues, which include the addition of PTS1 variant sequences onto the C-terminus of the non-peroxisomal protein β-glucuronidase. Immunoelectron-microscopy was then used to look at the cellular location of these gene products (Hayashi et al., 1997). Yeast-two-hybrid screens have also been performed with large libraries of peptides representing variant PTS1 sequences (Lametschwandtner et al., 1998). The results of these studies were collectively used to define a PTS1 consensus sequence.

1.5.1 Variation to the PTS1 by mutation, and the effect on peroxisomal import and on PEX5 binding

Many mutations to the PTS1 sequence have been performed and, as this study will focus on altering the two C-terminal PTS1 residues, mutants of interest and their effects on peroxisomal import are shown in **Table 4**.

{protein}-C-terminus	Localised to peroxisomes?	Reference
{β-glucuronidase}-SKL	✓ (efficiently)	Hayashi et al., 1997
{β-glucuronidase}-SRL	✓ (efficiently)	Hayashi et al., 1997
{β-glucuronidase}-SIL	**x**	Hayashi et al., 1997
{β-glucuronidase}-SGL	**x**	Hayashi et al., 1997
{β-glucuronidase}-SSL	**x**	Hayashi et al., 1997
{β-glucuronidase}-SHL	**x**	Hayashi et al., 1997
{β-glucuronidase}-SRI	✓ (detectably)	Hayashi et al., 1997
{β-glucuronidase}-SRV	**x**	Hayashi et al., 1997
{β-glucuronidase}-SRM	✓ (efficiently)	Hayashi et al., 1997
{β-glucuronidase}-SRS	**x**	Hayashi et al., 1997
{β-glucuronidase}-SRE	**x**	Hayashi et al., 1997
{β-glucuronidase}-SRK	**x**	Hayashi et al., 1997
{eYFP}-VHIQVRHSSM	✓ (weakly)	Lingner et al., 2011
{eYFP}-TENERIKSML	✓	Lingner et al., 2011
{eYFP}-TEGDRIRSLL	✓	Lingner et al., 2011
{eYFP}-VQVRVGHSNM	✓	Lingner et al., 2011
{eYFP}-SQINQAKSQL	✓	Lingner et al., 2011
{eYFP}-DTSPRTKSTL	✓	Lingner et al., 2011
{eYFP}-NNTPLIASRV	✓ (weakly)	Lingner et al., 2011
{eYFP}-VAKTTRPSRV	✓	Lingner et al., 2011
{eYFP}-LHKEDLKSHI	✓	Lingner et al., 2011
{eYFP}-LSRDVIPSEL	✓ (weakly)	Lingner et al., 2011
{eYFP}-LATPDLRSFM	✓	Lingner et al., 2011
{eYFP}-ASIPLLISRF	✓	Lingner et al., 2011
{eYFP}-DKFSAIPSGL	✓ (weakly)	Lingner et al., 2011
{eYFP}-SLFNKLRSKV	✓	Lingner et al., 2011
{eYFP}-SDIFPKPSEM	**x**	Lingner et al., 2011
{eYFP}-ISVPFLISPL	✓ (weakly)	Lingner et al., 2011
{eYFP}-DFQPLPPSPL	**x**	Lingner et al., 2011
{eYFP}-NNIPMSPSGI	**x**	Lingner et al., 2011
{eYFP}-SSIKAMLSTI	✓ (weakly)	Lingner et al., 2011
{eYFP}-VAKTTRPSRM	✓ (efficiently)	Skoulding et al., 2015
{eYFP}-VAKTTRPSRI	✓ (efficiently)	Skoulding et al., 2015
{eYFP}-VAKTTRPSRV	✓ (weakly)	Skoulding et al., 2015
{eYFP}-VAKTTRPSRY	✓ (detectably)	Skoulding et al., 2015
{eYFP}-VAKTTRPSRK	**x**	Skoulding et al., 2015
{eYFP}-VAKTTRPSNV	✓ (weakly)	Skoulding et al., 2015
{eYFP}-VAKTTRPSTV	**x**	Skoulding et al., 2015

Table 4| Point mutations in PTS1 and the resulting effect on peroxisomal import in plants. Many more PTS1 sequences than those displaying the properties [small]-[basic]-[hydrophobic]-CO_2H are targeted to the peroxisome.

Binding experiments have also been performed with plant PTS1 variants, and the binding affinities measured. These PTS1 variants are summarised in **Table 5**.

C-terminal sequence	K_i < 5,000 nM?	Reference
VAKTTRPSKL	✓	Skoulding et al., 2015
VAKTTRPSRL	✓	Skoulding et al., 2015
VAKTTRPSRM	✓	Skoulding et al., 2015
VAKTTRPSRI	**x**	Skoulding et al., 2015
VAKTTRPSRV	**x**	Skoulding et al., 2015
VAKTTRPSRY	**x**	Skoulding et al., 2015
VAKTTRPSNV	**x**	Skoulding et al., 2015
YQSKL	✓	Skoulding et al., 2015
YQSKV	**x**	Skoulding et al., 2015
VAKTTRPSNM	**x**	Skoulding, 2011
YQSEL	**x**	Skoulding, 2011

Table 5| Point mutations in PTS1 and the resulting effect on plant PEX5 binding. Only 'strong' PTS1 sequences (as determined *in vivo*) had a K_i below 5 μM, as determined by fluorescence anisotropy competition assays, where the fluorescent peptide being out-competed was lissamine-YQSKL.

1.5.2 Extension of the PTS1 consensus through *in silico* predictions

As previously stated, the PTS1 sequence had been thought of as having a clear consensus of [small]-[basic]-[hydrophobic]-CO_2H. With the increased availability of genomic and proteomic data, it has been possible for researchers to identify which proteins are naturally present in the peroxisome and, therefore, to predict and test the elements needed for a functional PTS1 sequence (Neuberger et al., 2003a; Neuberger et al., 2003b; Lingner et al., 2011; Reumann et al., 2012). For *A. thaliana* PTS1 sequences, the website AraPerox provides a comprehensive list of major, minor, and rare PTS1 sequences (Reumann et al., 2004). A server called PredPlantPTS1 (Lingner et al., 2011; Reumann et al., 2012) uses a positional weight matrix (PWM) scoring system to be able to predict whether a protein will be imported into the peroxisome *via* the PTS1 import pathway or not, when given the C-terminal 14 amino acids of the primary sequence of the protein. Peroxisomal targeting prediction scores, as determined by PredPlantPTS1, have recently been compared to *in vitro* binding assays and *in vivo* peroxisomal protein import efficiency of sequences carried out in

onion epidermal cells (Skoulding et al., 2015). This work showed that there was agreement between these methods, although in some cases the *in vivo* import experiments were more sensitive than *in vitro* assays (Skoulding et al., 2015).

A recent paper summarising all possible PTS1 sequences stated that 'canonical' residues that can be found in each position are: [S/A]-[R/K]-[L/M/I]-CO_2H, and 'non-canonical' residues that can be found in each position are: [P/C/F/V/G/T/L/K/I/Q]-[S/N/L/M/H/G/E/T/F/P/Q/C/Y/D/A]-[V/Y/F]-CO_2H (Reumann et al., 2016). Reumann and colleagues state that, for a peptide to function as a plant PTS1, at least two canonical residues must be present. This means that, even when only considering the three C-terminal residues, there are over 80 possible native PTS1 sequences.

1.6 Peroxisomes and synthetic biology

Peroxisomes are attractive organelles for synthetic biology, as they maintain a barrier to the cytosol while allowing the import of folded, large and sometimes oligomeric proteins. Import of the majority of these proteins is determined only through a short signal sequence (reviewed in Baker et al., 2016). Peroxisomal protein import, therefore, appears to be an exploitable system for the creation of a synthetic organelle within the cell.

Peroxisomal protein import has previously been manipulated for optical control of peroxisomal protein trafficking (Spiltoir et al., 2016). This was performed using a construct with a 'caged' PTS1, which becomes uncaged from the rest of the protein when blue light is used so the PTS1 is then free to bind to PEX5 and transport the full protein into the peroxisome. Very recently, the fungal peroxisome was used as a model in which non-peroxisomal proteins were imported into the peroxisome and then carried out a non-peroxisomal pathway inside peroxisomes (DeLoache et al., 2016). This was achieved through out-competition of a natural PTS1 sequence using enhancing residues on a PTS1 sequence to create an 'enhanced PTS1' (DeLoache et al., 2016). Peroxisomes have also been suggested as suitable organelles for engineering in a recent review (Kessel-Vigelius et al., 2013), which highlights their potential to: accommodate novel pathways and compartmentalise these within the cell, allow fully folded proteins to be transported across the peroxisome membrane, and proliferate in response to their environment.

1.7 Directed evolution and orthogonality

Directed evolution has been used by other research groups to achieve orthogonality, for example in the case of a v-Src tyrosine kinase which was re-designed by mutation to bind to a radioactive analogue of ATP rather than the natural ATP substrate (Liu et al., 1998). Downstream targets of the kinase could then be identified by their labelling with the radioactive ATP analogue. Another example of achieving orthogonality through directed evolution is in the re-design of a conserved protein-protein interaction to produce an interaction that would not have any cross-interference with native cell pathways (Speltz et al., 2015; Pratt et al., 2016). This interaction was successfully used as a method of specific in-cell imaging, when one of the binding partners was genetically fused to a fluorescent protein and the other binding partner was genetically fused to the protein of interest for visualisation (Pratt et al., 2016).

A classic example of an orthogonal system is the orthogonal ribosome. Orthogonal ribosomes were created to allow the incorporation of unnatural amino acids (UAAs) into protein sequences during protein synthesis. This process can then introduce desired properties to a protein. The creation of orthogonal ribosomes has been carried out using saturation mutagenesis of selected residues in a specific region of the ribosomal A-site. This region is important for tRNA (transfer RNA) binding so, with the introduction of tRNA molecules aminoacylated with UAAs, orthogonal ribosomes can be screened to recognise these. UAAs are then incorporated into the protein sequence (Wang et al., 2007; Bain et al., 1989). Orthogonal tRNA-synthetases have also been produced (Xie and Schultz, 2006). The method used was random mutagenesis which can often result in vast library numbers: the library of tRNA-synthetases produced was reported at ~10^9 mutants.

In order to hijack the peroxisome with a new non-PTS1 sequence which is imported preferentially by a mutant of the PEX5 protein, an orthogonal interaction will have to be created. The importance of an orthogonal interaction is that it can function alongside, yet independently of, the natural pathway. We aim to use directed evolution to create this orthogonal interaction, using PEX5:PTS1 as a template.

1.8 Objectives of this study

The aim of this study is to evolve a novel PEX5:PTS1-style interaction, which will be termed PEX5*:PTS1*. PEX5*:PTS1* will be distinct from interactions seen for the natural import of PTS1-cargo proteins. This interaction will allow for the direction of non-peroxisomal proteins to the matrix of the peroxisome. As the function of peroxisomes is determined by their protein content (Reumann and Bartel, 2016), the overall aim will be to re-purpose the peroxisome. Binding of a range of non-PTS1 sequences (PTS1* candidates) will be tested with a polyhistidine-tagged truncated construct of *Arabidopsis thaliana* PEX5 (*At*PEX5) comprising only the C-terminal, PTS1-binding domain of PEX5. By creating a mutant library of truncated PEX5, or His_6-*At*PEX5(340-728), the objective is to find a variant of His_6-*At*PEX5(340-728) that binds a non-PTS1 yet shows minimal, or ideally no, binding activity with natural PTS1 sequences. The binding will be optimised by subjecting the successful PTS1* binder to further mutations. The optimised PEX5*:PTS1* pair will then be provided to Dr Rupesh Paudyal for *in vivo* testing.

Specific goals of the study are to:

- Develop a screen that will allow the testing of His_6-*At*PEX5(340-728) variants with PTS1 and non-PTS1 sequences
- Use the above steps to create and optimise an orthogonal PEX5*:PTS1* interaction and test binding using fluorescence anisotropy
- Provide the optimised PEX5*:PTS1* pair to R. Paudyal for *in vivo* testing in a model plant organism (the moss *Physcomitrella patens*).

Success in this project could lead to the targeting of user-defined proteins to the peroxisome. If orthogonal binding of PEX5 and PTS1 can be determined *in vitro*, the next objective will be to move this system *in vivo* (for experiments to be carried out by R. Paudyal) and incorporate a switch between natural PTS1-mediated import and the designed orthogonal pathway.

Chapter 2

Attempted Rational Design of the PEX5*:PTS1* Orthogonal Interaction and Initial Screen Development

2.1 Introduction

As described in Chapter 1 (**Figure 3**), the peroxisomal targeting signal-1 (PTS1) mediated import pathway allows the peroxisomal import of proteins possessing a C-terminal PTS1 sequence. This process requires the binding of the PTS1 by the receptor peroxin-5 (PEX5). The binding site for the PTS1 is located in the C-terminal domain of PEX5, and PEX5 is a modular protein, so a construct comprising only this part of the protein has been used for binding studies (termed PEX5-C). Previous work at the University of Leeds had shown that *A. thaliana* PEX5-C had a similar PTS1-affinity to the full-length *A. thaliana* PEX5 protein (Skoulding, 2011; Skoulding et al., 2015). Additionally, the full-length PEX5 protein was much less stable than the C-terminal domain alone (Skoulding, 2011), so the *A. thaliana* PEX5-C construct (*At*PEX5-C) was used for all work presented in this chapter. This construct had previously been cloned into vector pET-28b by Sarah Gunn (former student of the University of Leeds) to include an N-terminal His_6 tag for purification (Gunn, 2008).

The aim of this study is to create an orthogonal *At*PEX5-C:PTS1 interaction, which will be termed *At*PEX5-C*:PTS1*. Rational design of the *At*PEX5-C binding site was attempted, which required a structural model in order to predict mutations that would change the binding specificity of the protein. No crystal structure of *A. thaliana* PEX5 has yet been obtained; however, there are a number of crystal structures showing PEX5-C, in complex with PTS1 sequences, from *H. sapiens* and *T. brucei* (Gatto et al., 2000; Stanley et al., 2006; Stanley et al., 2007; Sampathkumar et al., 2008; Fodor et al., 2012). The first *Hs*PEX5-C:PTS1 structure (Gatto et al., 2000) was used to perform homology modelling of the *At*PEX5-C protein (Skoulding, 2011), which showed that the two proteins (*Hs*PEX5-C and *At*PEX5-C) are predicted to have almost identical structure in the tetratricopeptide repeat (TPR) helices of the PTS1-binding site. The sequence identity of *At*PEX5-C with the *Hs*PEX5-C construct used by Gatto and

colleagues is 38% (sequence similarity = 53%) with 96% coverage (performed using BLAST (blastp)). Sequence identity of *Hs*PEX5-C TPRs 1-3 and 5-7, as stated by UniProtKB (TPRs 1-3 = 335-436 and TPRs 5-7 = 488-589), with all equivalent sequence in *At*PEX5-C (TPRs 1-3 ≈ 457-558 and TPRs 5-7 ≈ 590-691) is 47% (sequence similarity = 65%) with 100% coverage (Altschul et al., 1990 (BLAST); Apweiler et al., 2004 (UniProtKB)) (alignments were performed using BioEdit (sequence alignment editor). An alignment of all *At*PEX5-C residues predicted by I-TASSER as being involved in PTS1 binding (18 residues), with equivalent residues in the *Hs*PEX5-C construct used by Gatto and colleagues was performed. This alignment was performed using BLAST (blastp), after equivalent residues had been determined using BioEdit (sequence alignment editor), and revealed a sequence identity of 83% and a sequence similarity of 88% (Altschul et al., 1990).

This chapter will describe the expression and purification of *At*PEX5-C and a binding study carried out with this protein and a representative PTS1 sequence. Also explained will be the selection of potential PTS1* sequences and the subsequent selection of target amino acids in the *At*PEX5-C protein for mutagenesis. *In vitro* testing of combinations of *At*PEX5-C variants and their designed binding peptides will then be covered, along with the development of screening methodologies to allow investigation into the binding of multiple *At*PEX5-C variants to peptides in one assay.

2.2 Expression and purification of *At*PEX5-C

Expression of the *At*PEX5-C protein from the *At*PEX5-C gene in pET-28b plasmid was performed by autoinduction (section 6.10.2). Purification of *At*PEX5-C was then achieved by affinity chromatography using cobalt-agarose resin (section 6.10.4). The stages of purification of *At*PEX5-C were analysed by SDS-PAGE (section 6.11) and this analysis is shown in **Figure 13**.

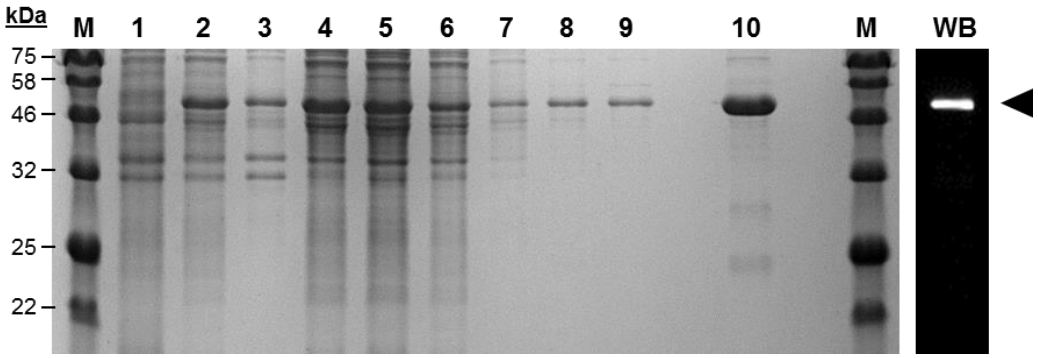


Figure 13| Purification SDS-PAGE gel for *At*PEX5-C. A band at approximately 50 kDa was seen on SDS-PAGE with Quick Blue coomassie stain (10), which was confirmed as AtPEX5-C using western blotting (WB) of a duplicate gel with α-PEX5 antibody. Filled arrowhead highlights the presence of *At*PEX5-C. M, protein standard markers; 1, non-induced BL21-Gold (DE3) cells containing His_6-*At*PEX5-C.pET-28b; 2, autoinduced BL21-Gold (DE3) cells containing His_6-*At*PEX5-C.pET-28b; 3, pellet after cell lysis; 4, supernatant after cell lysis; 5, supernatant after incubation with cobalt-agarose resin; 6-9, buffer washes 1-4 of cobalt-agarose resin; 10, elution of protein using 200 mM imidazole; WB, elution lane in western blot of duplicate gel.

Analysis of purification fractions by SDS-PAGE and western blotting (**Figure 13**) confirmed that the *At*PEX5-C protein was being isolated. Gel filtration was then performed to assess the purity of the protein, and electrospray ionisation mass spectrometry was used to confirm the mass of the purified protein (**Figure 14**). The calculated mass of *At*PEX5-C with a hexahistidine tag was 45,580.3 Da, and the observed mass was 45,580.3 Da.

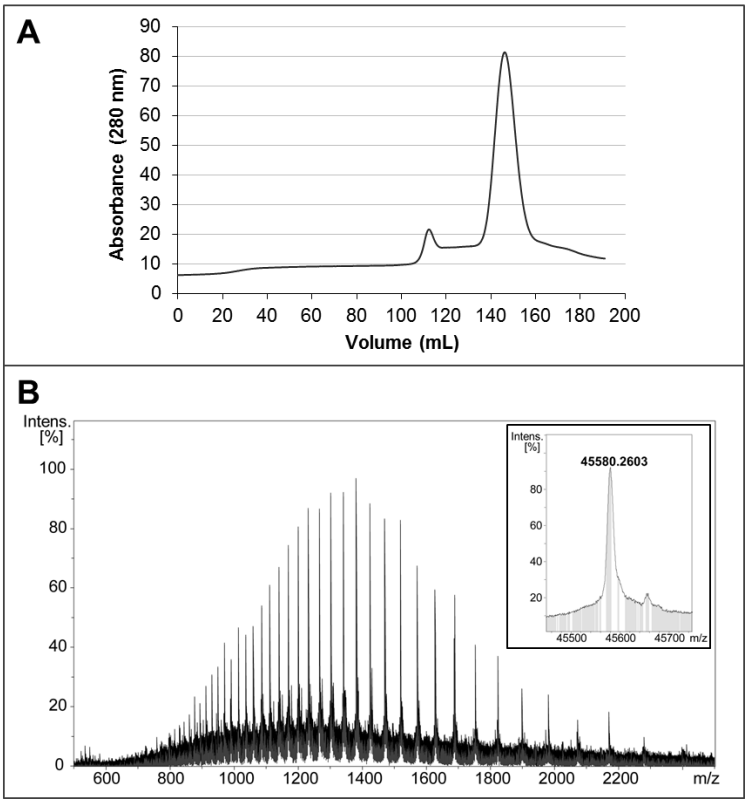


Figure 14| A) Gel filtration trace for *At*PEX5-C. B) Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C. Predicted elution volume based on protein monomer mass: 145.4 mL. Actual elution volume (large peak): 146.2 mL. Expected mass: 45,580.3 Da. Observed mass: 45,580.3 Da.

The SDS-PAGE gel, western blot, and mass spectrum together confirmed that the *At*PEX5-C protein had been successfully isolated. Two peaks could be observed in the gel filtration trace, the larger of which corresponded to the approximate size of monomeric *At*PEX5-C (predicted elution volume: 145.4 mL, and actual elution volume: 146.2 mL). The smaller peak in the gel filtration trace corresponded to the approximate size of the dimeric species of *At*PEX5-C (predicted elution volume: 114.0 mL, and actual elution volume: 112.3 mL). This could be due to the presence of five WxxxF/Y (typically PEX14-binding) repeats (where x is any amino acid) in the protein construct which could promote self-association: PTS1 import has been observed in plants in the absence of PEX14, so it is possible that PEX5 could be self-associating in this case to form a pore through which PTS1 proteins can be imported into peroxisomes (Monroe-Augustus et al., 2011). *At*PEX5-C could also be forming disulfide bonds between monomers. With this in mind, the gel filtration trace confirmed that the protein was pure.

With purity and a correct mass confirmed, the *At*PEX5-C protein was carried forward for *in vitro* binding studies using synthetic peptides corresponding to PTS1 and non-PTS1 sequences.

2.3 Synthesis of a 'native' PTS1 peptide and a subsequent binding study with *At*PEX5-C

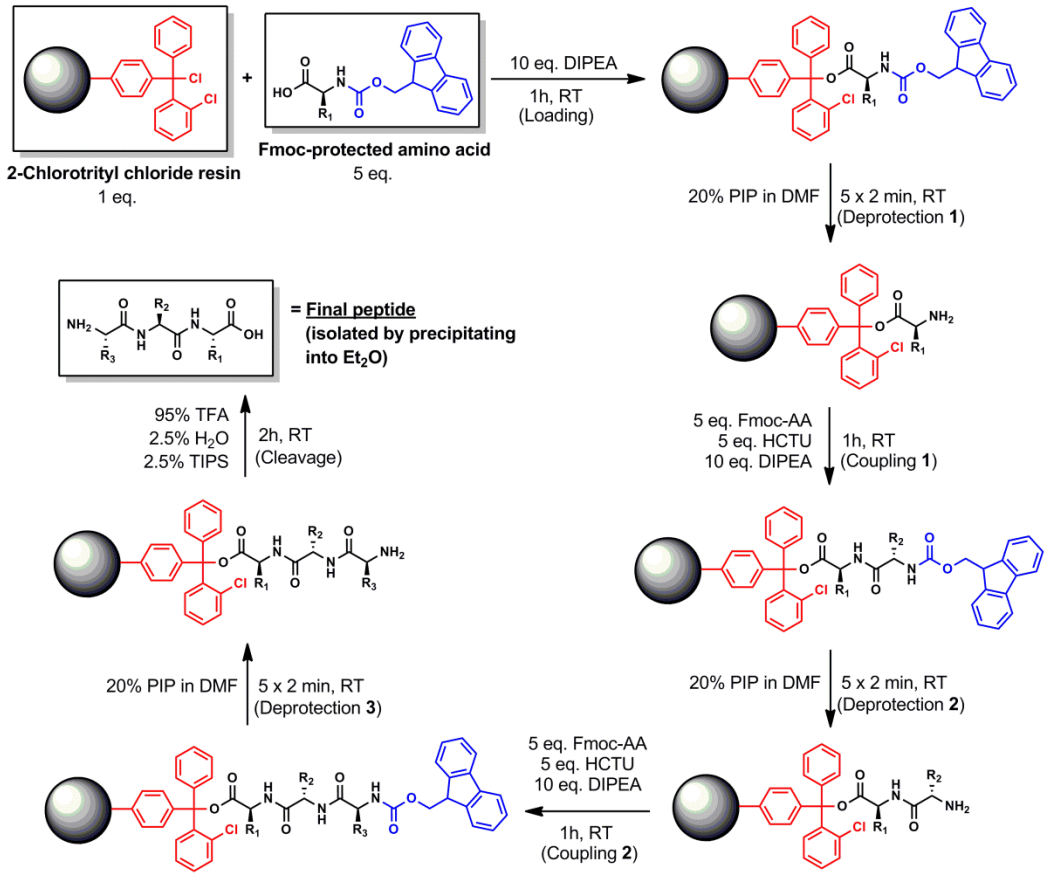
The first PTS1 sequence was identified as –Ser-Lys-Leu-CO_2H, or –SKL>[1], and this was found at the C-terminus of firefly luciferase (Gould et al., 1987; Gould et al., 1988). This short signal sequence was discovered to be responsible for the import of the luciferase protein into peroxisomes, where luciferase is found in the lantern organ of the firefly (Keller et al., 1987). The fusing of the –SKL> sequence at the C-terminus of a non-peroxisomal protein, chloramphenicol acetyltransferase (CAT), was able to redirect this protein to the peroxisome (Mullen et al., 1997). As peroxisomal import research developed, –SKL> was found to be one of the most common PTS1 sequences found on proteins destined for the peroxisome.

To date, several studies on the PEX5:PTS1 interaction have been performed. The first crystal structure of PEX5 in complex with a PTS1 sequence was solved by Gatto and colleagues (Gatto et al., 2000 (1FCH)). This structure featured the C-terminal domain of *H. sapiens* PEX5 (*Hs*PEX5-C) in complex with the pentapeptide H_2N-Tyr-Gln-Ser-Lys-Leu-CO_2H, or YQSKL. Three crystal structures of PEX5-C in complex with a PTS1 sequence terminating in –SKL> have now been published (Gatto et al., 2000 (1FCH); Sampathkumar et al., 2008 (3CVP and 3CVQ)). Isothermal titration calorimetry (ITC) has also been used to study the PEX5-C: –SKL> interaction using peptides with varying sequences upstream of the C-terminal –SKL> and the dissociation constants calculated were between 14 nM and 145 nM (Mesa-Torres et al., 2015). As YQSKL is the peptide sequence best characterised in complex with PEX5-C, YQSKL has been used to represent a native PTS1 sequence in this work.

In order to study the PEX5:PTS1 interaction *in vitro*, the peptide sequence YQSKL was synthesised on solid support using Fmoc solid phase peptide synthesis (SPPS)

[1] The '>' symbol represents the C-terminus of a protein where only a portion of the polypeptide sequence is shown.

(Merrifield, 1963) (**Scheme 1**). Characterisation of this peptide is shown in **Figure 15** and confirms that a peptide of the correct mass had been synthesised and that this peptide was pure.



Scheme 1| Outline of synthesising peptides by Fmoc SPPS, using a tripeptide as an example. In this technique, peptides are synthesised from C-terminus to N-terminus. After synthesis peptides are cleaved from the solid support (resin), yielding the free carboxylic acid group. Shown in red are acid-labile groups (the activated 2-chlorotrityl chloride on resin), and shown in blue are base-labile groups (the Fmoc protecting group). PIP, piperidine; TIPS, triisopropylsilane; Fmoc-AA, Fmoc-protected amino acid.

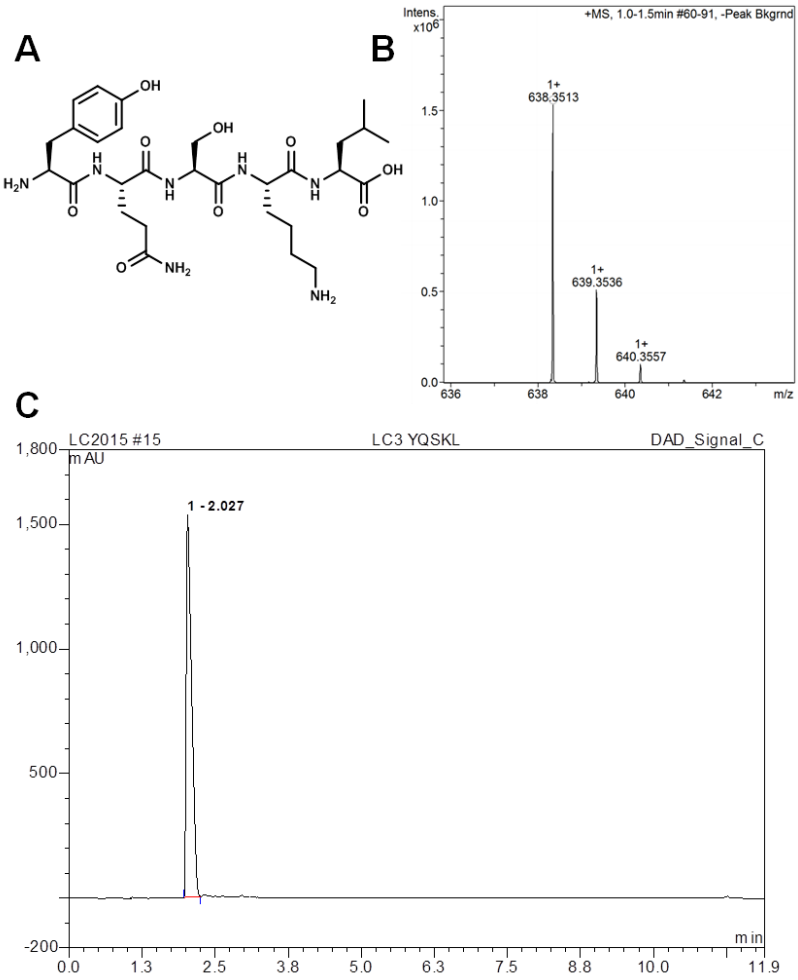


Figure 15| A) Structure of the PTS1 pentapeptide YQSKL. B) Isotopic distribution in mass spectrum of YQSKL. C) Analytical HPLC trace of YQSKL. Expected $[M+H]^+$: 638.3508 g/mol. Observed $[M+H]^+$: 638.3513 g/mol. The analytical HPLC trace confirms peptide purity.

An N-terminal fluorescent label (lissamine-rhodamine) was added to the peptide in order to use the resulting fluorescent peptide in a binding assay. From this point onwards, lissamine-rhodamine will be referred to as lissamine. Lissamine was added before cleavage of the peptide from the solid support resin (**Scheme 1**). Characterisation of lissamine-YQSKL is shown in **Figure 16** and confirms that a pure peptide of the correct mass was obtained.

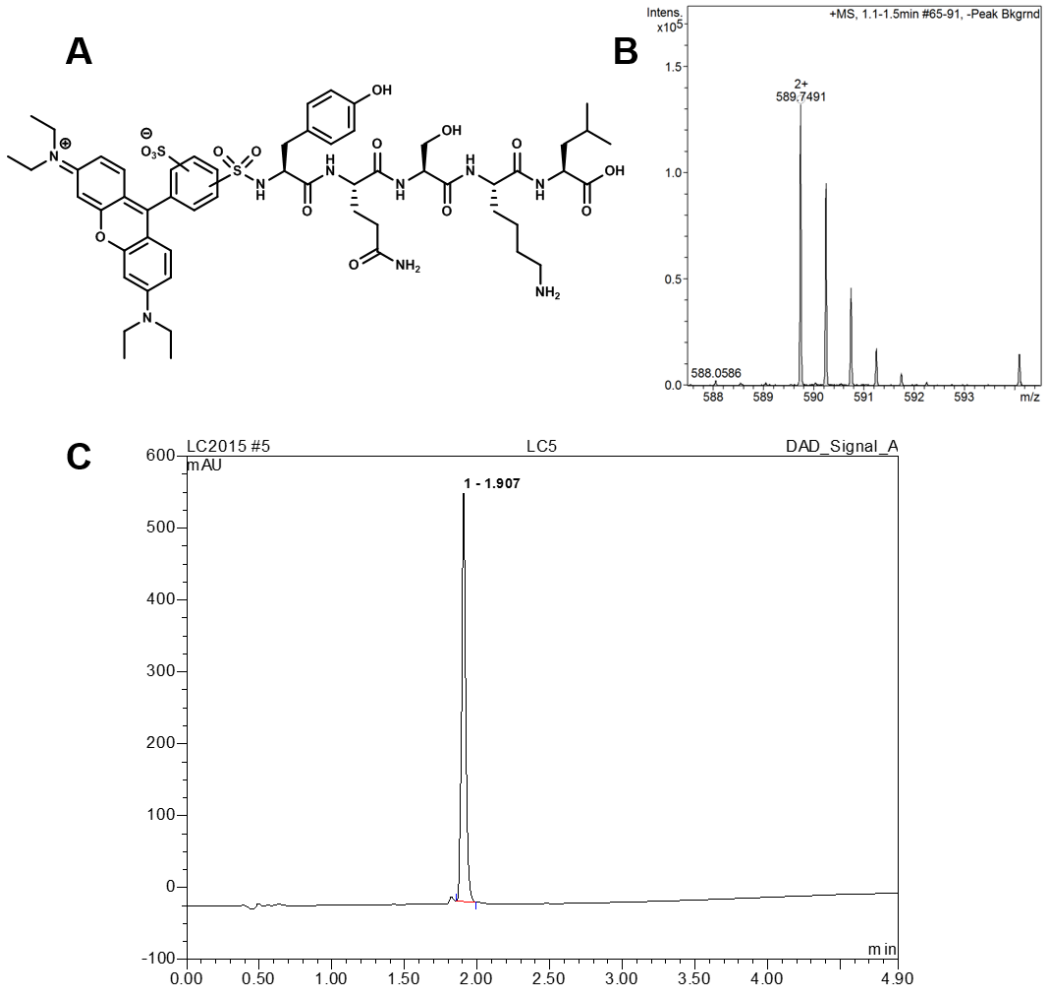


Figure 16| A) Structure of the fluorescent PTS1 pentapeptide lissamine-YQSKL. B) Isotopic distribution in mass spectrum of lissamine-YQSKL. C) Analytical HPLC trace of lissamine-YQSKL. Expected $[M+2H]^{2+}$: 589.7485 g/mol. Observed $[M+2H]^{2+}$: 589.7491 g/mol. The analytical HPLC trace confirms peptide purity.

With *At*PEX5-C purified and YQSKL synthesised and labelled, fluorescence anisotropy could be used to assess binding. Fluorescence anisotropy is a technique which can be used to calculate the dissociation constant of an interaction when one binding partner is significantly smaller than the other and fluorescently labelled. A small fluorescently labelled ligand will tumble quickly in solution, relative to the fluorescence lifetime, due to its small size. When the fluorophore attached to the ligand is excited by light in a particular orientation of polarisation, the light emitted by the fluorophore will be in various orientations of polarisation. When an increasing concentration of ligand-binding protein is titrated into the ligand solution, more of the ligand will become increasingly bound to the protein. This slows down the tumbling rate of the fluorescent ligand until the maximum amount of ligand is bound to the protein. At this point the fluorescent molecule is now effectively much larger and results in little change in the

polarisation orientation of light emitted by the fluorophore when compared to the polarisation orientation of light used to excite the fluorophore (**Figure 17**).

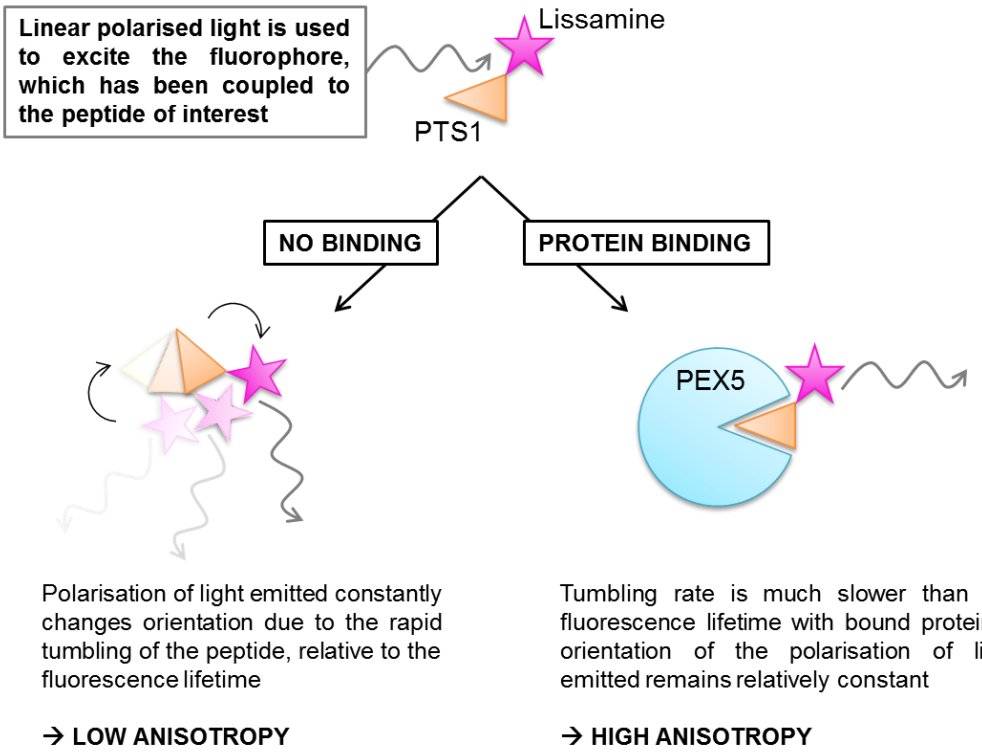


Figure 17| The principle behind fluorescence anisotropy as a technique for the detection of protein-peptide interactions. The fluorophore attached to a binding peptide is excited with linear polarised light. If light emitted is in various orientations of polarisation, this suggests that the fluorophore is bound to a small ligand. If, however, polarised light emitted remains linear, the fluorophore is part of a large complex which indicates protein binding of the ligand. Figure adapted from (Skoulding, 2011).

For the fluorescence anisotropy carried out in this work, the final ligand concentration used was constant at 100 nM and protein concentration was increased from 0 µM to 13.3 µM final concentration (**Figure 18**). The equation which was used to calculate anisotropy is shown in Chapter 6 (Experimental) section 6.14.7. The K_d of the *At*PEX5-C:Lissamine-YQSKL interaction calculated from **Figure 18** was 1.1 ± 0.6 nM.

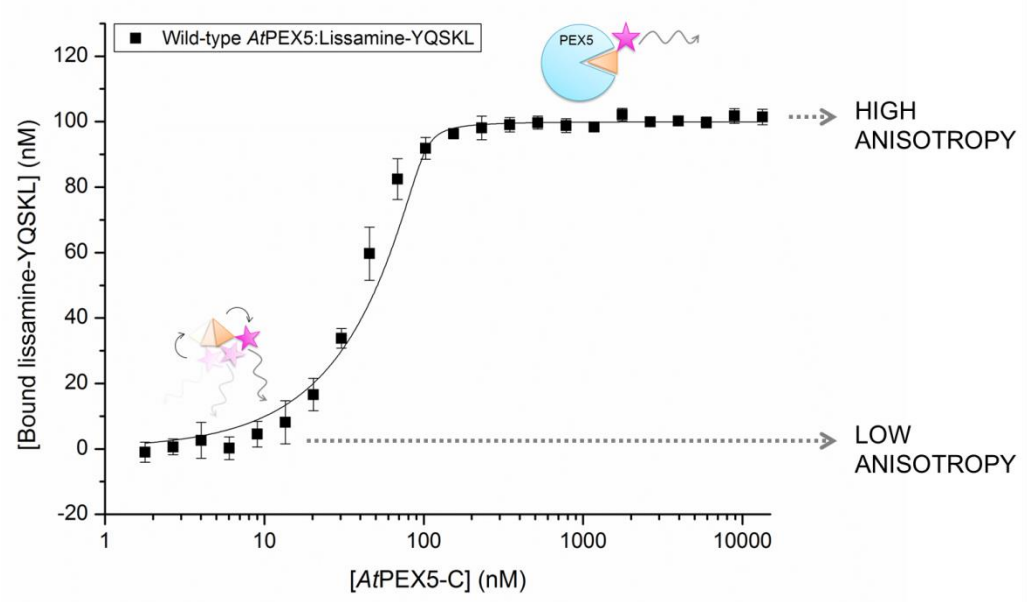


Figure 18| *At*PEX5-C (0–13.3 µM final concentration) versus amount of lissamine-YQSKL bound in the presence of 100 nM peptide, measured using fluorescence anisotropy. 'Low anisotropy' and 'high anisotropy' examples are shown on the graph to demonstrate where protein binding is seen. A quadratic equation was used to fit the curves using OriginPro 9.1, assuming a one-to-one model of binding.

Fluorescence anisotropy has been used by other researchers to measure the affinity of the PEX5-C:PTS1 interaction (Gatto et al., 2000; Harper et al., 2003; Maynard et al., 2004; Gunn, 2008; Ghosh and Berg, 2010; Lanyon-Hogg et al., 2014; Skoulding et al., 2015). The low nanomolar K_d obtained in this study is consistent with previously published PEX5:PTS1 dissociation constants (Ghosh and Berg, 2010; Gunn, 2008). The tightest affinity seen for human PEX5 binding to lissamine-YQSKL is a K_d of 2.7 nM (Ghosh and Berg, 2010), and for *Arabidopsis* PEX5 with lissamine-YQSKL, the tightest affinity found was a K_d of 3.10 nM (Gunn, 2008).

2.4 Potential PTS1* peptide sequences

In order to evolve an orthogonal PEX5*:PTS1* interaction, both binding partners of the PEX5:PTS1 interaction must be altered. As this was a large multidimensional problem, it was decided to start by designing potential PTS1* sequences which would not function as PTS1 sequences and use these for the evolution of the PEX5* protein.

Taking the most common native PTS1 sequence –SKL> as an example, an obvious substitution to make in the peptide sequence in an attempt to disrupt binding would be the exchange of the positively charged lysine for a negatively charged residue, for example glutamic acid. This was one of the first substitutions to be made in the PTS1

sequence of firefly luciferase by the Subramani laboratory in 1989 in order to investigate PTS1 binding requirements (Gould et al., 1989). Another substitution made at the time was the exchange of the C-terminal leucine to valine. Both of these substitutions resulted in firefly luciferase residing in the cytosol (Gould et al., 1989). More recent work has expanded on the reasons behind this lack of peroxisomal targeting when these substitutions are made. For example the crystal structure solved of *Hs*PEX5-C:YQSKL in 2000 (Gatto et al., 2000) revealed a pocket of acidic side chains surrounding the lysine residue in the PTS1, and a hydrophobic region of side chains surrounding the side chain carbons in the C-terminal leucine of the PTS1 (Gatto et al., 2000).

Quantitative binding studies have been performed for the peptide sequences YQSEL and YQSKV with *At*PEX5-C (Skoulding, 2011; Skoulding et al., 2015). The inhibition constant (K_i) of YQSKV with *At*PEX5-C, when out-competing lissamine-YQSKL in a fluorescence anisotropy competition assay, was determined as 32,400 ± 4800 nM, and the K_i of YQSEL was >100,000 nM (Skoulding, 2011; Skoulding et al., 2015). These K_i values are in comparison to the calculated K_i of YQSKL being 166 ± 23 nM, demonstrating that YQSKV has approximately 195-fold weaker binding to *At*PEX5-C than YQSKL, and YQSEL has over 600-fold weaker binding to *At*PEX5-C than YQSKL. The peptides YQSEL and YQSKV, therefore, were chosen as potential PTS1* sequences in this work.

Along with published *in vivo* and *in vitro* data on the PEX5:PTS1 interaction (reviewed in Smith and Aitchison, 2013, and Lanyon-Hogg et al., 2010), *in silico* work has also been carried out in order to determine what defines a PTS1 sequence (Reumann, 2011; Lingner et al., 2011; Reumann et al., 2012; Chowdhary et al., 2012). The prediction server PredPlantPTS1 was developed (Lingner et al., 2011; Reumann et al., 2012) to allow the prediction of whether a sequence of amino acids will be imported into the peroxisome if it were located at the C-terminus of a protein. The two chosen potential PTS1* sequences were predicted to be non-peroxisomal using the PredPlantPTS1 prediction.

In this chapter, fluorescence anisotropy was used to assess the K_d of fluorescently labelled peptides directly, rather than using unlabelled peptide to compete bound peptide from the binding site of *At*PEX5-C. A peptide that combined the two previous changes to the PTS1 sequence, resulting in lissamine-YQSEV, was also used as this was predicted to be a non-PTS1 sequence (using PredPlantPTS1) and it was

reasoned that this peptide would have an even lower affinity to *At*PEX5-C than either lissamine-YQSEL or lissamine-YQSKV.

YQSEL, YQSKV and YQSEV were synthesised on solid support using Fmoc solid phase peptide synthesis as was performed for YQSKL (section 2.3), and lissamine was then coupled to the N-terminus of each of these peptides. Fluorescence anisotropy was carried out using each of these peptides with *At*PEX5-C in order to determine dissociation constants (**Figure 19**).

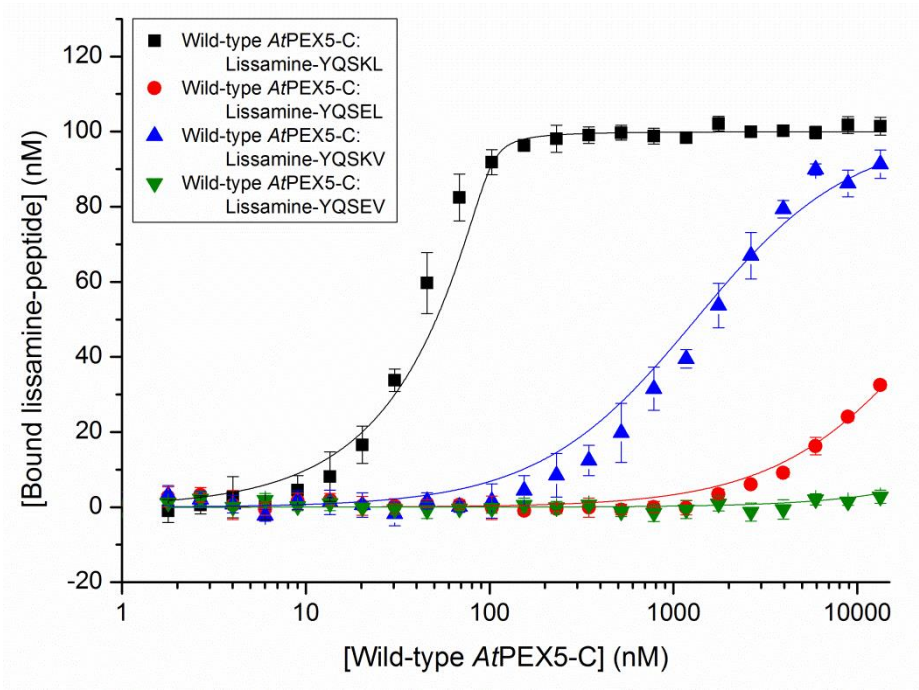


Figure 19| Wild-type *At*PEX5-C (0–13.3 µM final concentration) versus amount of lissamine-YQSKL, -YQSEL, -YQSKV, or -YQSEV bound in the presence of 100 nM (final concentration) peptide, measured using fluorescence anisotropy. Order of peptide binding affinity with *At*PEX5-C, from high to low: lissamine-YQSKL > lissamine-YQSKV > lissamine-YQSEL > lissamine-YQSEV. A quadratic equation was used to fit the curves using OriginPro 9.1, assuming a one-to-one model of binding.

When compared to the representative PTS1 sequence YQSKL, the three potential PTS1* sequences tested behaved as expected. The dissociation constants for these peptides are shown in **Table 6**.

Fluorescent peptide	Dissociation constant (K_d)(nM)	Standard error (nM)
Lissamine-YQSKL	1.1	0.6
Lissamine-YQSEL	29,000	1400
Lissamine-YQSKV	1300	100
Lissamine-YQSEV	>100,000	-

Table 6| Dissociation constants and standard errors for binding of a representative PTS1 sequence and three potential PTS1* sequences to *At*PEX5-C. Dissociation constants were determined using fluorescence anisotropy and show that lissamine-YQSKV displays >1000-fold lower binding affinity than lissamine-YQSKL, lissamine-YQSEL displays >25,000-fold lower binding affinity than lissamine-YQSKL, and lissamine-YQSEV displays >90,000-fold lower binding affinity than lissamine-YQSKL.

The three peptides selected as potential PTS1* sequences were all determined to have much lower binding affinities to *At*PEX5-C than the 'native' PTS1 sequence YQSKL. The potential PTS1* sequence with the highest affinity for the *At*PEX5-C was subject to *in vivo* testing in moss (*Physcomitrella patens*) by Dr Rupesh Paudyal, the postdoctoral researcher working on the *in vivo* aspects of this project (**Figure 20**).

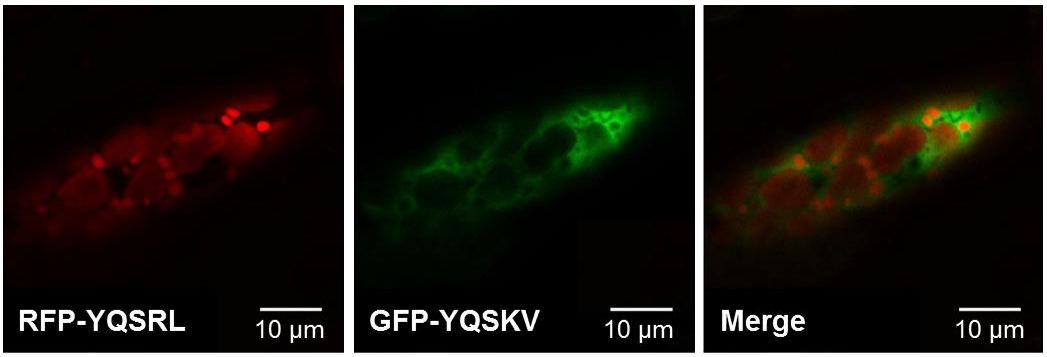


Figure 20| *In vivo* import experiment in *Physcomitrella patens*, with images captured 24 h after particle bombardment of vectors containing genes coding for RFP-YQSRL or GFP-YQSKV. Peroxisomes are shown as the bright regions on the RFP-YQSRL image, and these regions are not highlighted by GFP-YQSKV so it can be concluded that YQSKV as a C-terminal sequence does not induce peroxisomal targeting of GFP. Images courtesy of Rupesh Paudyal.

It can be concluded, as a result of **Figure 20**, that all three potential PTS1* sequences are appropriate sequences to use as potential PTS1*. This means that the evolution of an *At*PEX5-C* protein to bind one of these sequences could begin.

2.5 Variants of *At*PEX5-C designed for altered peptide-binding specificity

Variants of the *At*PEX5-C protein were produced with the aim of enhancing binding of *At*PEX5-C to the potential PTS1* peptides determined in section 2.4. A homology model of *At*PEX5-C had previously been produced at the University of Leeds (Skoulding, 2011) which revealed the similarity between the *Hs*PEX5-C and *At*PEX5-C structures. In this work, the web server I-TASSER was used to obtain a predicted *At*PEX5-C protein structure model in order to model the effects of particular mutations (Zhang, 2008; Roy et al., 2010; Yang et al., 2015; Yang and Zhang, 2015). The I-TASSER model of *At*PEX5-C was aligned with the *Hs*PEX5-C:YQSKL crystal structure determined by the Berg laboratory (Gatto et al., 2000) to determine the approximate position of the PTS1 peptide in the *At*PEX5-C binding site (**Figure 21**).

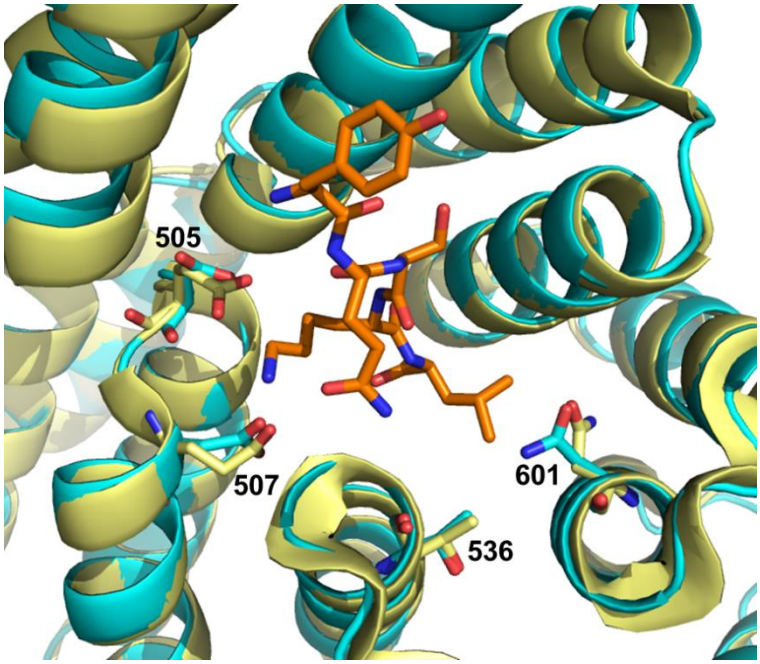


Figure 21| Predicted model of *At*PEX5-C aligned with *Hs*PEX5-C:YQSKL crystal structure. The predicted structural model of *At*PEX5-C (I-TASSER) is shown in cyan, the *Hs*PEX5-C crystal structure is shown in pale yellow and the YQSKL crystal structure is shown in orange (Gatto et al., 2000 (1FCH); Zhang, 2008; Roy et al., 2010; Yang et al., 2015; Yang and Zhang, 2015 (I-TASSER)). Structural overlay was performed in PyMOL by the overlay of α-carbons in the polypeptide chain. Residues 505, 507, 536 and 601 were chosen for mutation due to their proximity to the PTS1 sequence in this structural prediction, and these residues of interest are shown in 'stick' form.

Four residues were initially selected as being potentially important for binding to PTS1 due to their proximity to the PTS1 sequence: D505, D507, T536 and N601 (**Figure 21**). Conservation of these residues across species is displayed in **Figure 22**.

Figure 22| Alignment of relevant portions of PEX5 from a range of organisms to assess conservation of the highlighted residues. Organism names are coloured by kingdom: green text, plantae; blue text, excavata; pink text, animalia; orange text, fungi. Highlighted in yellow are matching residues, highlighted in green are residues with identical charge, and highlighted in cyan are residues with differing electronic or steric properties. Alignment was performed using BioEdit (sequence alignment editor).

Residues D505, D507, T536 and N601 were targeted by site-directed mutagenesis (**Figure 23**) in order to carry out a substitution mutation at each position. The variant E361K in *S. cerevisiae* was identified as binding to a non-PTS1 sequence –SEL> in a yeast-two-hybrid screen by the Distel laboratory (Klein et al., 2001). For this reason, it was decided to make the equivalent variant in *Arabidopsis* (D505K), along with another charge substitution mutation, D505H, in the hope that these *At*PEX5-C variants would bind to lissamine-YQSEL. D505R was not produced as it was decided that the arginine side chain would extend too far into the *At*PEX5-C binding pocket to allow for a PTS1 or PTS1* sequence to fit into the binding site. The double variant D505K-D507K was also produced, as residue 507 forms part of the acidic pocket surrounding the lysine residue of the PTS1, as seen by Gatto and colleagues (Gatto et al., 2000). The side chains of T536 and N601 lie in close proximity to the side chain of the C-terminal PTS1 amino acid, so it was decided to extend each of these side chains separately into the binding pocket, which would make the binding pocket smaller and

possibly encourage preferential acceptance of valine over leucine as the C-terminal PTS1 residue as a result of steric bulk. This could be achieved by making the variants T536N and N601Q.

Site-directed mutagenesis was used to produce selected variants of *At*PEX5-C. In order to create mutations in the *At*PEX5-C gene, overlapping primers were used in which the appropriate codon in the genetic sequence was substituted for a codon which would code for the desired amino acid. A QuikChange kit was used, along with linear PCR to produce all selected variants (**Figure 23**) (Chapter 6 (Experimental) section 6.5).

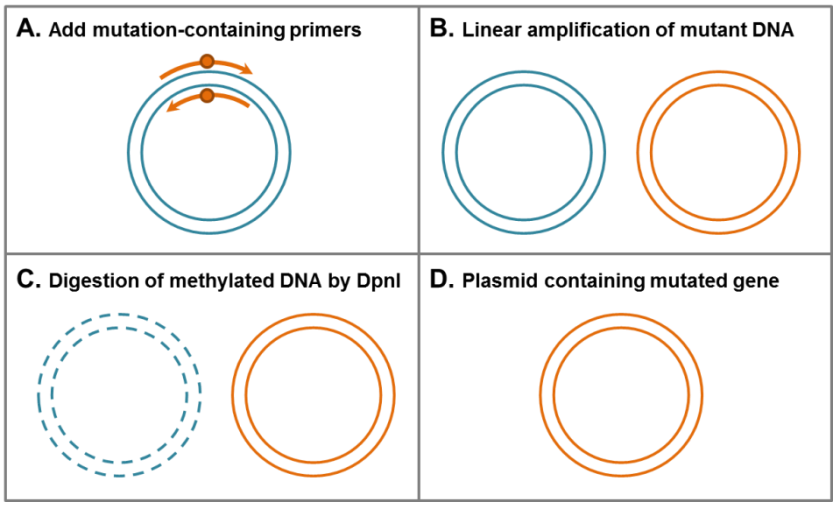


Figure 23| Schematic of site-directed mutagenesis. Overlapping primers are used, which contain the codon (which will insert the desired amino acid) at approximately the middle of each primer. Linear amplification of the entire plasmid is performed, followed by digestion of the parent DNA, which has been methylated by *E. coli* (plasmid containing the wild-type gene).

Selected *At*PEX5-C variant proteins were expressed and purified, and electrospray ionisation mass spectrometry was used to confirm masses of these *At*PEX5-C variants (**Figure 24**).

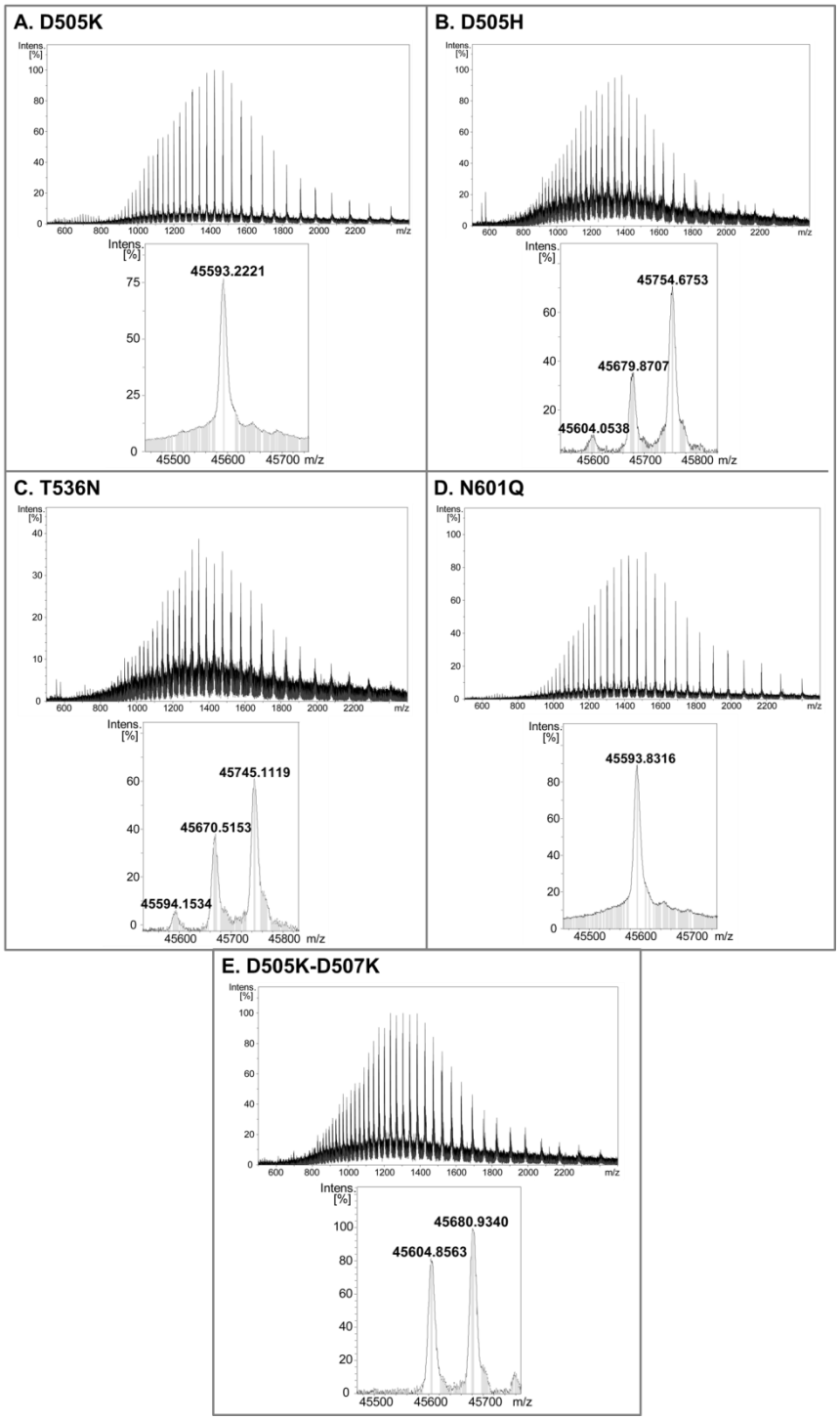


Figure 24| Mass spectra of *At*PEX5-C variant proteins. The deconvoluted masses are shown underneath the broad peak spectrum for each species. A) Expected mass for D505K: 45,593.4. Observed mass for D505K: 45,593.2. B) Expected mass for D505H: 45,602.4. Observed mass for D505H: 45,604.1. C) Expected mass for T536N: 45,593.3. Observed mass for T536N: 45,594.2. D) Expected mass for N601Q: 45,594.3. Observed mass for N601Q: 45,593.8. E) Expected mass for D505K-D507K: 45,606.5. Observed mass for D505K-D507K: 45,604.9. In spectra B, D, and E, the mass + approximately 76 g/mol, and the mass + approximately (2*76 g/mol) is observed. These peaks correspond to the mass of the protein + 2-mercaptoethanol. *At*PEX5-C has two surface-exposed cysteine residues (not near the PTS1-binding site), which are available to react with 2-mercaptoethanol and TCEP was used to reduce these disulfide bonds with varying degrees of success.

Circular dichroism was performed on the five variants of *At*PEX5-C in comparison to wild-type *At*PEX5-C to determine whether any major changes to secondary structure had occurred as a result of each of the mutations (**Figure 25**). As the C-terminal domain of *At*PEX5-C is predicted to be almost entirely α-helical, it was expected that troughs in the CD profiles at 208 nm and 222 nm would be seen.

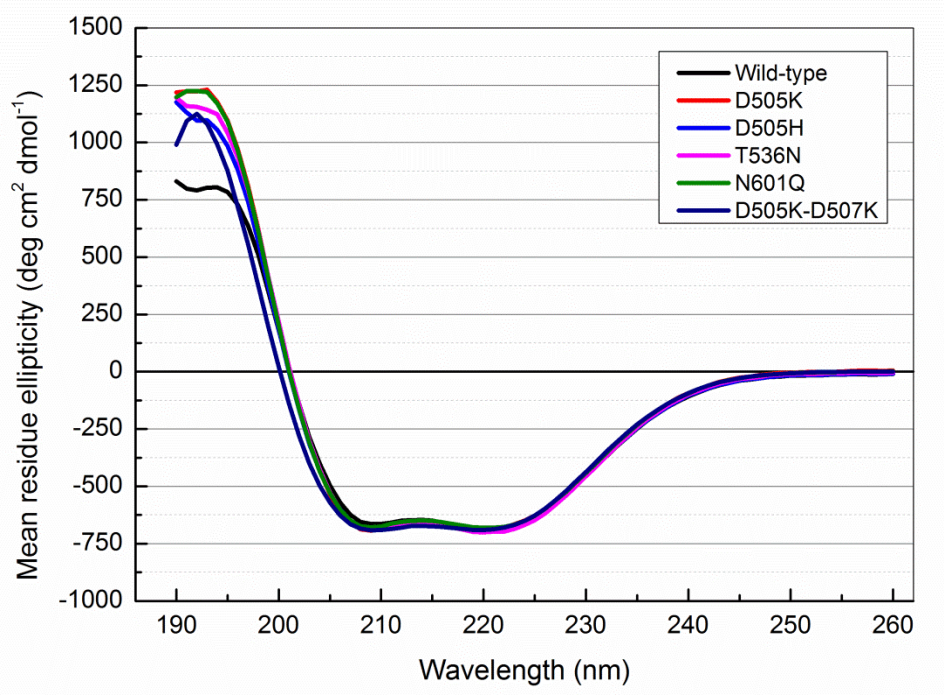


Figure 25| Circular dichroism profiles of five *At*PEX5-C variants produced in comparison to that of wild-type *At*PEX5-C. All variant proteins have a similar secondary structure to wild-type *At*PEX5-C. Troughs are apparent at 208 nm and 222 nm, indicating that all proteins adopt an α-helical secondary structure.

The variants of *At*PEX5-C produced showed almost identical CD spectra to wild-type *At*PEX5-C, suggesting that the mutations did not cause misfolding of the protein at a secondary structure level. Fluorescence anisotropy was therefore carried out on these five variants of the protein with the chosen potential PTS1* peptides (chosen in section 2.4). If enhanced binding of the *At*PEX5-C variants was seen with the potential PTS1* peptides YQSEL or YQSKV, the aim would then be to combine mutations in the protein in order to promote binding to YQSEV.

Fluorescence anisotropy results are plotted so that the binding of the 'native' PTS1 YQSKL and the potential PTS1* peptide are shown with wild-type *At*PEX5-C and the appropriate variant of *At*PEX5-C on the same plot (**Figure 26**). In the following graphs, with wild-type *At*PEX5-C, lissamine-YQSKL is shown in black squares, lissamine-

YQSEL is shown in red circles, and lissamine-YQSKV is shown in blue triangles. With *At*PEX5-C variants, lissamine-YQSKL is shown in green diamonds, lissamine-YQSEL is shown in purple stars, and lissamine-YQSKV is shown in orange triangles.

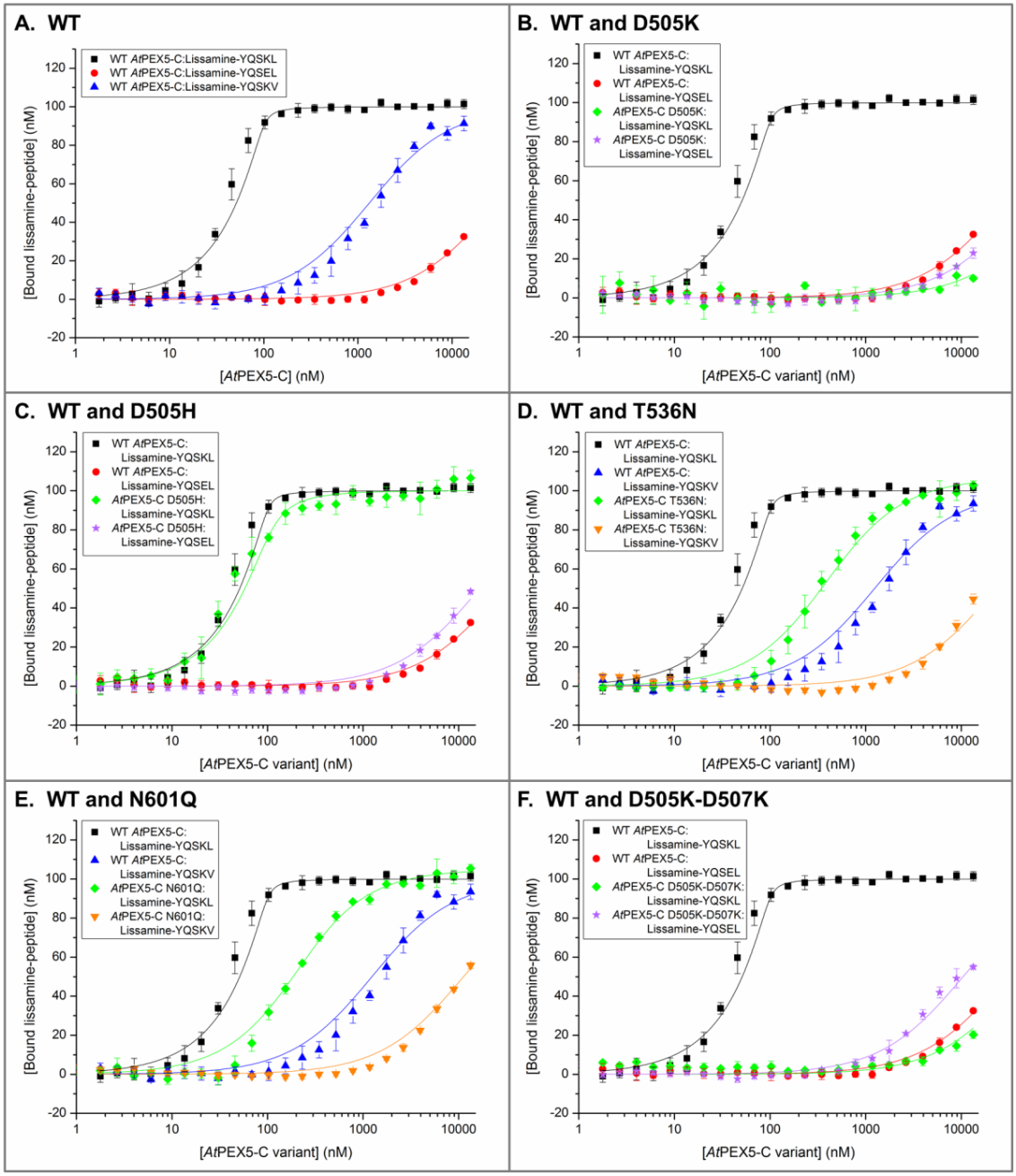


Figure 26| Wild-type or variant *At*PEX5-C (ranging from 0–13.3 µM final concentration) versus amount of lissamine-YQSKL, -YQSEL, or -YQSKV bound in the presence of 100 nM peptide, measured using fluorescence anisotropy. Binding curves for wild-type *At*PEX5-C with lissamine-YQSKL are shown in black squares, those for wild-type *At*PEX5-C with lissamine-YQSEL are shown in red circles, and those for wild-type *At*PEX5-C with lissamine-YQSKV are shown in blue triangles. Binding curves for variants of *At*PEX5-C with lissamine-YQSKL are shown in green diamonds, those for variants of *At*PEX5-C with lissamine-YQSEL are shown in purple stars, and those for variants of *At*PEX5-C with lissamine-YQSKV are shown in orange downward triangles. A quadratic equation was used to fit the curves using OriginPro 9.1, assuming a one-to-one model of binding.

The *At*PEX5-C variant D505K did not appear to bind either the 'native' PTS1 peptide YQSKL or the potential PTS1* peptide YQSEL with any notable affinity (**Figure 26, B**). This was in contrast to results from the Distel laboratory, as work by the Distel laboratory had demonstrated binding of this variant of the *Sc*PEX5 protein to a C-terminal –SEL> sequence (Klein et al., 2001). However, it may be that there is a distinct difference between the binding interactions for *S. cerevisiae* PEX5:PTS1 versus *A. thaliana* PEX5:PTS1 (Lametschwandtner et al., 1998; Kragler et al., 1998), or the interaction seen by the Distel laboratory could have been a false positive from the yeast-two-hybrid assay.

The D505H variant of *At*PEX5-C does not show a great deal of variance from wild-type *At*PEX5-C in terms of its peptide binding profile (**Figure 26, C**). In proteins, the pK_a of the imidazole of the histidine side chain is approximately 6.5–7.0 (Cantor and Schimmel, 1980). As the fluorescence anisotropy assay was carried out at pH 7.5, this meant that approximately 15% of the histidine side chains would have been protonated under these conditions. The H505 side chain, therefore, predominantly removed charge from the acidic pocket in the PTS1-binding site. It appeared that this removal of charge had a very slight effect on binding specificity but the binding of the protein to YQSEL was still much weaker than the binding to YQSKL.

The mutation T536N did result in the binding affinity to the two peptides being changed; however, this change appeared to shift the binding curves of both peptides to the right (**Figure 26, D**). It appears that the effects of this mutation are just to impair overall PTS1 binding. *At*PEX5-C variant N601Q showed a similar peptide binding profile to T536N (**Figure 26, E**). The overall PTS1 peptide binding affinity, however, was slightly higher with variant N601Q than with T536N.

The variant of *At*PEX5-C with a double mutation, D505K-D507K, shows the desired pattern of peptide binding specificity. This *At*PEX5-C variant results in the impairment of PTS1 binding affinity, while the binding affinity for a non-PTS1 sequence is increased (**Figure 26, F**). Peroxisomal import of a protein possessing a particular C-terminal PTS1 sequence can be predicted based on the affinity for that PTS1 sequence to PEX5 *in vitro* (the K_d for the interaction) (Skoulding et al., 2015). The affinity of the *At*PEX5-C D505K-D507K variant to lissamine-YQSEL appears to fall far short of the predicted affinity threshold for strong peroxisomal protein import *in vivo* (~500 nM) (Maynard and Berg, 2007; Skoulding et al., 2015). YQSKV is a non-PTS1 sequence as determined by *in vivo* study of peroxisomal protein import (**Figure 20**). In

this work, the calculated K_d for the interaction of wild-type *At*PEX5-C with lissamine-YQSKV was 1300 ± 100 nM. The calculated K_d for the interaction of *At*PEX5-C D505K-D507K with lissamine-YQSEL was 10,600 ± 400 nM (approximately 8-fold weaker affinity than wild-type *At*PEX5-C with lissamine-YQSKV) (**Table 7**). It can therefore be predicted that this interaction will not lead to peroxisomal protein import *in vivo*.

***At*PEX5-C variant**	**Fluorescent peptide**	**Dissociation constant (K_d)(nM)**	**Standard error (nM)**
Wild-type	Lissamine-YQSKL	1.1	0.6
	Lissamine-YQSEL	29,000	1000
	Lissamine-YQSKV	1300	100
D505K	Lissamine-YQSKL	96,000	12,000
	Lissamine-YQSEL	46,600	5200
D505H	Lissamine-YQSKL	7.8	1.9
	Lissamine-YQSEL	16,500	1300
T536N	Lissamine-YQSKL	330	40
	Lissamine-YQSKV	23,000	5200
N601Q	Lissamine-YQSKL	150	10
	Lissamine-YQSKV	11,200	800
D505K-D507K	Lissamine-YQSKL	42,900	5300
	Lissamine-YQSEL	10,600	400

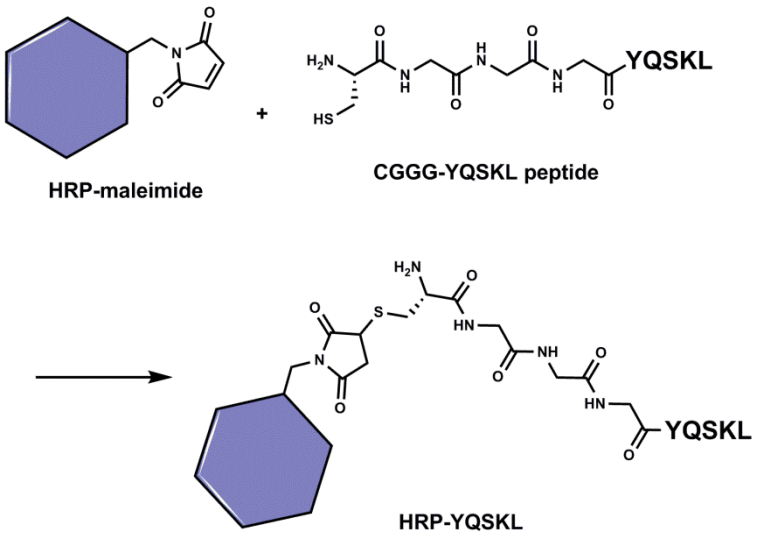
Table 7| Summary of dissociation constants and standard errors for binding of a representative PTS1 sequence and two potential PTS1* sequences to *At*PEX5-C and variants. Dissociation constants were determined using fluorescence anisotropy.

Through the generation of five *At*PEX5-C variants, it was possible to alter peptide binding specificity and slightly enhance binding to a new, non-PTS1, peptide sequence (lissamine-YQSEL). Unfortunately this interaction was predicted to be much too weak to allow import of this sequence into peroxisomes *in vivo*, so it became apparent that this interaction could not easily be changed by rational design through binding predictions. For this reason, it was necessary to develop a screen by which a larger number of *At*PEX5-C variants could be studied for peptide binding properties.

2.6 Screen development for the study of PEX5:PTS1 binding by colony blotting

As the initial rational design experiments had not resulted in an orthogonal pair, it was decided that a higher throughput screen would be developed in order to allow coverage of a larger number of protein-peptide combinations. Fluorescence anisotropy was initially used as a screen idea; however, it was not possible to express and purify each *At*PEX5-C variant from a library on a small enough scale to allow for differentiation between positive and negative controls. Colony blotting was chosen as the technique to develop into a screen, as it had previously been shown by Tom Lanyon-Hogg (former student of the University of Leeds) that horseradish peroxidase (HRP) conjugated to the peptide YQSKL (HRP-YQSKL) binds to *At*PEX5-C in a dot blot where HRP-YQSEL does not (Lanyon-Hogg, 2012).

In order to test whether colony blotting would be a viable method for identifying *At*PEX5-C variants that bind to a particular peptide, an initial dot blotting experiment was carried out using purified protein and HRP-conjugated peptides (**Figure 27**). Each HRP-conjugated peptide was produced by the addition of a Cys-Gly-Gly-Gly peptide linker at the N-terminus of the pentapeptide of interest, and reaction of the resulting peptide with HRP-maleimide (**Scheme 2**). Maleimide reacts covalently with the free thiol group of cysteine, so this reaction will leave HRP-CGGG-pentapeptide. HRP-CGGG-peptide, from this point onwards, will be referred to as HRP-peptide.



Scheme 2| HRP-maleimide coupling to peptide with an N-terminal cysteine. Here, the peptide CGGGYQSKL is shown as an example. Peptides were purified under reducing conditions and the conjugation reaction with HRP-maleimide was carried out in the presence of TCEP to prevent the formation of disulfide bonds between peptides.

All of the potential PTS1* peptides tested by fluorescence anisotropy in section 2.4 (lissamine-YQSEL, lissamine-YQSKV, and lissamine-YQSEV) were conjugated to HRP for use in colony blot screening. Reactions were carried out using an excess of peptide, and unreacted peptide was dialysed out of the reaction (and monitored by the disappearance of the UV trace corresponding to peptide) after capping of any unreacted maleimide sites using 2-mercaptoethanol.

Dot blots were initially used to test whether colony blotting was viable as a screening concept. These dot blots involved the dotting of purified protein onto nitrocellulose membrane, and the blotting of this membrane with HRP linked to peptide. If the peptide bound to the protein in question, a bright dot would be observed following luminol-based HRP detection. In the optimisation of colony blotting, it was decided that the positive control would be wild-type *At*PEX5-C with HRP-YQSKL, as the wild-type *At*PEX5-C:Lissamine-YQSKL interaction had a K_d of 1.1 ± 0.6 nM (section 2.3). The negative peptide control was decided as HRP-YQSEL, as the lissamine-YQSEL peptide shows very weak binding affinity with either wild-type *At*PEX5-C (section 2.4) or the variant D505K (section 2.5). The negative protein control would be the *At*PEX5-C variant D505K, which was found in section 2.5 not to bind to lissamine-YQSKL or to lissamine-YQSEL.

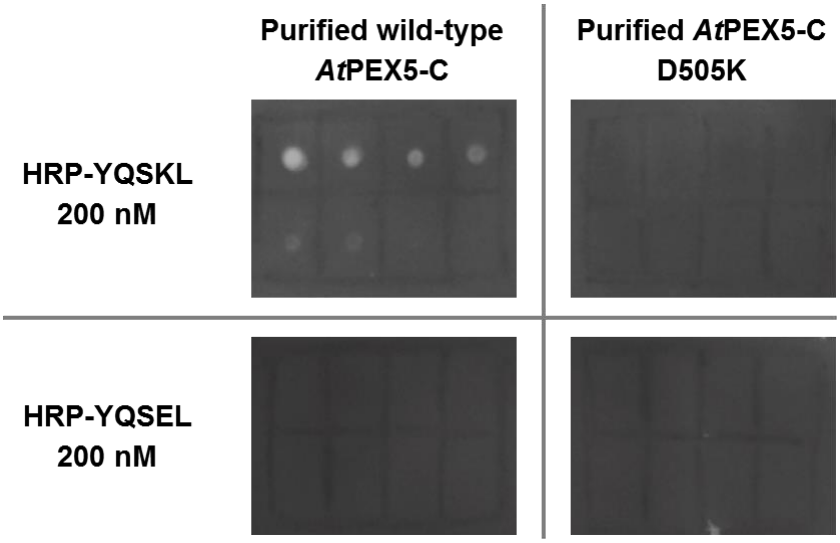


Figure 27| Dot blot of two *At*PEX5-C variants with two different peptides conjugated to HRP. HRP-YQSKL binds selectively to wild-type *At*PEX5-C over a non-PTS1 binding variant of the protein. The amount of purified protein used was spotted on nitrocellulose membrane in serial dilutions ranging from 500 ng to 3.91 ng (7 serial dilutions; 8 dots). The presence of HRP was visualised by luminol-based detection.

This dot blotting experiment proved that an HRP-conjugated peptide could selectively identify a peptide-binding variant of the *At*PEX5-C protein. The positive control of wild-type *At*PEX5-C with HRP-YQSKL showed evidence of binding whereas negative controls, wild-type *At*PEX5-C with HRP-YQSEL, and *At*PEX5-C D505K with HRP-YQSKL and HRP-YQSEL, did not appear to show binding. The next stage was to test whether this selectivity could still be seen when blotting colonies as opposed to purified protein. In order to test this, a published method of colony blotting was used (The QIAexpressionist, 2001). This protocol involves the transfer of transformed cells onto nitrocellulose membrane and subsequent expression of genes by incubation of this nitrocellulose membrane on solid selection media containing IPTG. Cells on the nitrocellulose membrane are then lysed by transfer of the nitrocellulose onto filter paper soaked with a range of buffers (see Chapter 6 (Experimental) section 6.12.4), and blotting is carried out after a blocking step, as per western or dot blotting. In the initial experiments for optimising colony blotting as a screen, protein controls with different known binding affinities to YQSKL were used and all colonies were tested with HRP-YQSKL (**Figure 28** and **Figure 29**).

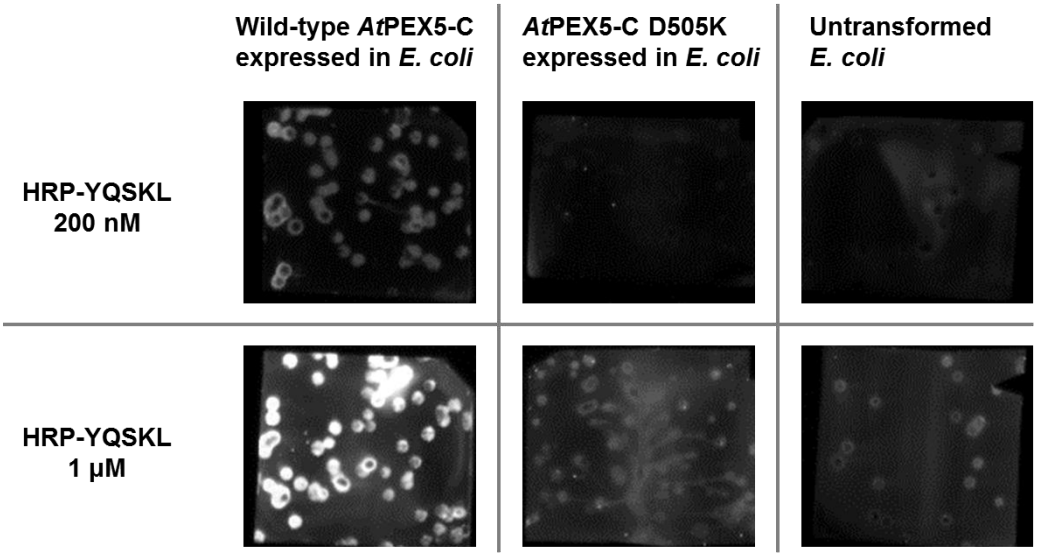


Figure 28| Colony blot of induced BL21 (DE3) cells containing either wild-type *At*PEX5-C, or *At*PEX5-C D505K compared to untransformed induced BL21 (DE3) cells. HRP-YQSKL appears to bind more strongly to cells containing wild-type *At*PEX5-C, with low background seen for cells containing *At*PEX5-C D505K and for untransformed cells. The presence of HRP was visualised by luminol-based detection. Two different concentrations of HRP-YQSKL were used (200 nM or 1 µM) for blotting.

As can be seen in **Figure 28**, cells containing the wild-type variant of *At*PEX5-C can easily be identified as those expressing a protein that binds to HRP-YQSKL. The difference between 'binder' and 'non-binder' is clearer when using a higher concentration of the HRP-conjugated peptide, so a final concentration of 350 nM for HRP-conjugated peptides was used (**Figure 29**). Further optimisation of the colony blotting protocol described above was required in order to use this technique as a screen. Optimisations included a lower percentage of milk used in blocking buffer, along with a longer blocking step, and a crucial optimisation was the use of a colony picker to perform re-arraying and gridding of cells for much easier identification of specific colonies (method described in Chapter 6 (Experimental) section 6.12.4). **Figure 29** shows the result of this optimisation.

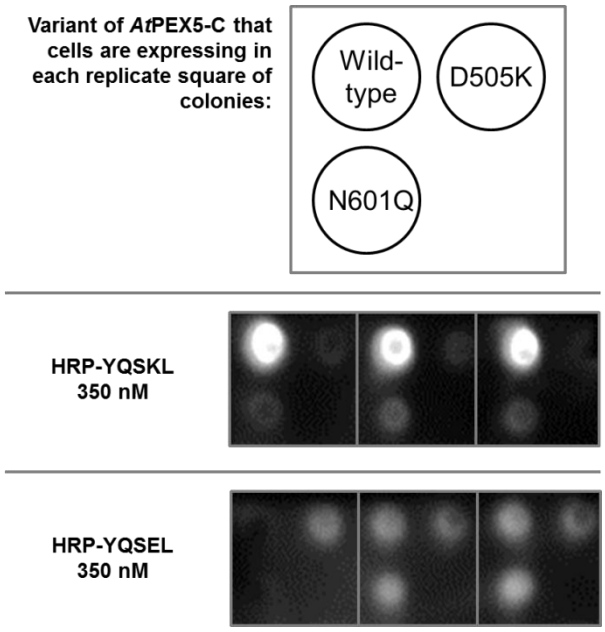


Figure 29| Colony blot optimisation using variants of *At*PEX5-C and different HRP-conjugated peptides. Labelled in the figure are protein variants expressed within each colony tested. These colonies were tested in replicate by incubation with either HRP-YQSKL or HRP-YQSEL. The presence of HRP was visualised by luminol-based detection.

The protein variants tested in **Figure 29** were chosen as they displayed a range of dissociation constants when binding to YQSKL (section 2.5). Wild-type *At*PEX5-C had a dissociation constant of 1.1 ± 0.6 nM, *At*PEX5-C N601Q had a dissociation constant of 150 ± 10 nM, and *At*PEX5-C D505K had a dissociation constant of 96,000 ± 12,000 nM. This allowed testing of the sensitivity of the assay. Although slight detection of colonies expressing *At*PEX5-C N601Q using HRP-YQSKL was seen, the background levels of staining by HRP-YQSEL surpassed this. This result was not seen as a problem as it could mean that only the strongest peptide-binding proteins would be identified. This method provided a way of selectively identifying colonies which contained protein able to bind strongly to a maleimide-peptide. The colony blotting screen could then be taken forward to screen a library of *At*PEX5-C variants.

The aim of developing a screen was to enable the study of a library of *At*PEX5-C mutants. It was decided to create randomised libraries of the proteins rather than the peptides, as this would give a greater level of variation in the libraries due to protein folding possibilities. Saturation mutagenesis was employed in order to produce protein variant libraries. Saturation mutagenesis works in the same way as site-directed mutagenesis (**Figure 23**); however, a larger number of mutagenesis primers are used, in which the codon in each mutagenesis primer coding for the desired amino acid is

replaced with a degenerate codon. This means that various different specified codons can be inserted (in the manufacturing of the primer library) in place of the original codon at that position in the DNA sequence (**Figure 31**). Multiple degenerate codons can be included in the same mutagenesis primer, providing the resulting residues are close to one another in the primary sequence of the protein.

Four positions in the protein were initially chosen for saturation mutagenesis: D505, D507, V533 and T536. Position D505 was chosen because previous research into mutation of the equivalent residue in *S. cerevisiae* PEX5 had shown that it is possible for this variant of the protein to bind to a non-PTS1 sequence, YQSEL (Klein et al., 2001). Positions D507, V533 and T536 in the protein were chosen due to apparent proximity to the binding peptide in existing crystal structures (Gatto et al., 2000; Stanley et al., 2006; Stanley et al., 2007; Sampathkumar et al., 2008; Fodor et al., 2012). If these four residues were targeted in the same library and replaced with any of the 20 amino acids, a library of 160,000 protein variant combinations would be produced (including all wild-type possibilities). A reduction in library size was possible by splitting the four residues into two libraries with 200 protein variant combinations each. If hits were found in each of these saturation libraries, the mutations from both libraries could be combined, which may improve binding to PTS1*. This approach is known as CASTing (Combinatorial Active site Saturation Testing) (Reetz et al., 2006). Library 1 was varied at positions D505 and D507, and library 2 was varied at positions V533 and T536 (**Figure 30**). It was decided that library 1 could be produced to allow variation within the acidic pocket of the PTS1-binding site, and library 2 could be produced to allow size variation at the hydrophobic region accommodating the C-terminal PTS1 residue.

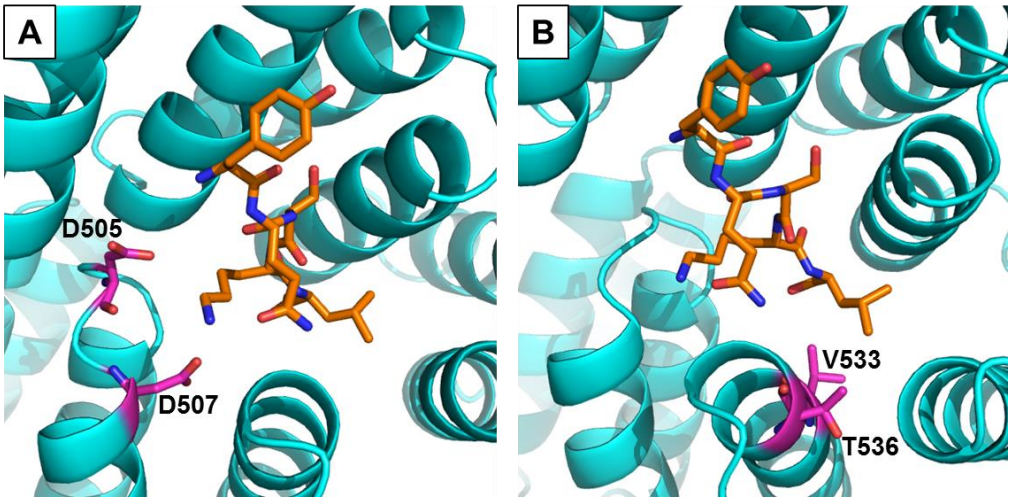


Figure 30| Models of the *At*PEX5-C protein variant libraries 1 and 2. Library 1 (A) contains two acidic residues surrounding the basic side chain of the -2 amino acid of the representative PTS1 peptide. Library 2 (B) contains two residues which appear to form part of the hydrophobic pocket for interaction with the hydrophobic side chain of the -1 (C-terminal) amino acid of the representative PTS1 peptide (Gatto et al., 2000 (1FCH); Zhang, 2008; Roy et al., 2010; Yang et al., 2015; Yang and Zhang, 2015 (I-TASSER)).

In order to produce these protein variant libraries, degenerate codons were used in the mutagenesis primers. Saturation mutagenesis primers used for making these libraries are shown in Appendix B. Three different degenerate codons were selected in order to give an appropriate level of variation in the library whilst keeping library sizes manageable. This reduced library sizes further, from 200 protein variant combinations in each library (if any of the 20 amino acids were used) to less than 90 protein variant combinations in each.

Positions 505 and 507 in the protein (targeted in library 1) form an acidic region by which *At*PEX5-C has been found to interact with a basic residue in the PTS1 sequence, YQSKL (Gatto et al., 2000; Stanley et al., 2006; Stanley et al., 2007; Sampathkumar et al., 2008; Fodor et al., 2012). With one of the potential PTS1* sequences being YQSEL, it was decided to include all possible basic residues at both positions in library 1. For this reason, ‘VRN’ was selected as the degenerate codon to use at both positions. ‘V’ means either C, A or G as a base in that position, ‘R’ means either A or G, and ‘N’ means any of the bases can be placed at that position. Using the codon ‘VRN’ gave the possibility of nine amino acids, Gln, His, Arg, Lys, Asn, Ser, Glu, Asp, or Gly, occurring at positions 505 or 507 in the resulting protein.

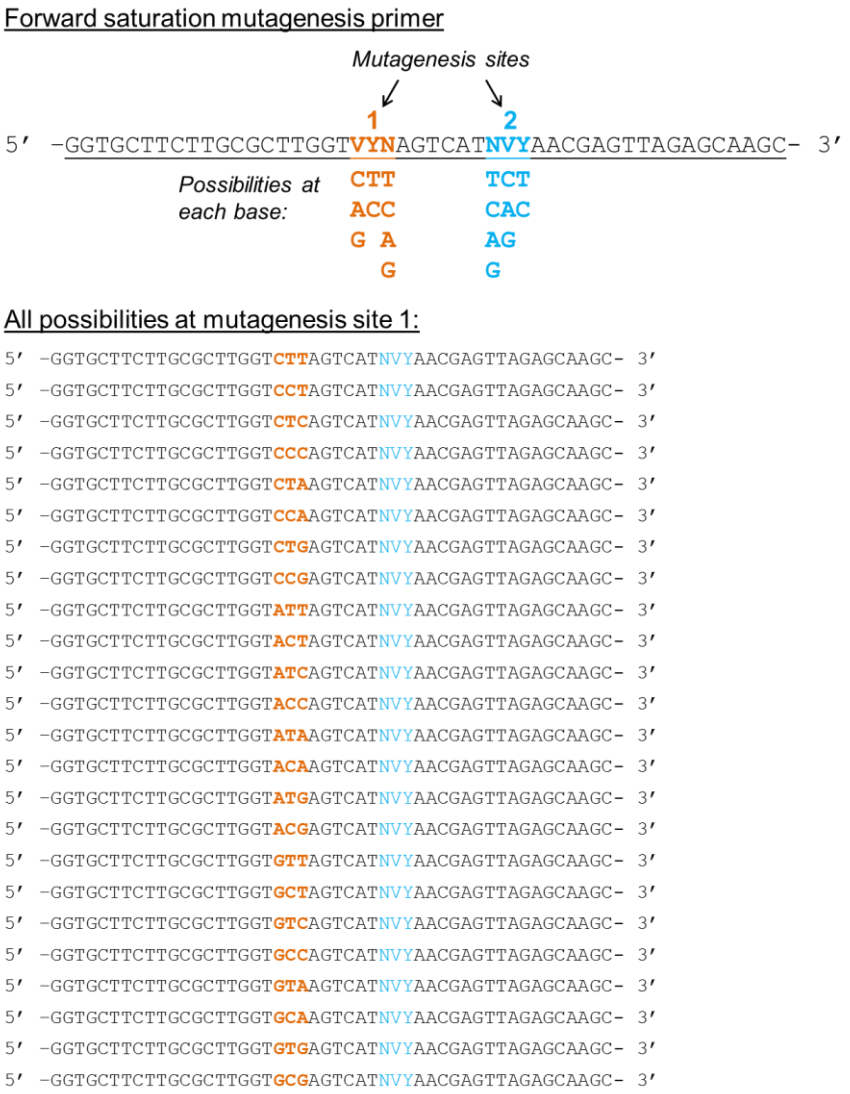


Figure 31| All possible codons produced from a degenerate codon in a saturation mutagenesis primer (to generate library 2). Here, the forward primer for the generation of protein variant library 2 is shown as an example.

From study of existing crystal structures of PEX5-C:PTS1, positions 533 and 536 both appear to be part of the hydrophobic region by which the C-terminal PTS1 residue can interact (Gatto et al., 2000; Stanley et al., 2006; Stanley et al., 2007; Sampathkumar et al., 2008; Fodor et al., 2012). For V533, the intention was to replace this with any non-polar hydrophobic amino acid. The degenerate codon 'VYN' was therefore used at this position in the mutagenesis primer (**Figure 31**). 'Y' means either T or C as a base in that position. Using the codon 'VYN' gave the possibility of seven amino acids, Leu, Ile, Met, Val, Pro, Thr, or Ala, occurring at position 533 in the resulting protein. For Thr536, it was decided that largely uncharged hydrophilic or hydrophobic residues would be beneficial at this position in the protein, so the degenerate codon 'NVY' was used at this position. Using the codon 'NVY' gave the possibility of eleven amino acids,

Gly, Ser, Arg, Cys, Asp, Asn, His, Tyr, Ala, Thr, or Pro, occurring at position 536 in the resulting protein.

In the resulting protein libraries, library 1 should have contained 81 possibilities, and library 2 should have contained 77 possibilities. As library 1 was the largest, a number of colonies were taken from a transformation, grown in selection media, and cells sent for DNA sequencing in a 96-well plate to assess coverage of the library (Beckman Coulter Genomics). A good sequencing read was obtained for 88 of the 96 colonies, and mutations found in these 88 colonies are shown in **Figure 32**.

Position 505

Position 507	Arg	Asn	Asp	Gln	Glu	Gly	His	Lys	Ser
Arg	7		1	5	3	2	2	2	
Asn	2							1	
Asp	1		4		2	2	3		
Gln	3			1		2	3		1
Glu	2		1		1	3	2		
Gly	2	2			2	4		1	1
His	3	1			2		2		
Lys	1				1			1	
Ser	2			1	1	3		1	1

Figure 32| Sequencing results from 88 colonies of saturation library 1. Some combinations appear to be favoured over others, but overall there is a good distribution of mutations.

As 43 unique mutants of the *At*PEX5-C protein were identified in only 88 colonies, the coverage achieved was 53.1% ((43/81)*100). Using the web server GLUE-IT, it is possible to calculate the expected coverage of a library obtained using variable codons (Firth and Patrick, 2008 (GLUE-IT)). For library 1, two variable codons 'VRN' were used. With a library size of 88, the expected library coverage, as calculated by GLUE-IT, was 59.3% (48.04 expected distinct amino acid variants). This predicted coverage is 6.2% higher than the actual coverage obtained so library sizes were increased from predictions made by GLUE-IT in order to allow for sufficient coverage.

The predicted size of the library to screen which would be sufficient to achieve 95% coverage of the library was calculated as at least 362 colonies for library 1, and at least 341 colonies for library 2 (predicted by GLUE-IT) (Firth and Patrick, 2008 (GLUE-

IT)). Screens were performed using 384-well plates, with 4 wells used for a control region (**Figure 33**), so 380 colonies were screened per 384-well plate. It was decided that sufficient coverage of each of the libraries could be obtained by screening two 384-well plates (760 colonies) for each library, as the resulting expected coverage of library 1 was 99.7% and expected coverage of library 2 was 99.1% (Firth and Patrick, 2008 (GLUE-IT)). Cells from each of the colonies from the library were grown in 384-well plates, and were then transferred to nitrocellulose membrane, resulting in roughly uniform colony growth in grid format on nitrocellulose membrane. The overall screening method is outlined in Chapter 6 (Experimental) section 6.12.4. A brief outline of the colony blotting method developed, along with results for a small number of colonies from library 1 screened with HRP-YQSKL (including the control region) are shown in **Figure 33**.

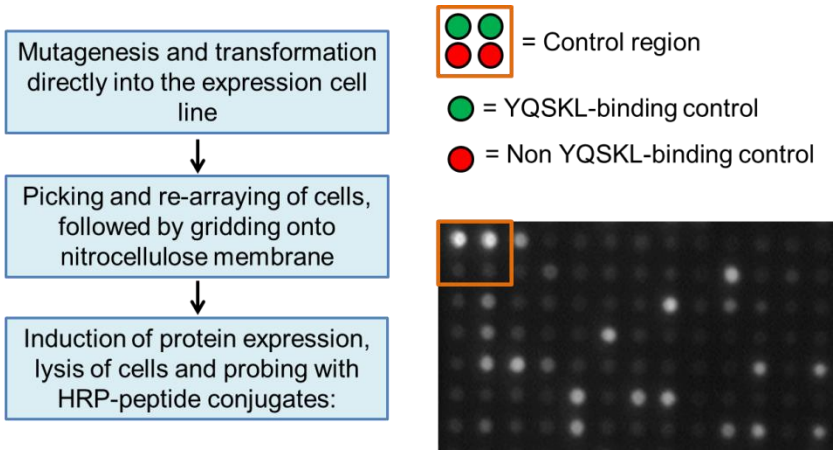


Figure 33| An outline of the colony blotting method and a screen result when testing with HRP-YQSKL. The control region shows that 'YQSKL-binding control' colonies expressing wild-type *At*PEX5-C bind to HRP-YQSKL, whereas 'non YQSKL-binding control' colonies expressing *At*PEX5-C D505K do not.

The colony blotting method successfully identified proteins able to bind the HRP-conjugated peptide. Sequencing of a number of 'positive' colonies from this screen revealed that these colonies contained any combination of the two acidic amino acids at positions 505 and 507. With this in mind, colony blot screening was performed using all potential PTS1* sequences (determined in section 2.4), conjugated to HRP. A selection of results from this screening is shown in **Figure 34**.

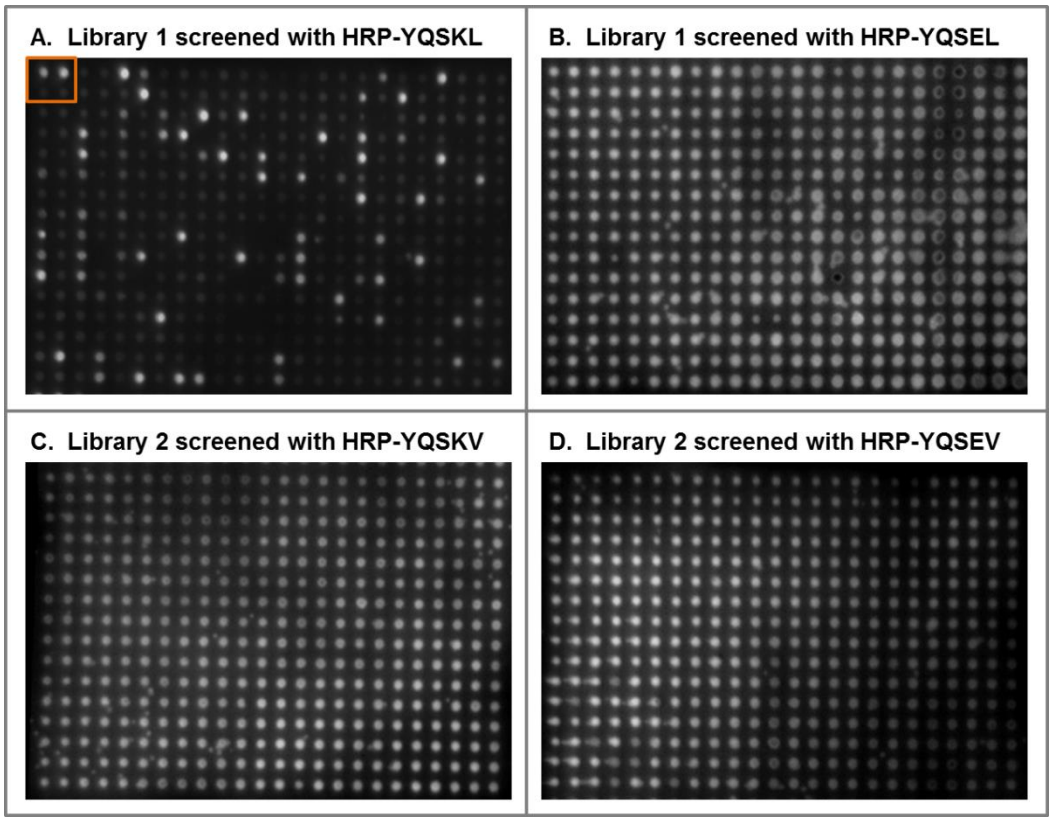


Figure 34| A selection of colony blotting screen results. In the example shown of library 1 screened with HRP-YQSKL, specific colonies are identified by HRP-YQSKL. In the other three examples, where HRP-conjugated to any of the potential PTS1* peptides was used, all colonies appear to be identified by the HRP-peptide. The control region of colonies is highlighted by an orange outline for the screen treated with HRP-YQSKL.

The control region of four colonies established and shown in **Figure 33** are located in the top left-hand corner of each of the screen experiments shown in **Figure 34** (highlighted in **Figure 34, A**). This control region only applied for membranes treated with HRP-YQSKL as it was not possible to establish variants of *At*PEX5-C that could bind to the other three HRP-peptide conjugates before the screening process. The results for 384 colonies are shown for each screen (2 x 384-colony screens were performed for each, so 760 colonies minus the control regions). Libraries 1 and 2 were screened with all HRP-conjugated peptides. Results of colony blotting with HRP-YQSKL (**Figure 34, A**) show that selective identification of a small number of colonies from the 380 colonies shown was possible, demonstrating that the screen worked well when blotting with HRP-YQSKL. Sequencing also confirmed that these colonies identified were all wild-type revertants (or similar) of the *At*PEX5-C protein.

Results of the colony blotting therefore showed that the screen was a viable method for identifying protein variants able to bind to HRP-YQSKL. With all potential PTS1* peptides used, however, the level of background binding was much too high to identify

any true binding proteins. This could be due to the potential PTS1* peptides having some form of binding with a component of the cell lysate or with the nitrocellulose membrane itself. Further screen optimisation in the time available did not result in any improvement on the level of high background binding seen with these peptides. The technique was also time consuming and unreliable, as it was not possible to achieve completely uniform growth of colonies and some colonies would become detached from the nitrocellulose membrane during the blotting process. Furthermore, it had not been demonstrated that this technique could be used to identify intermediate peptide-binding proteins. For these reasons, colony blotting was abandoned as a method of screening.

2.7 Summary

In order to examine the binding properties of the C-terminal TPR domain of *At*PEX5 (*At*PEX5-C), this portion of the *At*PEX5 protein was expressed in *E. coli* and purified. A peptide representing a native PTS1 sequence was synthesised and coupled to lissamine, which enabled the dissociation constant (K_d) of the interaction to be determined using fluorescence anisotropy. This K_d was consistent with published findings. Three potential PTS1* peptide sequences were synthesised for use in fluorescence anisotropy with variants of *At*PEX5-C. The five variants of *At*PEX5-C were produced by site-directed mutagenesis at selected residues within the protein, based on published crystal structures of homologous proteins. The K_d values obtained from these variants with potential PTS1* sequences are shown in **Table 7**. This allowed comparison of K_d values with *in vivo* targeting efficiency of these sequences as determined by Dr Rupesh Paudyal (the postdoctoral researcher working on the *in vivo* aspects of this project).

It became apparent that a higher throughput method than rational design would be needed to identify any *At*PEX5-C* proteins so screening by colony blotting was developed and optimised. This screen successfully identified variants of the protein able to bind to a representative PTS1 sequence conjugated to HRP. Colony blotting was, however, an unsuitable technique for use with our potential PTS1* peptide sequences so an alternative method of screening was required.

Chapter 3

Development of a Pull-Down Method of Screening *At*PEX5-C for Binding to Peptides within a Library

3.1 Introduction

Previous attempts to find an orthogonal PEX5*:PTS1* pair through screening a large library of protein variants against a small library of peptides were unsuccessful (Chapter 2, section 2.6). A new approach was needed, so the procedure was reversed: a large library of peptides would be used to try to identify binding interactions against a small library of protein variants. In order to do this, it would have to be possible to co-isolate *At*PEX5-C variants with their binding peptides. An example of successful published co-isolation of the PEX5-C:PTS1 complex is the pull-down of lissamine-YQSKL using *A. thaliana* PEX5-C in the presence of nickel-agarose resin (Lanyon-Hogg et al., 2014).

The idea of this pull-down method of screening was that if a library of pentapeptides representing PTS1 and non-PTS1 sequences was mixed with purified protein possessing a His_6 tag, binding peptides could be pulled out of the mixture *via* the His_6 tag of the protein, using resin with a coordinating metal ion (**Figure 35**). The next stage would be to separate these binding peptides from the resin used for pull-down, and then identify which peptides had been bound to the protein using LC-MS based methods.

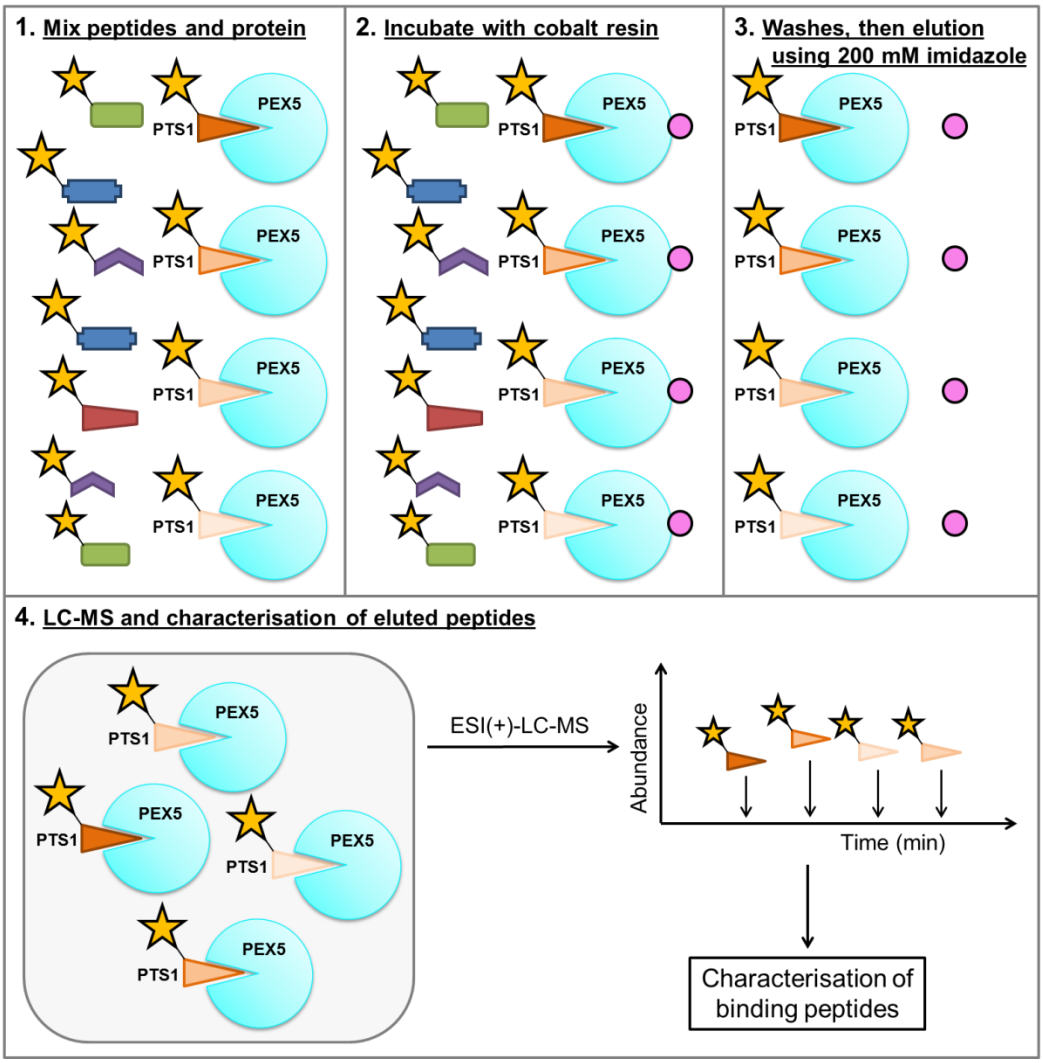


Figure 35| Scheme of the pull-down-LC-MS screening concept. Purified protein is mixed with a library of peptides. Cobalt-agarose resin is then used to capture protein using its His_6-tag. Protein, along with binding peptides, is eluted from cobalt-agarose resin. LC-MS analysis of the elution fraction allows the separation of binding peptides from one another and characterisation can then be performed for the identification of binding peptides.

This chapter will describe the development and optimisation of a pull-down screen in order to selectively identify protein-binding peptides from a library. A comparison of these identified peptides with PTS1 sequences predicted by bioinformatics will be presented for screen validation.

3.2 Synthesis of a library of N-terminally labelled peptides

3.2.1 Assembly of a peptide library

It was decided to randomise the two C-terminal amino acids of the pentapeptide library as, through analysis of PEX5-C:PTS1 crystal structures, this appeared to be where the

most variation in binding would be achieved (Chapter 2, section 2.4; Gatto et al., 2000). Randomisation of the two C-terminal residues of the pentapeptide YQSKL left 400 different peptide possibilities to synthesise. It was not necessary to prepare each peptide separately, as split-and-pool synthesis was used to prepare mixtures of peptides directly (**Figure 36**). The full peptide library was synthesised in four separate batches, for reasons that will be explained in this chapter, and then combined for screening.

Coupling reactions of each of the amino acids in the split-and-pool synthesis was carried out as before (for standard solid phase peptide synthesis) (Chapter 2, section 2.3). Following the assembly of the peptide libraries, a hydrophobic N-terminal fluorophore was used to label the peptides. This modification would increase the retention time of peptides on the C18 column used for LC-MS, improving separation of the peptides. The fluorescent label would also allow tracking of peptides during initial screen development.

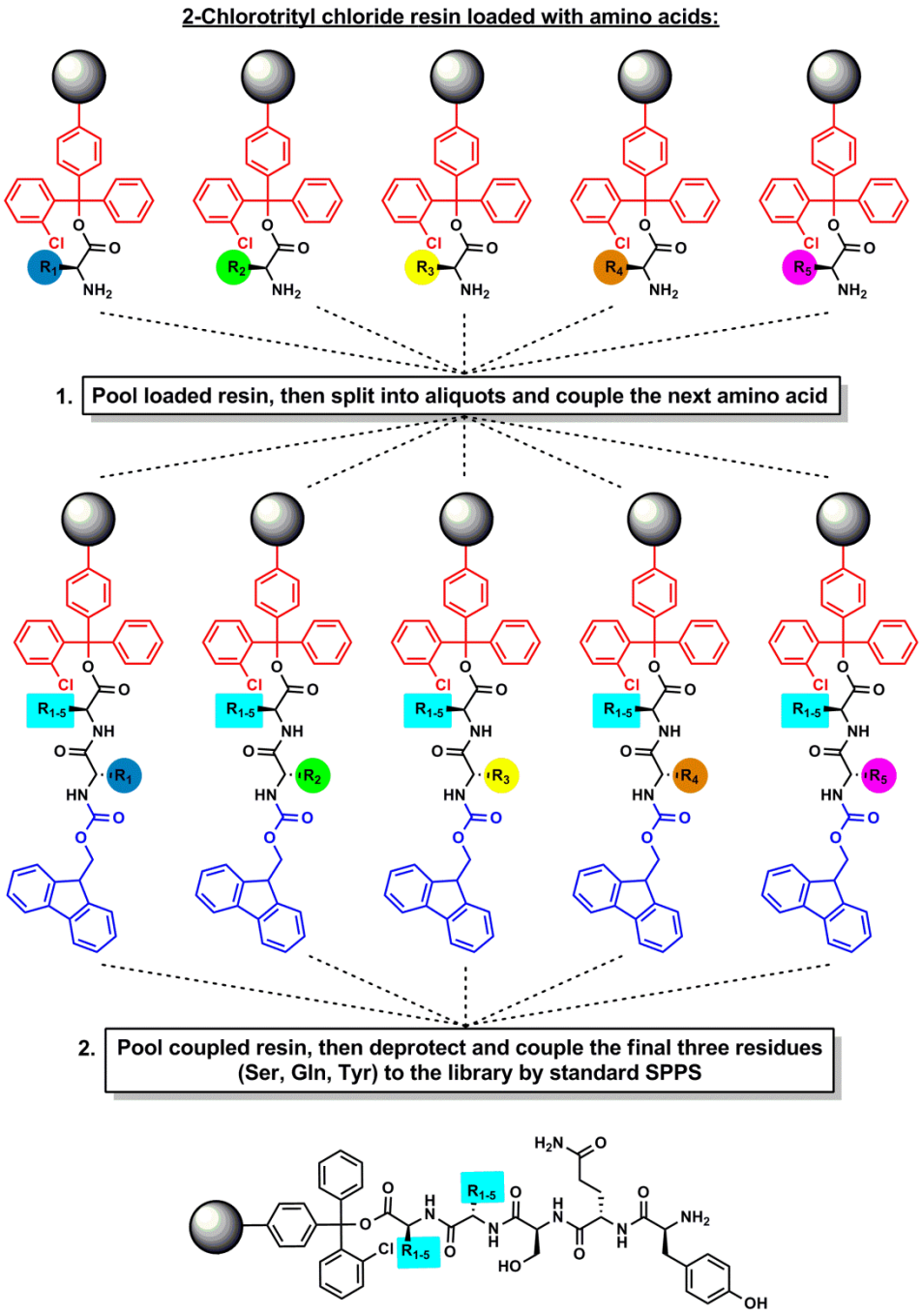


Figure 36| Main steps in the split-and-pool synthesis of a small peptide library. Here, 2-chlorotrityl chloride resin loaded with five different amino acids is pooled and split for coupling of the next five amino acids. Coupling of three N-terminal amino acids, Ser, Gln, and finally Tyr (at the N-terminus) leaves a library of pentapeptides with two randomised amino acids at the C-terminus. Acid-labile groups are in red, base-labile groups are in blue.

3.2.2 Determination of a suitable N-terminal label for peptides

Two fluorescent labels were considered for the N-terminus of the peptide library, a coumarin moiety or a dansyl moiety. The structures of these compounds, and the

schemes for coupling onto the N-terminus of a pentapeptide YQSxx (where x is any amino acid), are shown in **Figure 37**.

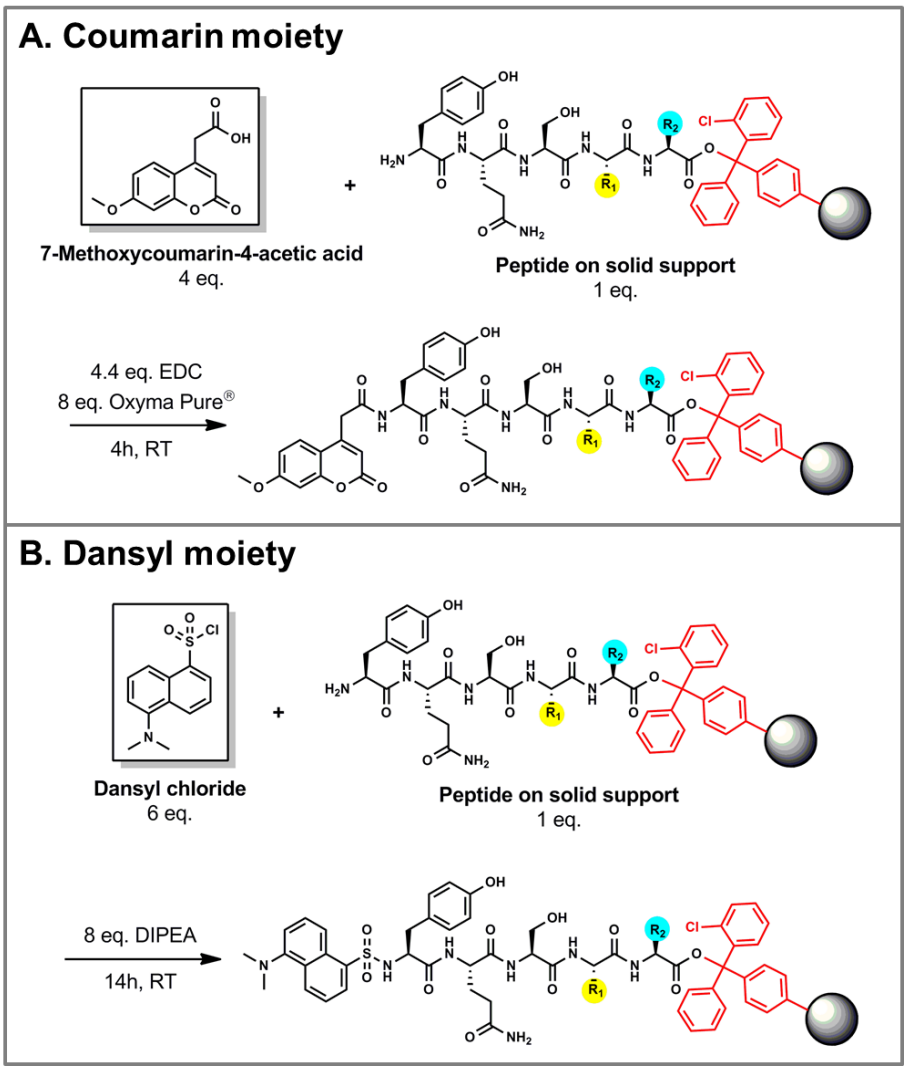


Figure 37| The schemes for the final coupling conditions for a coumarin moiety (A) and a dansyl moiety (B) to the N-terminus of a pentapeptide YQSxx, where x represents any amino acid. Both moieties contain aromatic groups, which will contribute to the hydrophobicity of the overall peptide. The acid-labile activated 2-chorotrityl chloride group on resin is shown in red.

After optimisation, the efficiency of the coumarin moiety coupling to peptides was only approximately 20%, whereas coupling of the dansyl moiety to peptides was approximately 90%. In order to determine which N-terminal label would be most suitable for the assay in terms of their effect on binding of the peptide to *At*PEX5-C, fluorescence anisotropy could be used. Competition assays were performed, monitoring the fluorescence anisotropy of lissamine-rhodamine-YQSKL upon its displacement from the *At*PEX5-C binding site by an increasing concentration of either unlabelled YQSKL or YQSKL labelled with either a coumarin or a dansyl moiety (for

details of the assay, see Chapter 6 (Experimental) section 6.14.6). By measuring the anisotropy it was possible to calculate the half-maximal inhibitory concentration (IC_{50}). The log(IC_{50}) was then used to calculate inhibition constant (K_i) (Nikolovska-Coleska et al., 2004; Skoulding et al., 2015) for coumarin-YQSKL and for dansyl-YQSKL in order to compare these with the K_i for unlabelled YQSKL (**Figure 38**). The error in K_i was calculated by a Monte Carlo error estimation approach (Skoulding et al., 2015), using the errors obtained for K_d (of the *At*PEX5-C:lissamine-YQSKL interaction) and for IC_{50} of the competitor peptide. Simulations were performed (2000 in total), based on simulated K_d and IC_{50} values around the observed mean. The standard deviation of each of these 2000 simulations was then used to estimate the error on K_i (Nikolovska-Coleska et al., 2004; Skoulding et al., 2015).

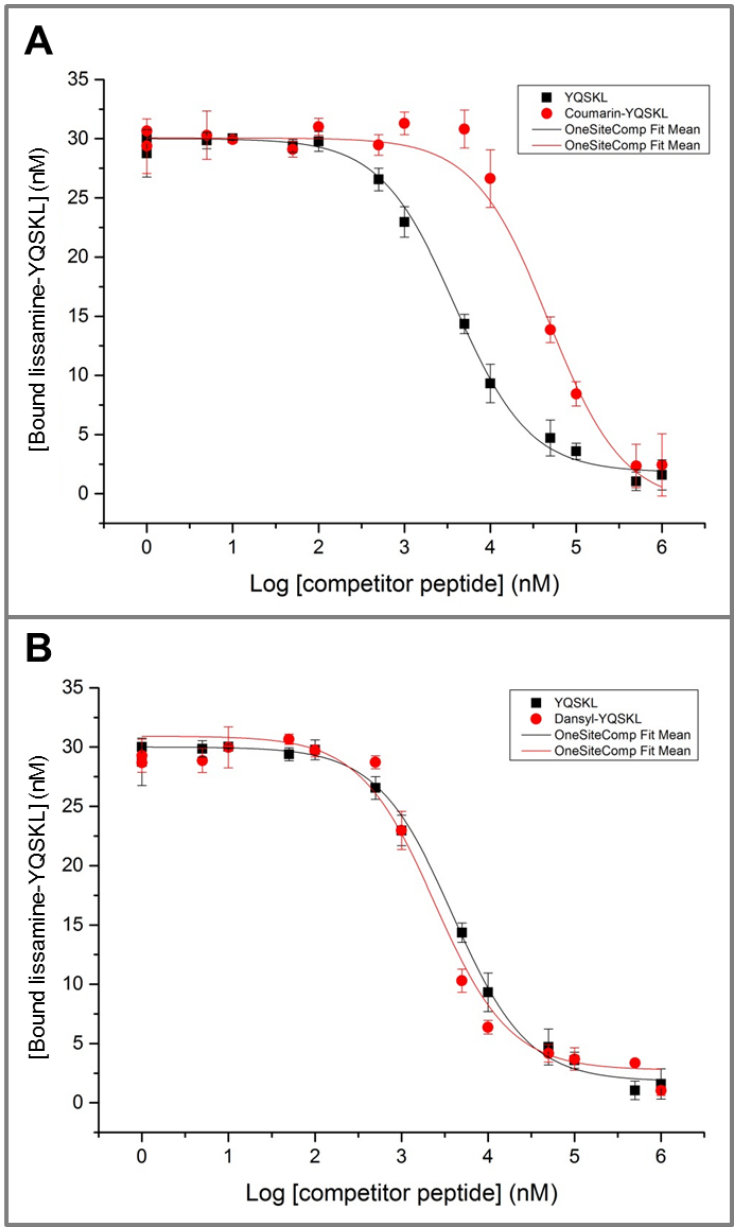


Figure 38| Fluorescence anisotropy competition assays using a fixed concentration of lissamine-YQSKL (30 nM final concentration), and of *At*PEX5-C (200 nM final concentration), and competitor peptide (ranging from 0–1 mM final concentration). A) Coumarin-YQSKL versus unlabelled YQSKL. B) Dansyl-YQSKL versus unlabelled YQSKL. Unlabelled YQSKL appears to out-compete lissamine-YQSKL more effectively than coumarin-YQSKL, whereas dansyl-YQSKL appears to out-compete lissamine-YQSKL slightly more effectively than unlabelled YQSKL. Data were fitted to a one-site competition model using OriginPro 9.1.

Using a fluorescence anisotropy competition assay with coumarin-YQSKL as the competing peptide, it was calculated that the log(IC_{50}) for coumarin-YQSKL was 4.68 ± 0.10 nM, whereas the log(IC_{50}) for unlabelled YQSKL was 3.59 ± 0.03 nM. This equated to a K_i of coumarin-YQSKL being 280 ± 170 nM, whereas the K_i of unlabelled YQSKL was 22 ± 12 nM. The coumarin appeared to be having an adverse effect on *At*PEX5-C binding of the peptide. The binding of YQSKL to *At*PEX5-C appeared to be

relatively unaffected by the N-terminal dansyl label, as the log(IC_{50}) for dansyl-YQSKL was 3.38 ± 0.10 nM. This equated to a K_i of dansyl-YQSKL being 13 ± 9 nM, which is very similar to the K_i of unlabelled YQSKL.

The use of a dansyl label at the peptide N-terminus was preferable to a coumarin label as a dansyl moiety at the N-terminus of the binding peptide did not appear to adversely affect binding of the peptide to *At*PEX5-C. A dansyl moiety was therefore coupled to the N-terminus of the library of peptides for screening.

3.3 Peptide library characterisation

The peptide library was analysed by LC-MS, using a C18 column prior to electrospray ionisation, and m/z determined using a Bruker maXis impact Q-TOF mass spectrometer (Chapter 6 (Experimental) section 6.15.2). As many peptides were injected into the mass spectrometer in one run, the individual spectra of all peptides were effectively combined with one another. A base peak chromatogram shows the most abundant peak at each time point across an eight minute LC-MS run. This highlights the problematic issue of separating individual peptides out of this mixture (**Figure 39**).

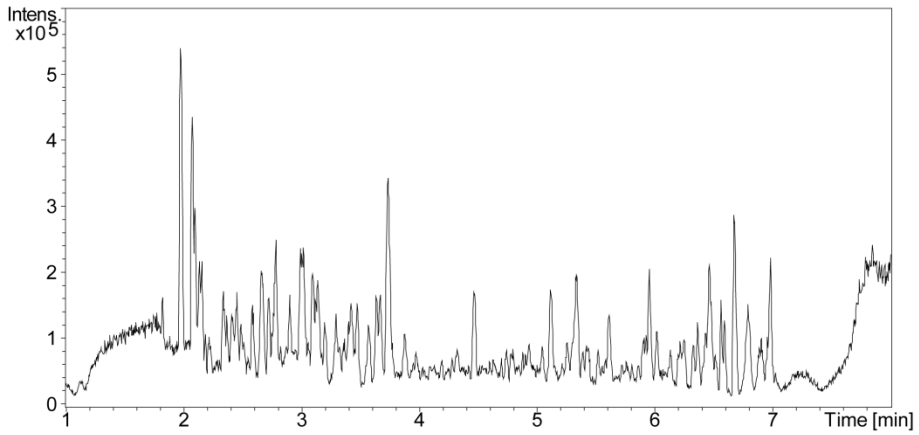


Figure 39| Base peak chromatogram for an LC-MS run of the entire peptide library, dansyl-YQSxx, where x is any amino acid. Many peaks are present and it is not possible to take useful characterisation information away from this analysis.

The raw data (the mass spectrum) for a selected time point, 3.1–3.3 minutes, within this LC-MS run is shown in **Figure 40**. Peptides were predicted to range from 743.3–1001.4 g/mol in molecular weight with an N-terminal dansyl label. It can be seen that a number of peptide $[M+H]^+$ values are present in this small range (**Figure 40**). Peptides

were seen between the range of 2 and 6.5 minutes for an eight-minute LC-MS run. This shows the difficulty faced in terms of characterising each of these peptides individually.

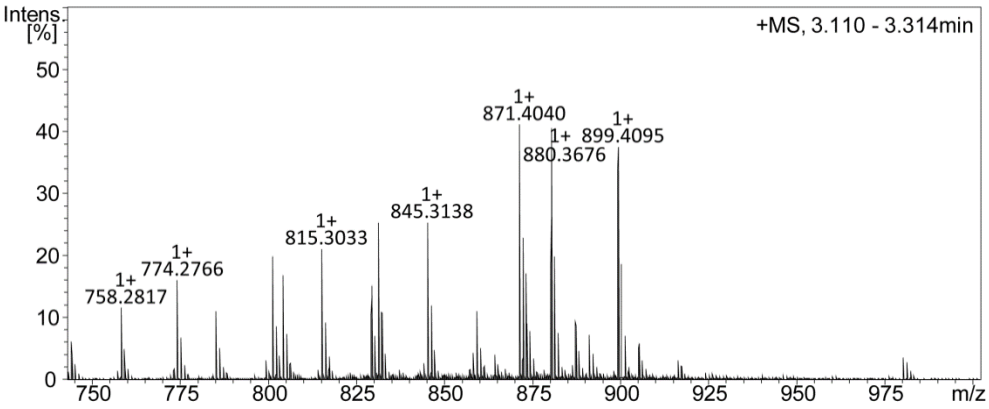


Figure 40| Masses observed for singly charged ions between 3.1 and 3.3 minutes for an eight minute LC-MS run of all peptides in the library. Singly charged ions between 743 and 1002 m/z are shown. There are many peaks to discriminate from one another.

The size range of peptides in the library was calculated by drawing the smallest peptide (dansyl-YQSGG) and the largest peptide (dansyl-YQSWW) in ChemDRAW and taking the mass values provided by ChemDRAW to define the mass range. In order to calculate the masses of all peptides in the library from their formulas, Pipeline Pilot was used (by Stuart Warriner) to quickly enumerate the library. With the formulas and, hence, masses of peptides in the library known, it was possible to look for each of these masses in the mass spectrum of the peptide library and begin to characterise each of the dansyl-labelled peptides separately.

3.3.1 Extraction of ion chromatograms based on exact masses of peptides

Although the masses of all peptides in the library had been obtained, it was apparent that some peptides had identical predicted nominal masses. The nominal mass of a peptide is the average of the mass of all possible isotopes. This problem of identical nominal masses is illustrated in **Table 8**, where a selection of four pairs of peptides are shown, along with the nominal masses and the exact masses of the peptides.

Pair	Peptide sequence	Nominal mass	Exact mass
1	Dansyl-YQSGQ	814.3	814.2956
	Dansyl-YQSGK	814.3	814.3320
2	Dansyl-YQSNN	857.3	857.3014
	Dansyl-YQSDI	857.3	857.3266
3	Dansyl-YQSQE	886.3	886.3240
	Dansyl-YQSAW	886.3	886.3393
4	Dansyl-YQSQW	943.4	943.3534
	Dansyl-YQSKW	943.4	943.3898

Table 8| Nominal masses and exact masses for four pairs of peptides. In the case of each peptide pair, the nominal masses are identical and only exact masses are unique.

As is apparent in **Table 8**, nominal masses could not be relied upon for efficient characterisation in this situation. Analysis of peptides by high resolution mass spectrometry was therefore crucial for identifying peptides uniquely within the library. Exact masses were needed and it was possible to obtain these using the Bruker maXis impact Q-TOF mass spectrometer. LC-MS runs could therefore be processed by extracting ion chromatograms for the exact mass of each of the peptides in the library.

The peptides dansyl-YQSQE and dansyl-YQSAW have the same predicted nominal mass (**Table 8**) and are the most similar in exact mass, so this pair was used for graphical representation in order to show the requirement for an accurate mass instrument in processing data generated from this screen (**Figure 41**).

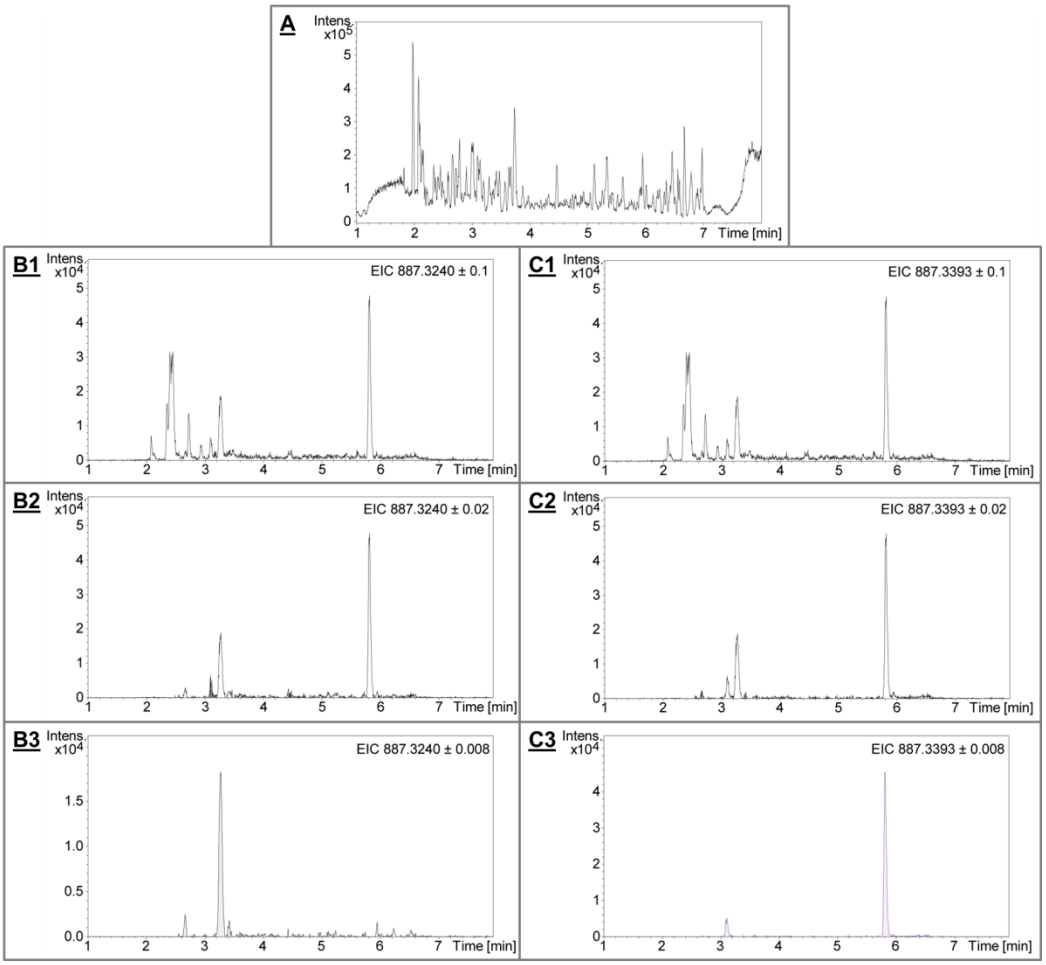


Figure 41| Distinct identification of two peptides with very similar exact masses. A) Base peak chromatogram of all peptides in the library. B1) Extracted ion chromatogram for the exact mass of dansyl-YQSQE with an error of 0.1 g/mol. B2) Extracted ion chromatogram for the exact mass of dansyl-YQSQE with an error of 0.02 g/mol. B3) Extracted ion chromatogram for the exact mass of dansyl-YQSQE with an error of 0.008 g/mol. C1) Extracted ion chromatogram for the exact mass of dansyl-YQSAW with an error of 0.1 g/mol. C2) Extracted ion chromatogram for the exact mass of dansyl-YQSAW with an error of 0.02 g/mol. C3) Extracted ion chromatogram for the exact mass of dansyl-YQSAW with an error of 0.008 g/mol.

Figure 41 shows that it was possible to identify individual peaks for each peptide. Adding extracted ion chromatograms (EICs) therefore provided a way to separate individual peptides from the mixture. After an LC-MS run of the full peptide library, data were processed so that an extracted ion chromatogram for the exact mass of each peptide, ± 0.008 g/mol, would be shown. This process was automated (Appendix D, section D.2) so that EICs were added for the exact masses of each of the peptides in the library. All EICs found from the LC-MS run of the full peptide library are shown in **Figure 42**.

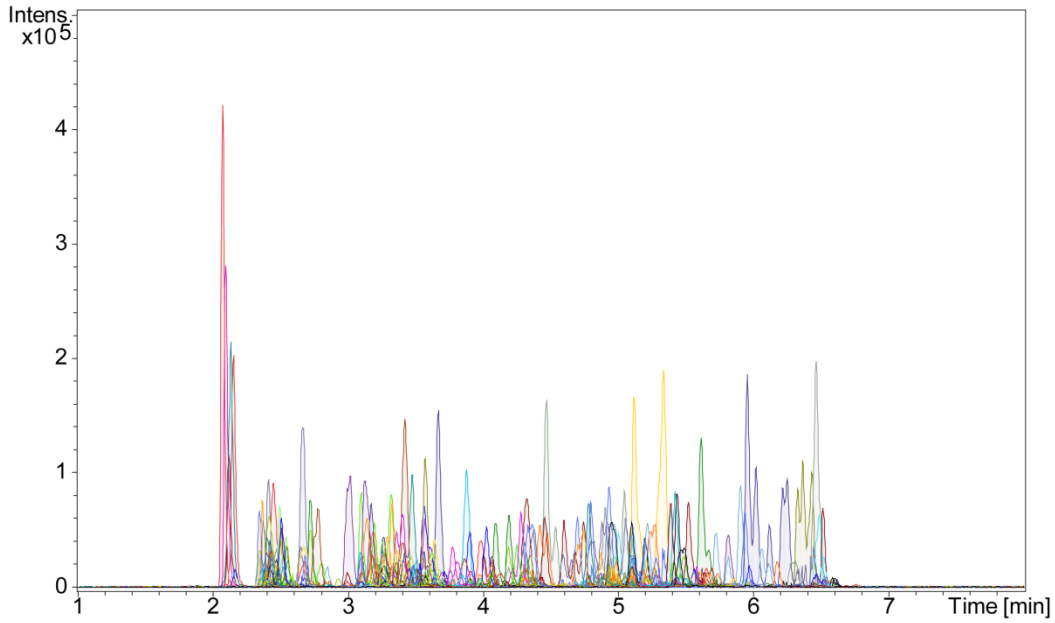


Figure 42| Extracted ion chromatograms for dansyl-labelled peptides in the full library run by LC-MS.

By using extracted ion chromatograms, even overlapping peaks could be detected (**Figure 42**). In a library of peptides with the two C-terminal residues randomised, there is also the problem of multiple peptides in the library having identical exact masses. For example, the positional isomers dansyl-YQSKW and dansyl-YQSWK have identical exact masses and there were many other positional isomers in the library. Tandem mass spectrometry was used for conclusive characterisation in these cases.

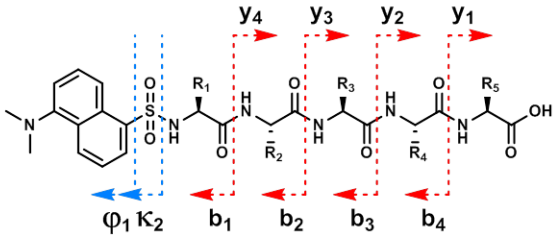
3.3.2 The use of tandem mass spectrometry for further characterisation

There were a number of peptides in the library with identical exact masses to one another. A selection of three groups of these peptides is shown in **Table 9**.

Group	Peptide sequence	Exact mass
1	Dansyl-YQSAP	797.3054
	Dansyl-YQSPA	797.3054
2	Dansyl-YQSDT	845.2902
	Dansyl-YQSTD	845.2902
	Dansyl-YQSES	845.2902
	Dansyl-YQSSE	845.2902
3	Dansyl-YQSKF	904.3789
	Dansyl-YQSFK	904.3789

Table 9| Three groups of peptides in the full peptide library along with their exact masses. It can be seen that each group of peptides have identical exact masses.

Provided the retention times for these peptides were different after LC-MS, it was possible to use tandem mass spectrometry to identify the unique retention time for each peptide. Tandem mass spectrometry is a fragmentation technique which involves the isolation of particular peptide 'precursor' ions (some of which are displayed in **Figure 40**) and the fragmentation of these *via* collisions with an inert gas (Steen and Mann, 2004). This leads to the production of 'product' ions which can then be analysed. The most likely fragmentation pattern for a peptide is by fragmentation at the amide bonds, which leads to the production of 'b' and 'y' ions (**Figure 43**). Unique product ions can then be identified. Taking group 3 (**Table 9**) as an example, product ions predicted after fragmentation of the two peptides dansyl-YQSKF and dansyl-YQSFK are shown in **Figure 43**.



	Dansyl-YQSKF	Dansyl-YQSFK
y_4	509.2724	509.2724
y_3	381.2138	381.2138
y_2	294.1818	294.1818
y_1	166.0869	147.1134
φ_1	170.0970	170.0970
κ_2	234.0589	234.0589
b_1	397.1301	397.1301
b_2	525.1887	525.1887
b_3	612.2207	612.2207
b_4	740.3157	759.2891

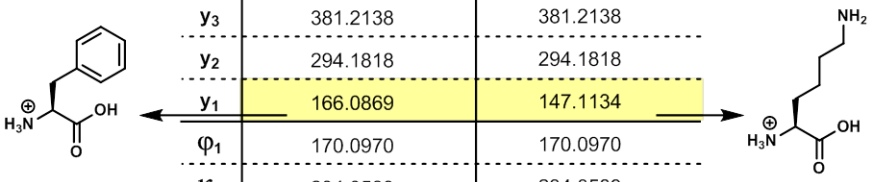


Figure 43| Illustration of the ions generated after the fragmentation of a generic peptide. A table of predicted masses of 'product' ions generated after the fragmentation of dansyl-YQSKF and dansyl-YQSFK is also shown, with masses in g/mol. Predicted masses of product ions were generated using the Fragment Ion Calculator (*db.systemsbiology.net:8080/proteomicsToolkit/FragIonServlet.html*).

Two different unique ions (y_1 and b_4) can be seen after the simulated fragmentation of dansyl-YQSKF and dansyl-YQSFK. These ions are both generated by fragmentation of the final amide bond in the peptide chain, the bond that will always be located between the two randomised C-terminal residues in this study. The y_1 ion was chosen for the separate identification of all peptides in the library with identical exact masses, as this was the smallest unique ion. **Figure 44** shows the peaks identified as either dansyl-YQSKF or dansyl-YQSFK, after extraction of ion chromatograms based on the exact mass of each peptide (section 3.3.1). These were seen as two distinct peaks because of the difference in retention times of the two peptides. Tandem mass spectra are shown for the time points in the mass spectrum where each of these two peaks were found. Here, unique product ion peaks are seen and these are highlighted in **Figure 44**.

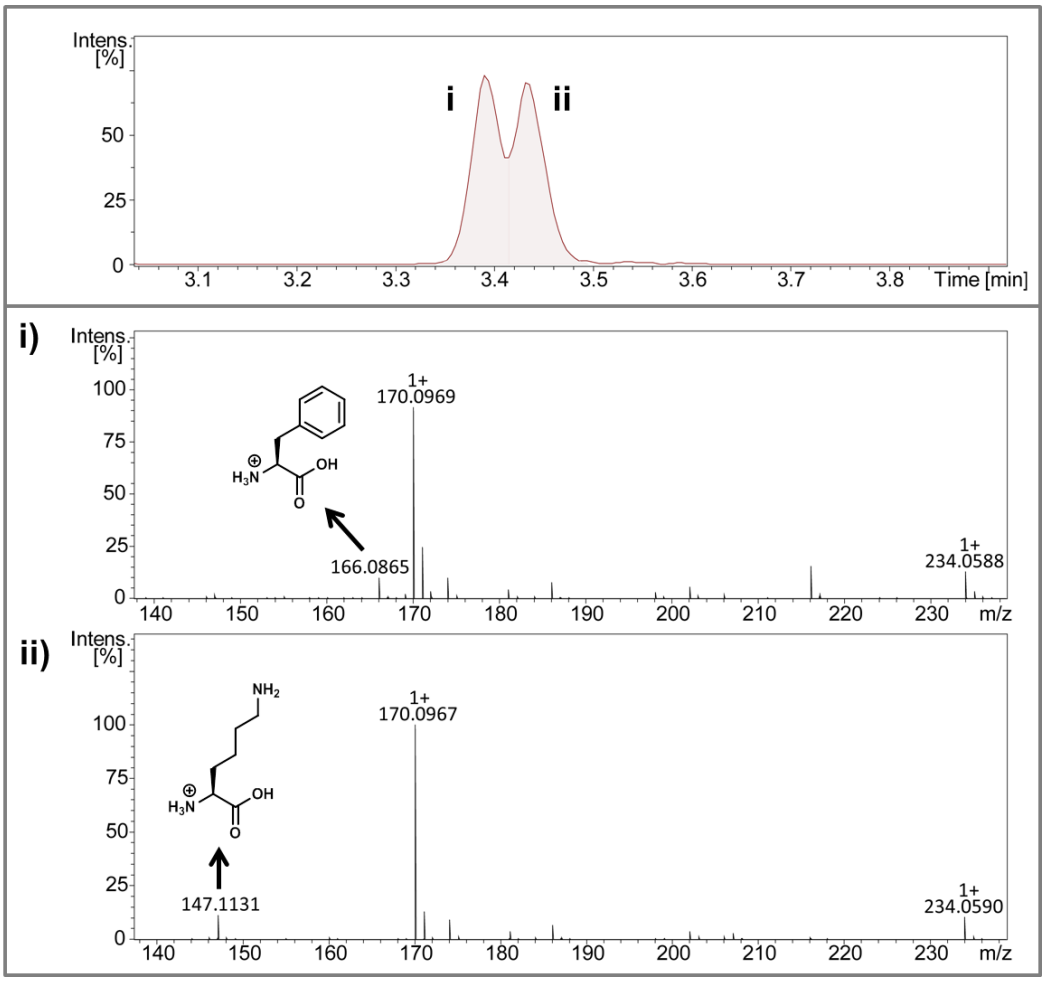


Figure 44| The use of tandem mass spectrometry for identification of unique peptides. Top panel: extracted ion chromatograms corresponding to dansyl-YQSKF and dansyl-YQSFK. i) Tandem mass spectrum for peak (i), in which the y_1 ion of phenylalanine can be seen. ii) Tandem mass spectrum for peak (ii), in which the y_1 ion of lysine can be seen.

It can be concluded from **Figure 44** that peak (i) corresponds to dansyl-YQSKF and peak (ii) corresponds to dansyl-YQSFK. By performing analysis in this way for all positional isomers, and indeed all peptides with identical exact masses, it was possible to acquire unique peptide identifiers using retention time and mass. These unique peptide identifiers were termed 'retention time & mass', or ReT-mass, identifiers in this study. In order to assign ReT-mass identifiers quickly, the process of searching for y_1 ions in tandem mass spectra was automated on the mass spectrometer (Appendix D, section D.2).

Tandem mass spectrometry was a valuable technique for the unique identification of a number of peptides. In some cases, however, even tandem mass spectrometry will not discriminate between two peptides. The amino acids leucine and isoleucine are

constitutional isomers, so the y_1 ions generated by the presence of either of these amino acids at the C-terminus of a peptide will have identical masses. For this reason, the entire library of peptides was synthesised as four separate sub-libraries. This ensured that full characterisation could be carried out, and as many ReT-mass identifiers as possible could be assigned, before the full-peptide library was assembled.

3.3.3 Sorting of peptides into sub-libraries based on mass

The entire library of peptides was split into four groups before synthesis. This was achieved by splitting amino acids into two groups based on mass. During the split-and-pool synthesis (**Figure 36**), these amino acids were combined in such a way that four sub-libraries of peptides were produced. Exact masses of peptides had already been enumerated virtually (by Stuart Warriner, using Pipeline Pilot) and calculated based on the formulas of the peptides. This allowed the virtual sorting of peptides into sub-libraries so that the maximum amount of peptides with identical exact masses to one another was two. If the entire peptide library had been synthesised in one mixture, the maximum amount of peptides with identical exact masses to one another would have been six. Sorting amino acids into two groups meant that amino acids with identical exact masses (leucine and isoleucine) and amino acids differing by a methyl group (for example, serine and threonine) could be separated. Cysteine was excluded from peptide sub-libraries due to its propensity to form disulfide bonds, which could complicate analysis. Amino acids in each of the groups (1 or 2), along with combinations of these groups required for the production of each sub-library, are displayed in **Figure 45**.

Sub-library 1.1: Dansyl-[Y]-[Q]-[S]-[D/F/G/H/I/K/N/P/S/V]-[D/F/G/H/I/K/N/P/S/V]

Sub-library 1.2: Dansyl-[Y]-[Q]-[S]-[D/F/G/H/I/K/N/P/S/V]-[A/E/L/M/Q/R/T/W/Y]

Sub-library 2.1: Dansyl-[Y]-[Q]-[S]-[A/E/L/M/Q/R/T/W/Y]-[D/F/G/H/I/K/N/P/S/V]

Sub-library 2.2: Dansyl-[Y]-[Q]-[S]-[A/E/L/M/Q/R/T/W/Y]-[A/E/L/M/Q/R/T/W/Y]

Figure 45| The amino acid composition of each of the four sub-libraries. Sequences are written from N- to C-terminus and sub-libraries were varied at the two C-terminal residues. In red, group 1 amino acids; in blue, group 2 amino acids.

The four peptide sub-libraries were run on the mass spectrometer separately by LC-MS. Taking all peptide sub-libraries into account, it should have been possible to

identify 361 peptides. It was, however, not possible to identify any of the methionine-containing peptides in their unoxidised state, and carrying out peptide library cleavage under reducing conditions compromised the yield of other peptides in the libraries, so methionine was excluded from analysis.

After exclusion of methionine-containing peptides, a total library of 324 peptides had been successfully synthesised. It was possible to obtain unique ReT-mass identifiers for 234 of these peptides, with the use of tandem mass spectrometry in some cases (**Figure 44**). This enabled the production of a master list comprising peptide sequences and retention times, termed the 'master Seq-ReT list' in this study. In the case of 45 pairs of peptides, the LC retention times for the positional isomers were identical so it was not possible to obtain unique ReT-mass identifiers for these 90 peptides; however, the peaks could be assigned as belonging to the pair. A table of all peptide retention times, along with sequences, are shown in Appendix D, section D.1.1. In the case of the 45 pairs of peptides indistinguishable from one another, one of the peptides from each pair was included in the master Seq-ReT list and is shown in the table in Appendix D, section D.1.1 in pale blue, and the other peptide from the pair is shown in the table in Appendix D, section D.1.2.

3.3.4 Data processing after an LC-MS run

In order to accelerate analysis, data were automatically processed. The extracted ion chromatograms (EICs) were automatically generated on the mass spectrometer and the peak data (mass, retention time, and area under EIC) were exported as a '.csv' file for each LC-MS run. VBA was then used to create macros in Microsoft Excel which would process the data from each exported '.csv' file. These macros compared the mass and retention time data for each LC-MS run with the reference data obtained during the development of the assay (the master Seq-ReT list), in some cases applying a small compensating retention time drift. This allowed rapid and consistent assignment of each peak to the corresponding peptide sequence.

In the case of peptide pairs without unique ReT-mass identifiers, the macros would assign the appropriate area-under-EIC to the sequence of the peptide in the master Seq-ReT list. This area-under-EIC would also then automatically be assigned to the other peptide from the pair. VBA was also used to plot the area-under-EIC data against peptide sequences corresponding to these EICs. This was presented in a heat map format to show only the two C-terminal residues, where the x-axis displays the C-

terminal or '-1' amino acid and the y-axis displays the '-2' amino acid of the peptide (**Figure 46**). Details of the function of each of the macros created and used in this study, and the code for each, are shown in Appendix D, section D.3.

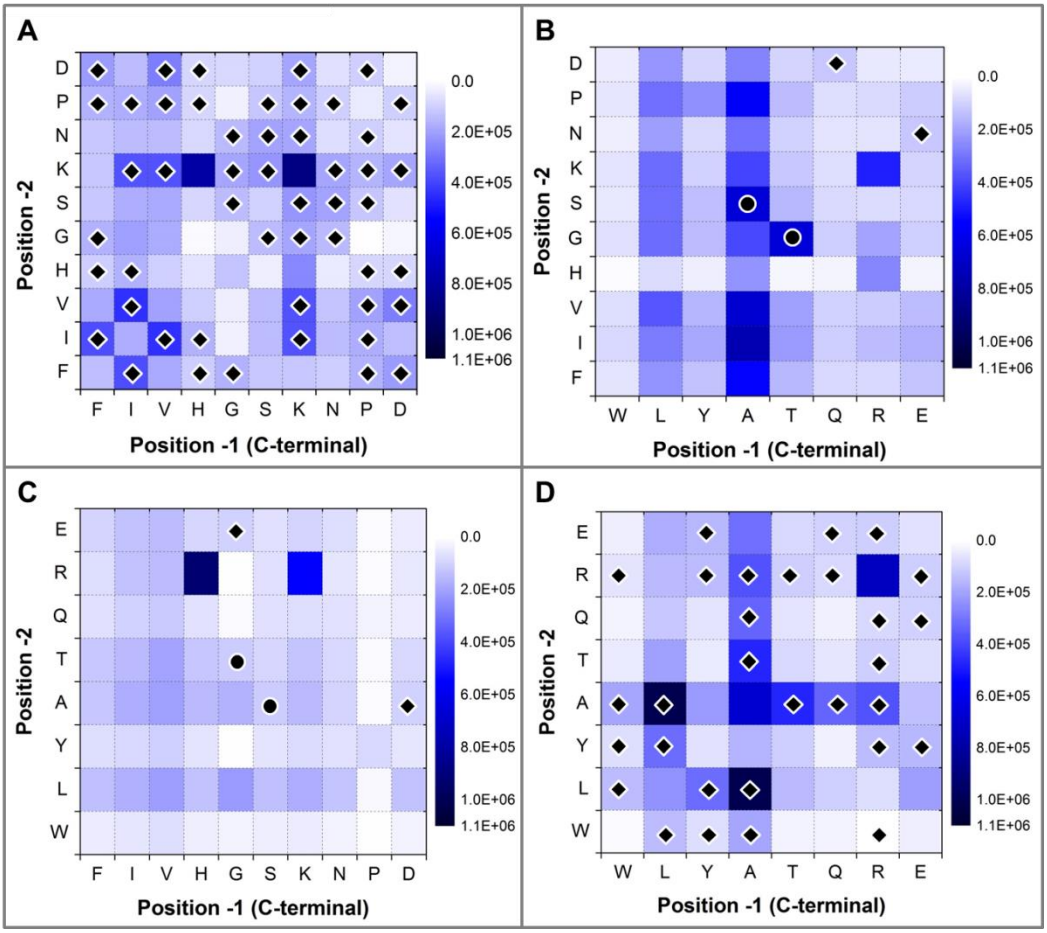


Figure 46| Peptides that were present in each sub-library after LC-MS analysis. A) Sub-library 1.1, B) Sub-library 1.2, C) Sub-library 2.1, and D) Sub-library 2.2. For sub-libraries 1.1 and 2.2, diamonds represent positional isomer peptides with identical retention times. For sub-libraries 1.2 and 2.1, each type of symbol (either diamond or circle) represents a pair or peptides with identical retention times. These symbols therefore represent peptides without unique retention time & mass (ReT-mass) identifiers. The squares are shaded according to the area-under-EIC for each peptide.

It was possible to assign an area-under-EIC to all peptides in each of the sub-libraries, which meant that all peptides in the full library were characterised. It was then decided to run the full peptide library by LC-MS in order to test the characterisation of this larger mixture of peptides after an LC-MS run.

3.3.5 An LC-MS run and processing of the full peptide library

A sample containing all peptide sub-libraries combined was run by LC-MS (**Figure 42**) and results showing the area-under-EIC for each peptide in the sample are shown in heat map format (**Figure 47**).

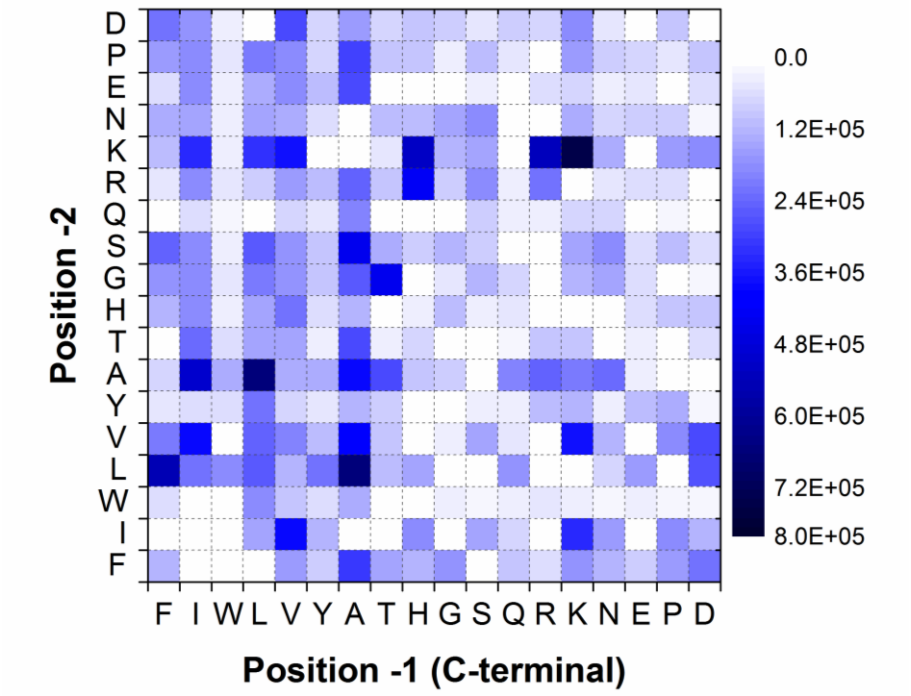


Figure 47| Peptides present after processing an LC-MS run of the full peptide library. Peptides were detected at varying intensities; however, all peptides in the full library were detected. Amino acids on each axis are sorted from hydrophobic to hydrophilic, as determined by the Sigma-Aldrich Hydrophobicity Index (Sereda et al., 1994; Monera et al., 1995).

As can be seen from the heat maps of the LC-MS runs of peptides alone, the signal of each peptide present in the full peptide library varied greatly from one another. This could have been due to the mass difference between peptides, as peptide sub-libraries were resuspended based on the average mass in the sub-library. However, more importantly, the absolute area would be affected by how well each of the peptides ionised, with peptides containing positively charged basic side chains ionising much more effectively. It was therefore important that, when using this technique to interpret screening results, the semi-quantitative presence or absence of each of the peptides being pulled down was treated as more significant than the absolute intensity of the peaks.

3.4 Development and optimisation of a peptide pull-down screen

An initial test pull-down-LC-MS experiment was carried out using 40 µM purified wild-type *At*PEX5-C and 5 µM of each peptide in peptide sub-library 1.2. This allowed us to investigate whether YQSKL, the representative PTS1 sequence used in this study, was pulled down by the wild-type protein. The screen was carried out by incubating purified protein with peptide sub-library 1.2 for 1 hour at 4°C. This protein-peptide mixture was then added to cobalt resin followed by a 1h incubation at 4°C. Five buffer wash steps were performed, followed by a 30 minute (4°C) incubation with buffer containing 200 mM imidazole to release protein from the cobalt resin. When carrying out these initial screens, fluorescence intensity was measured at 550 nm to identify the amount of dansyl fluorophore in each fraction (**Figure 48**). Controls in the absence of protein or peptide were treated identically.

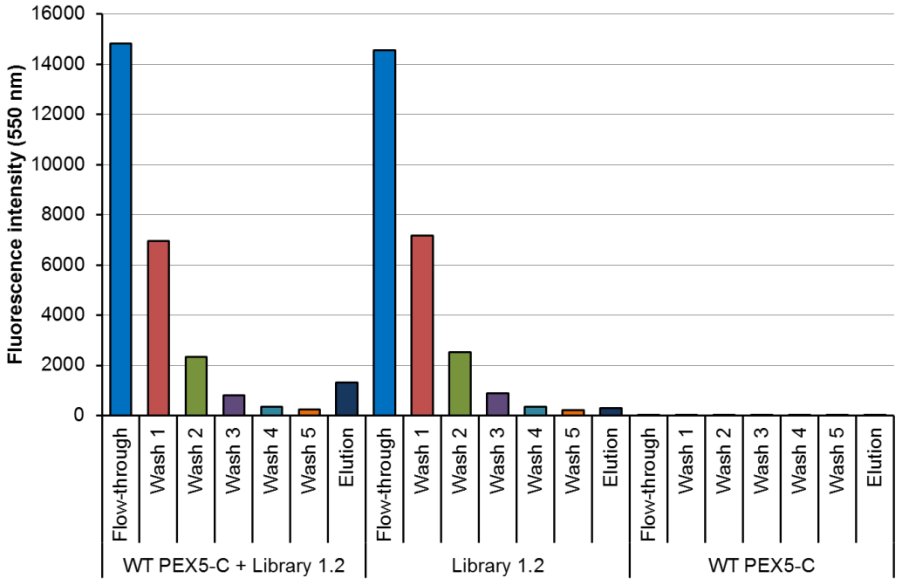


Figure 48| Fluorescence intensity measurements at 550 nm in order to track dansyl levels in all stages of preliminary pull-down screen experiments.

By tracking the dansyl fluorescence through the stages of the screen (**Figure 48**), it could be observed that the wild-type *At*PEX5-C protein appeared to be releasing some peptide upon elution from the cobalt resin. This increase in fluorescence was not observed in the controls. This elution fraction was analysed by LC-MS to identify which peptides of library 1.2 were present after having presumably been bound to the wild-type *At*PEX5-C protein. Results of the LC-MS run are shown in **Figure 49**, compared to the LC-MS run of sub-library 1.2 before screening.

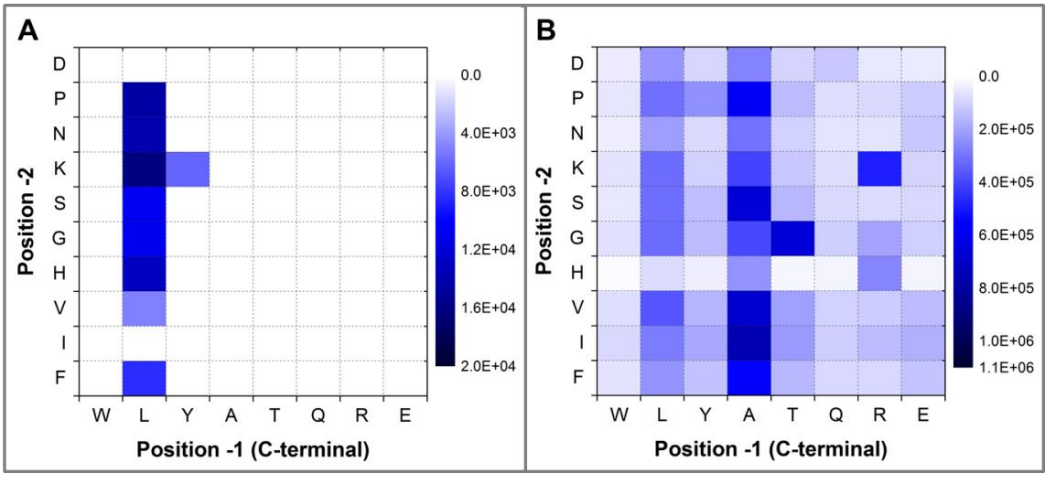


Figure 49| A) LC-MS of elution fraction of a pull-down-LC-MS screen for wild-type *At*PEX5-C with sub-library 1.2. B) Sub-library 1.2 before screen. Wild-type *At*PEX5-C is almost exclusively isolating peptides with a C-terminal leucine, which shows selectivity when compared with peptide sub-library 1.2 before screening.

Selective pull-down of peptides was clearly occurring. By comparing the results from this pull-down-LC-MS screen with previously published *in vivo* peroxisomal import experiments (Lingner et al., 2011; Skoulding et al., 2015), it was possible to reveal whether the wild-type *At*PEX5-C protein was selectively identifying PTS1 sequences. Within the entire sub-library 1.2 (comprising 81 peptides), there are only six experimentally validated C-terminal tripeptide PTS1 sequences, and a total of nine predicted PTS1s (Lingner et al., 2011; Chowdhary et al., 2012): –SKY, –SDL, –SPL, –SFL, –SGL, –SHL, –SNL, –SSL, and –SKL. Of these nine sequences, eight were identified in the pull-down-LC-MS screen, demonstrating that the experiment does indeed enable identification of PTS1 sequences.

3.5 Further optimisation of the pull-down screen and validation

The pull-down screen protocol was further modified in order to allow the pull-down of a larger number of peptides. As the initial experiment had used 40 µM purified protein and 5 µM of each peptide in the library, this in theory would mean that 8 very tight protein-binding peptides could occupy all available protein binding sites. This would leave only the tightest protein-binding peptides available for analysis. It was expected that any new peptide binding seen with an *At*PEX5-C variant would not have a particularly high affinity so the concentration of purified protein used in the screen was varied and analysed. A service of the mass spectrometer also greatly improved

sensitivity of the instrument so lower concentrations of overall peptide libraries were needed.

The pull-down-LC-MS protocol was modified slightly to include three extra wash steps, to account for the full peptide library being used, and elution of the peptides by use of 6M urea in the buffer to unfold the protein coordinated to resin to release bound peptides. The unfolded protein remained bound to the cobalt resin, enabling more straightforward LC-MS analysis (**Figure 50**).

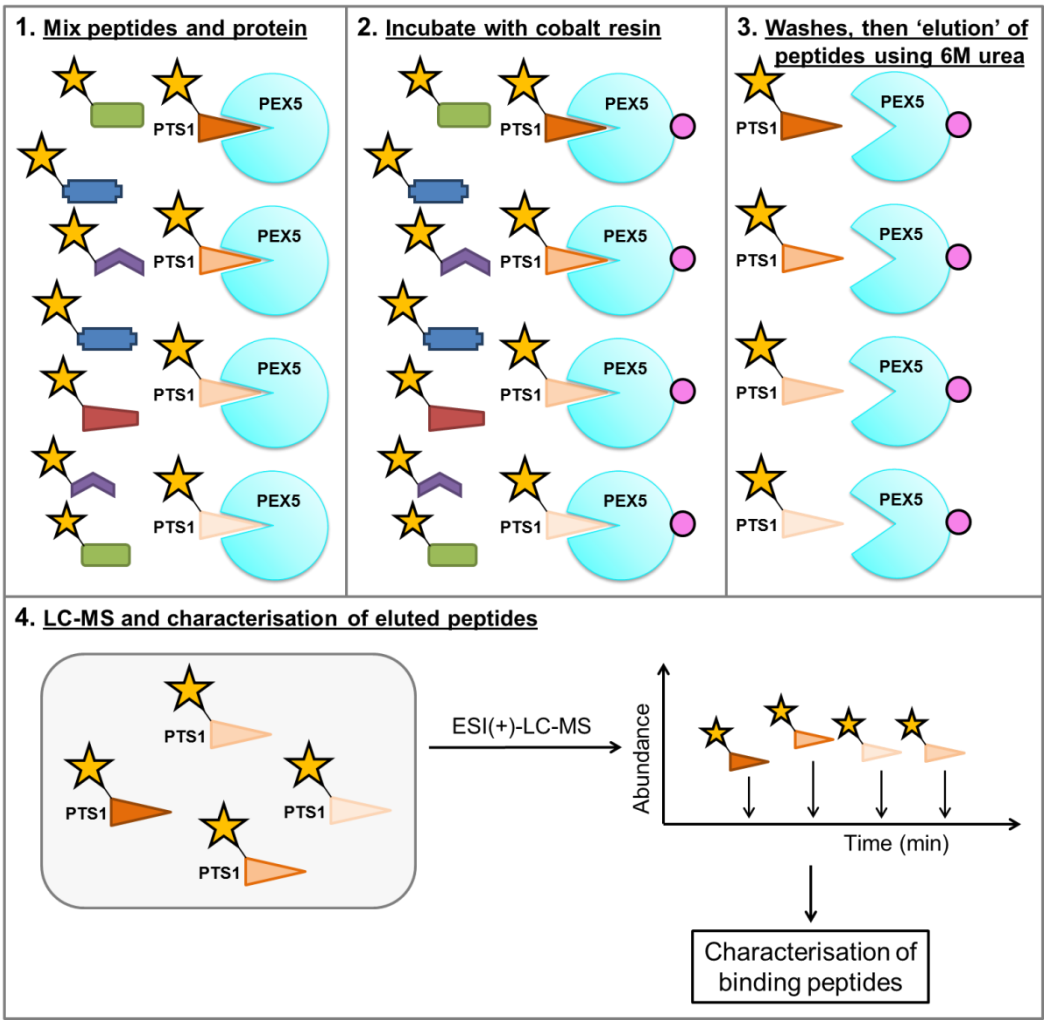


Figure 50| Modified pull-down-LC-MS protocol, using 6M urea to isolate peptides. Cobalt-agarose resin is used to capture protein, after incubation with a library of peptides, using its His_6-tag. Binding peptides are released by unfolding the protein, which allows LC-MS analysis of binding peptides.

Concentrations of peptides in the libraries were set at approximately 500 nM of each peptide for all preliminary experiments in this section. The concentration of purified wild-type *At*PEX5-C protein was varied from 2.5 μM to 25 μM. The peptide sub-

libraries were combined for these experiments, enabling screening of the full library of 324 peptides, and the peptides pulled down with each concentration of purified protein are shown in **Figure 51**.

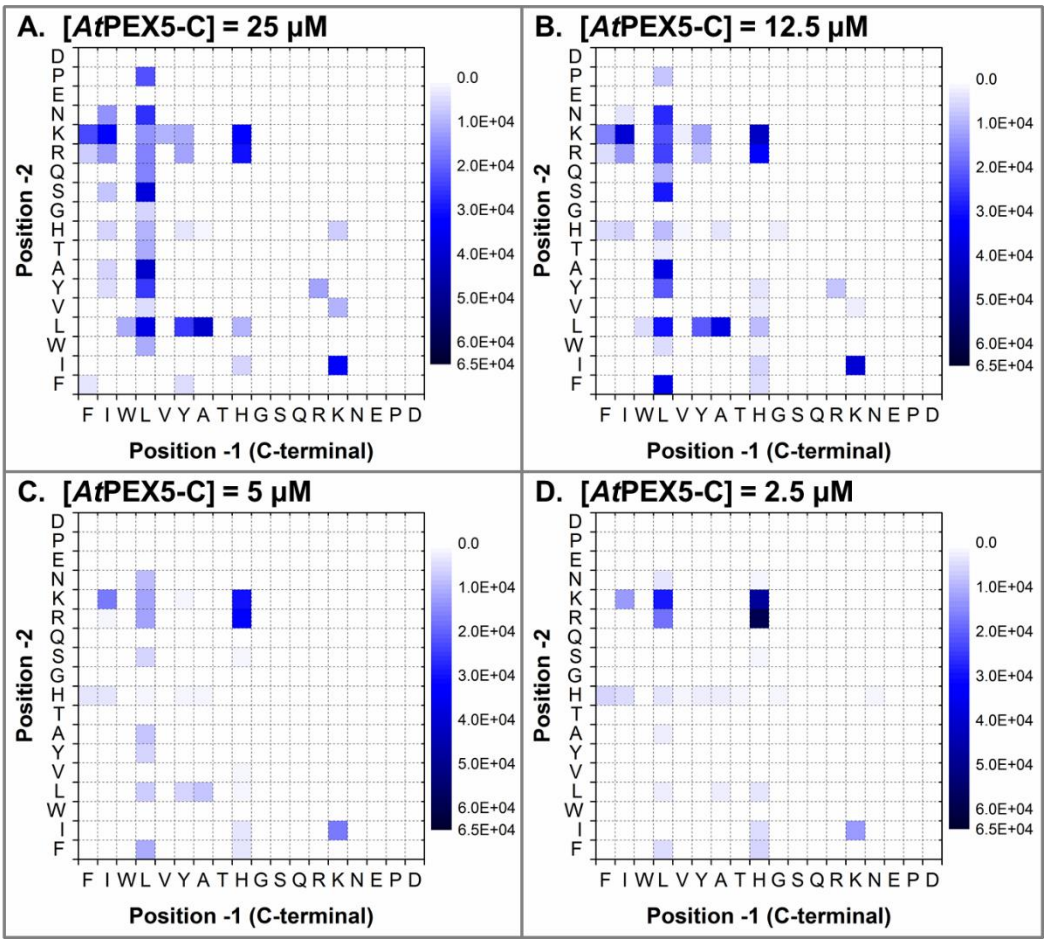


Figure 51| A) Pull-down-LC-MS screen using 25 µM wild-type *At*PEX5-C and approximately 500 nM each peptide in the full library. B) Pull-down-LC-MS screen using 12.5 µM wild-type *At*PEX5-C and ~500 nM each peptide in the full library. C) Pull-down-LC-MS screen using 5 µM wild-type *At*PEX5-C and ~500 nM each peptide in the full library. D) Pull-down-LC-MS screen using 2.5 µM wild-type *At*PEX5-C and ~500 nM each peptide in the full library. Screens (A) and (B) show a very similar overall peptide binding profile. The full-range of peptide binding appears to be lost in screen (C) and appears to be further limited in screen (D).

Following the range of concentration tests performed with wild-type *At*PEX5-C, it was decided that the optimum concentrations that would be tested while providing sufficient information on the peptide binding profile was 12.5 µM purified protein and ~500 nM each peptide. This allowed a wide range of peptides to be pulled down by the protein while using a low enough concentration of protein to allow any screen repeats from one protein purification for consistency. The screen was carried out in a total volume of

500 µL, meaning that 284.9 µg (6.25 nanomoles) protein and approximately 0.217 µg (0.25 nanomoles) of each peptide was used.

These data also allowed the first opportunity to fully compare the sequences that were pulled down across the library with experimentally validated and predicted PTS1 sequences. In order to carry out this comparison, the pull-down-LC-MS data were compared with *in silico* PTS1 prediction software (Lingner et al., 2011; Reumann et al., 2012) to assess whether the peptides identified by pull-down-LC-MS were also predicted PTS1 sequences. The PredPlantPTS1 server uses a positional weight matrix (PWM) to assign an amino acid at a particular position in the extreme C-terminal sequence of a given protein with a prediction score (Skoulding et al., 2015, supplementary information). The PWM created by the Reumann research group, based on *in silico* predictions, is displayed in **Figure 52**.

	Position														
Res.	**-14**	**-13**	**-12**	**-11**	**-10**	**-9**	**-8**	**-7**	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**	**Res.**
A	-0.1	-0.1	-0.1	0.01	-0	-0.1	-0.1	-0.1	-0.1	-0	-0	0.34	-0.2	-0.2	**A**
R	-0.1	-0.1	0.01	-0.1	-0.1	-0.1	-0.1	-0.1	0.03	-0.1	0.03	-0.2	0.46	-0.2	**R**
N	-0.1	-0.1	-0.1	-0.1	-0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.2	0.01	-0.2	**N**
D	-0.1	-0.1	-0.1	0.02	-0	-0	-0.1	-0.1	-0.1	-0.2	-0.2	-0.2	-0.2	-0.1	**D**
C	-0.1	-0.1	-0.2	-0.1	-0.2	-0.2	-0.1	-0	-0.2	-0.1	-0.2	0.12	-0.1	-0.2	**C**
Q	-0.1	-0.1	-0.1	-0	-0.1	-0.1	0	-0	-0	-0.1	-0	-0.1	-0.1	-0.2	**Q**
E	-0.1	-0.1	-0.1	-0.1	-0.1	-0	-0.2	0	-0.1	-0.1	-0.2	-0.2	-0.1	-0.1	**E**
G	-0.1	-0.1	-0.1	-0	-0.1	-0.1	-0	-0.1	-0.1	-0.1	-0.2	-0.2	-0.2	-0.1	**G**
H	-0.1	-0	-0.1	-0.1	-0	0.01	-0	-0	-0.1	-0.1	0.03	-0.1	-0.1	-0.2	**H**
I	-0	-0.1	-0.1	-0.1	-0.1	-0.1	-0	-0.1	-0	0	-0.1	-0.2	-0.1	0.33	**I**
L	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0	-0	-0.1	-0.2	-0.1	0.66	**L**
K	-0.1	-0.1	-0.1	-0	-0.1	0	-0.1	-0	-0	-0.1	0	-0.1	0.44	-0.2	**K**
M	-0.1	-0	0	-0.1	-0.1	-0.1	-0.1	-0.1	0.07	-0	-0.1	-0.1	-0.1	0.64	**M**
F	-0.1	-0.2	-0.1	-0.1	-0.1	-0	-0.1	-0.2	-0.1	-0	-0.1	-0	-0.2	-0.1	**F**
P	-0.1	-0.1	-0.1	-0.1	0.03	-0.1	-0.1	0.02	-0.1	0.03	0	0.13	-0.2	-0.2	**P**
S	-0.1	-0.1	-0.1	-0	-0.1	-0.1	0	-0.1	-0	-0.1	-0.1	0.48	-0.1	-0.2	**S**
T	-0.1	-0.1	-0	-0	0	-0.1	-0.1	-0.1	-0	-0.1	-0.1	-0.1	-0.2	-0.2	**T**
W	0.15	0.15	0	0.01	-0.1	-0.1	-0.2	-0.1	-0.3	-0.2	-0.1	-0.3	-0.2	-0.2	**W**
Y	0.01	-0	-0.1	-0.3	-0.1	0.02	-0.1	-0.1	0	-0.1	0.01	-0.1	-0.1	-0.2	**Y**
V	-0.1	-0	-0	-0.1	0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.2	-0.1	**V**
	-14	**-13**	**-12**	**-11**	**-10**	**-9**	**-8**	**-7**	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**	

Figure 52| Positional weight matrix (PWM) used by the Reumann research laboratory. This is used to assign a PTS1 prediction score to each amino acid in the 14 C-terminal amino acids of a protein. (Reumann et al., 2012; Skoulding et al., 2015).

A comparison between the data obtained from the pull-down-LC-MS results with the PredPlantPTS1 server was carried out. The positional weight matrix (PWM) algorithm used by PredPlantPTS1 was applied to each of the peptide sequences in the full library screened with wild-type *At*PEX5-C. The way in which PWM score is assigned to a peptide sequence is by addition of the individual amino acid scores at each position, as determined by the PWM (**Figure 52**). A graph was plotted to show the peptides ranked by 'PWM score' and to highlight which peptides were found by the pull-down-LC-MS screen (**Figure 53**).

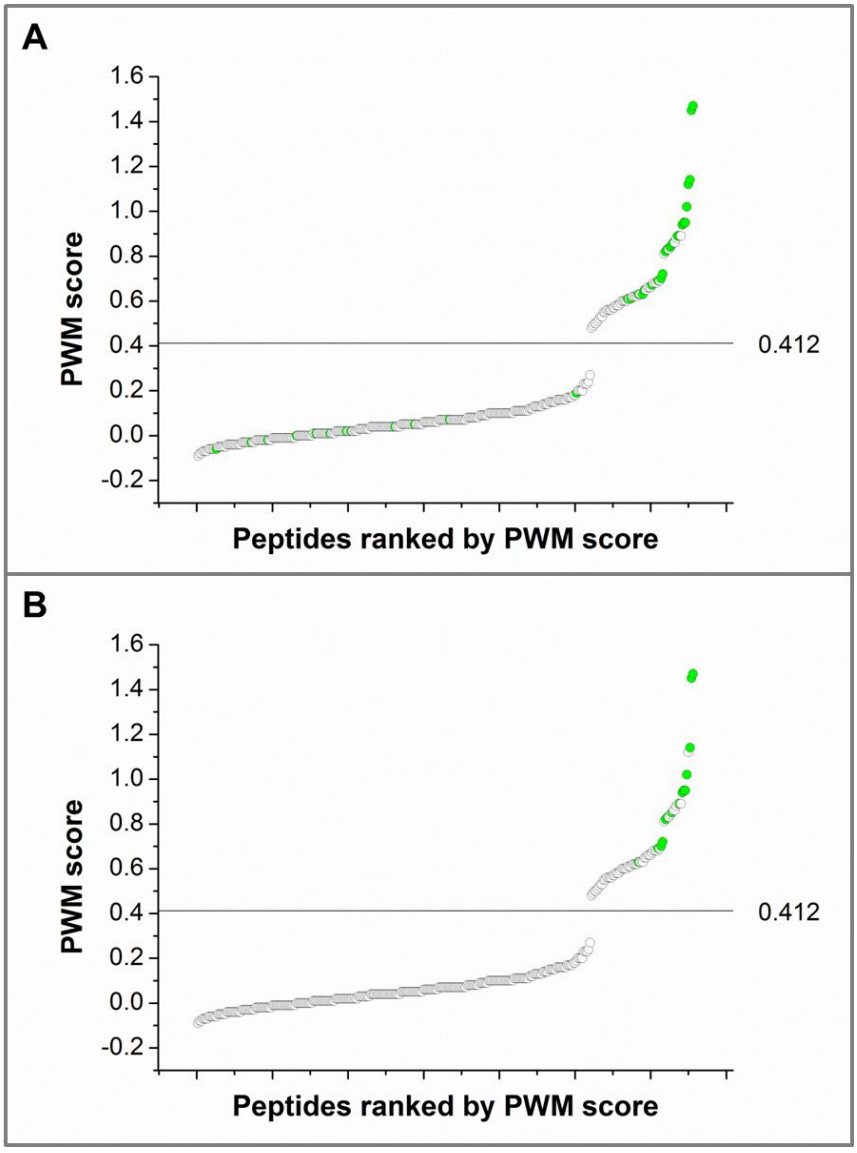


Figure 53| (A) and (B) All peptides in the full library ranked by positional weight matrix (PWM) score. A) In green are peptides identified in figure 17B. B) In green are peptides identified in figure 17B, excluding those with an identical ReT-mass identifier to another peptide, and excluding those with a C-terminal histidine. A higher PWM score designates that the peptide sequence is more likely to act as a PTS1 sequence. A threshold score for peroxisomal import is given by PredPlantPTS1 as 0.412 and this is shown on graphs (A) and (B).

By ranking peptide sequences by their PWM score, it was possible to determine whether the expected peptide sequences were being pulled down by wild-type *At*PEX5-C. Overall, it appears that the peptides that were identified by pull-down-LC-MS were some of the highest PWM scoring peptides. This conclusion was made clearer when peptides with a C-terminal histidine and all peptides without unique ReT-mass identifiers were removed.

In order to study the peptides identified by the pull-down-LC-MS screen by comparison with more recent research, peptides in sub-library 1.2 that were predicted to be canonical PTS1 sequences irrespective of their PWM scores, based on experimental validation of each amino acid at each of the three C-terminal positions (Chowdhary et al., 2012), were recorded. This allowed the study of only predicted canonical PTS1 sequences in the context of the pull-down-LC-MS screen. These PTS1 sequences were then sorted by PWM score (Reumann et al., 2012). Of these sequences, the peptide sequences identified in the pull-down-LC-MS screen were then highlighted in green (**Figure 54**).

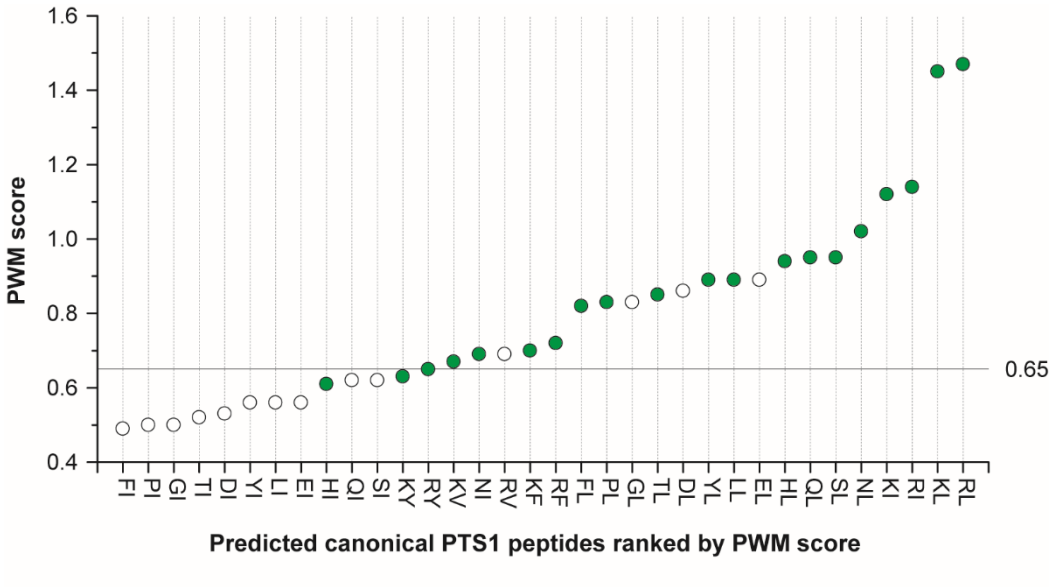


Figure 54| Predicted canonical PTS1 sequences in the full peptide library ranked according to their positional weight matrix (PWM) score. Displayed on the x-axis are the two C-terminal residues from each pentapeptide sequence. Upstream residues are H_2N-Y-Q-S- for all peptides. In green are peptides identified in the pull-down-LC-MS screen. The PWM threshold value for the pull-down-LC-MS screen was predicted as approximately 0.65 (shown on graph).

The conclusion from analysis shown in **Figure 54** was that the PWM score threshold for the screen was approximately 0.65, although it should be noted that YQSKY and

YQSRY are likely to be underscored according to PredPlantPTS1 (Skoulding et al., 2015). This is because the set of plant PTS1 sequences used to train PredPlantPTS1 did not contain any PTS1s with a C-terminal tyrosine, as this is a relatively new discovery in plants (Waller et al., 2010; Skoulding et al., 2015). A PWM score of 0.65 represents a moderate PTS1 sequence so this appeared to be a good threshold to have for finding the orthogonal PEX5*:PTS1* interaction. The screen, therefore, was able to identify a range of binding peptides which should be able to act as moderate and strong PTS1 sequences.

3.6 Summary

A pull-down-LC-MS screen was successfully developed and optimised, and was used to identify which peptides from a library were binding to a purified protein. The pull-down-LC-MS screen was carried out with wild-type *At*PEX5-C against the full peptide library synthesised in section 3.2.1. This screen correlates with *in silico* work which is based on a number of *in vivo* and *in vitro* PEX5:PTS1 studies (Lingner et al., 2011; Reumann et al., 2012; Skoulding et al., 2015). It is also a relatively recent discovery that the -2 residue in the PTS1 sequence is much more flexible than previously thought, with the previous knowledge being that only a basic residue could be accepted at this position (Chowdhary et al., 2012). The results from the pull-down-LC-MS screen developed in this study agree with this finding. At this point it was confirmed that the screen was producing reliable results, so the next stage was to use the pull-down-LC-MS screen to investigate the peptide-binding properties of a range of *At*PEX5-C variants.

Chapter 4

Pull-Down-LC-MS Screening of *At*PEX5-C Variants, and Quantitative Testing and Optimisation of *At*PEX5-C* Binding to PTS1*

4.1 Introduction

A selective pull-down-LC-MS screen had been established (Chapter 3) which could be used to successfully identify PTS1 sequences when using wild-type *At*PEX5-C as the target protein. This screen could now be used to aid in the development of an orthogonal *At*PEX5-C:PTS1 binding pair (termed *At*PEX5-C*:PTS1* in this study). Using the approach of screening a large library of peptides against a small library of protein variants, it could be possible to identify whether any mutations in *At*PEX5-C lead to a significant enough change in the peptide-binding profile to begin to define the AtPEX5-C* protein.

This chapter will describe the selection and production of various *At*PEX5-C protein mutants to use as a small protein library, and the screening of these *At*PEX5-C variants using the developed pull-down-LC-MS screen (Chapter 3). A summary of pull-down-LC-MS screening results will be presented, and the combination of mutations in certain *At*PEX5-C variants, to define a preliminary *At*PEX5-C*:PTS1* pair, will be explained. Also described will be approaches used to further enhance *At*PEX5-C* binding to the preliminary PTS1* sequence, including the calculation of dissociation constants for the interaction, further mutagenesis of the *At*PEX5-C* protein, and the addition of upstream residues to the preliminary PTS1*. This chapter will conclude with the *in vivo* testing of the optimised binding pair, which was carried out by Dr Rupesh Paudyal (A. Cuming laboratory, University of Leeds).

4.2 Screening of a small library of *At*PEX5-C variants using pull-down coupled with LC-MS

A small library of protein variants was produced by site-directed mutagenesis (Chapter 2, **Figure 23**; Chapter 6 (Experimental) section 6.5). Following over-expression, proteins were purified using affinity chromatography and buffer-exchanged ready for analysis. These protein variants were then tested with the synthesised dansyl-coupled peptide library. In the creation of the protein variant library, mutations were targeted to the PTS1-binding site of the *At*PEX5-C protein (**Figure 55**). Some protein variants had already been produced (Chapter 2, section 2.5), so these were also expressed and purified for screening.

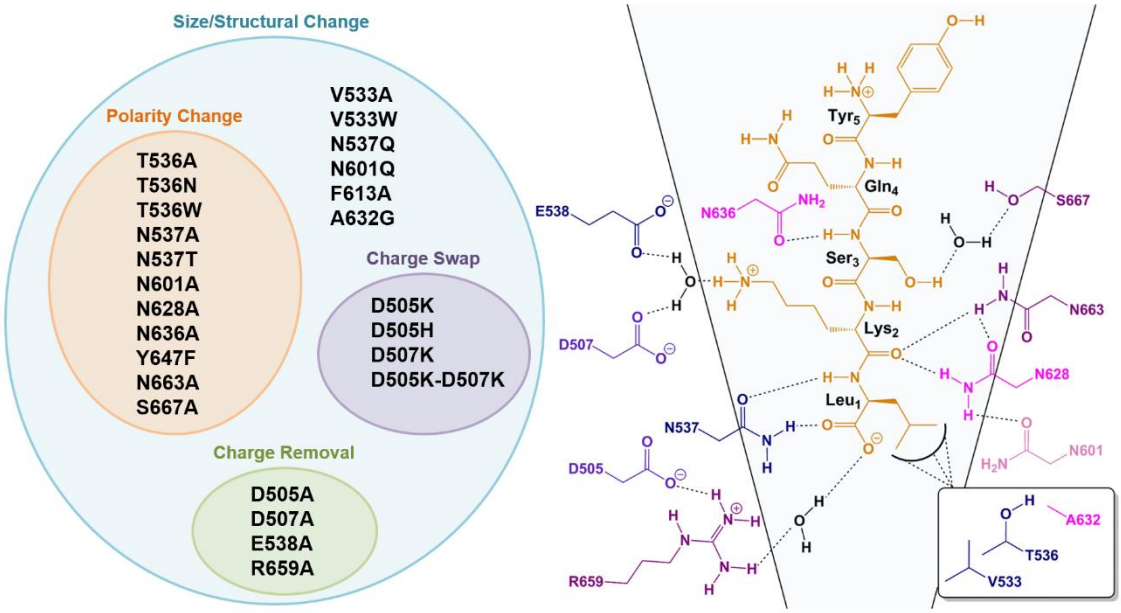


Figure 55| Mutations made in *At*PEX5-C for pull-down-LC-MS screening. The mutations are grouped according to the change they provided in the binding site, and a plot of predicted interactions of some of these protein residues towards a representative PTS1 (YQSKL) is shown for review. Residue numbers shown on the interaction plot above are representative of *A. thaliana* PEX5 residue numbers.

The *At*PEX5-C variants (**Figure 55**) were characterised by mass spectrometry (Appendix C) before the peptide-binding preferences were determined by screening with the previously developed pull-down-LC-MS method (Chapter 3). Pull-down-LC-MS screens were performed as previously described (Chapter 3, **Figure 51, B**). In total, 25 protein variants were analysed for their binding properties to the dansyl-labelled library of peptides (**Figure 55**), and four typical classes of behaviour were

seen. These four classes of behaviour are shown in **Figure 56**. All other *At*PEX5-C variant screen results are presented in Appendix E.

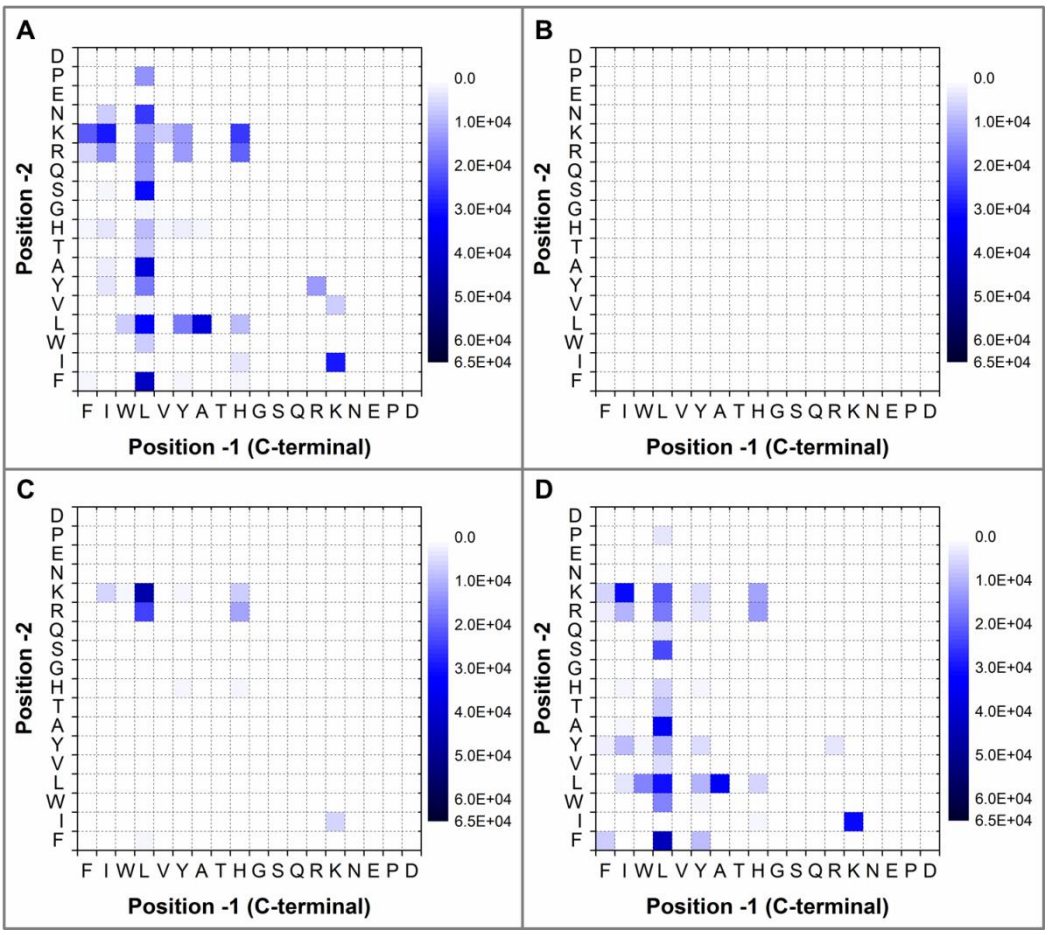


Figure 56| Pull-down-LC-MS screen results for four *At*PEX5-C variants, representing the four types of binding seen. A) S667A, similar peptide-binding profile to wild-type *At*PEX5-C. B) D505K, no peptides far above background level pulled down. C) F613A, only the strongest PEX5-binding peptides pulled down. D) D505H, binding to non-PTS1 peptide sequences seen. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

The 25 protein variants tested were grouped according to the four main binding patterns, and the result of these groupings is shown in **Table 10**. Group 1 (**Figure 56, A**) contains protein variants that pulled down the same peptides, generally, as wild-type *At*PEX5-C. Group 2 (**Figure 56, B**) contains protein variants that showed the same peptide-binding profile of the 'peptide only' control, so showed a background intensity of peptide pull-down. Group 3 (**Figure 56, C**) contains protein variants that showed pull-down of only the strongest PTS1 sequences, as defined by PredPlantPTS1 (Lingner et al., 2011; Reumann et al., 2012). Group 4 (**Figure 56, D**) contains protein variants that showed pull-down of non-PTS1 sequences, as defined by PredPlantPTS1 (Lingner et al., 2011; Reumann et al., 2012).

Group based on peptide-binding changes			
1. Similar profile to wild-type AtPEX5-C	***2. Background intensity of peptides pulled down***	***3. Only strong native PTS1s pulled down***	***4. Pull-down of non-PTS1 peptides seen***
D505A	*D505K*	*V533A*	*D505H*
D507A	*D505K-D507K*	*V533W*	*T536W*
D507K	*N537A*	*T536N*	*N601A*
T536A	*N628A*	*N537T*	
E538A	*R659A*	*N537Q*	
N601Q	*N663A*	*F613A*	
Y647F		*A632G*	
S667A		*N636A*	

Table 10| Summary of the peptide binding changes seen with each *At*PEX5-C variant. Proteins are grouped into four categories based on their apparent peptide-binding profile.

*At*PEX5-C protein variants in group 1 (**Table 10**) showed a similar peptide-binding profile to wild-type *At*PEX5-C. After sequence conservation analysis of PEX5 proteins across different organisms (**Figure 57**), it was found that none of the residues targeted in group 1, with the exception of Y647, were conserved across species. This suggested that these residues were not essential for the function of the protein, and so it was not particularly surprising that change was tolerated at each of these positions.

*At*PEX5-C protein variants in group 2 (**Table 10**) showed a very low level of peptide pull down, in a similar way to the peptide-only background control screen. Residues N537, N628, R659, and N663 show complete conservation across species (**Figure 57**). Residues D505 and D507 exhibit conservation of a negative charge, as only acidic residues are found in these positions. This could explain why peptide binding is abolished with mutation of residues 537, 628, 659, and 663 to alanine, whereas peptide binding is only abolished with mutation of residues 505 and 507 from an acidic to a basic residue.

Interestingly, *At*PEX5-C protein variants in group 3 (**Table 10**) showed pull-down of only the strongest PTS1 sequences, largely dansyl-YQSKL and dansyl-YQSRL. Changes made in the group 3 *At*PEX5-C variants, with the exception of N636A, are related to size and not property changes, so the maintenance of some PTS1 peptide binding in these variants could signify a compaction of the PTS1-binding site upon

PTS1 binding. This was seen by Fodor and colleagues, as mutation of C-terminal PTS1 residues of the protein AGT to smaller residues led to compensatory compaction of the PTS1-binding site, as confirmed by crystal structures (Fodor et al., 2015).

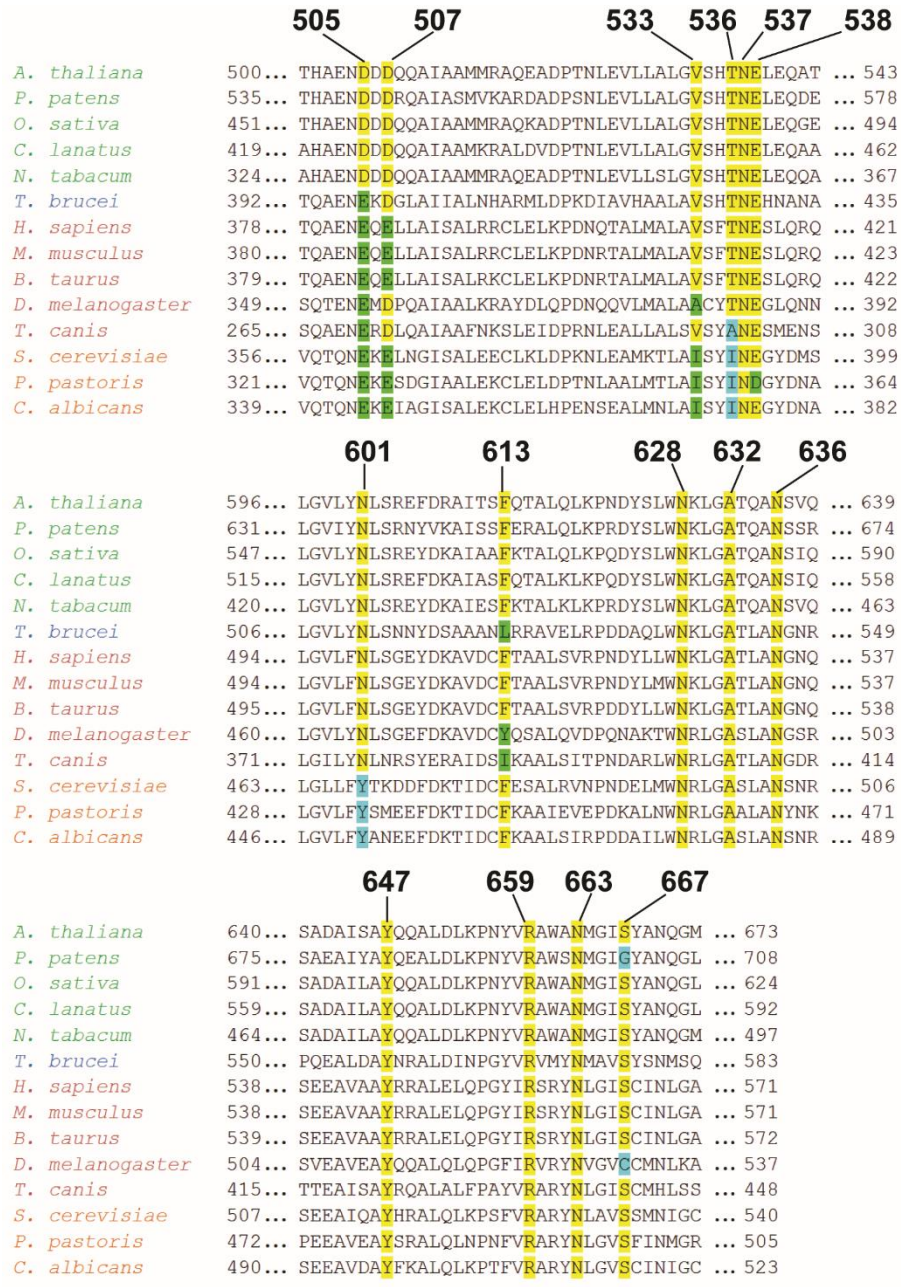


Figure 57| Sequence conservation analysis of the highlighted residues by alignment of PEX5 proteins. Highlighted in yellow are matching residues, highlighted in green are residues with similar properties, and highlighted in cyan are residues with differing properties. Alignments were performed using BioEdit (sequence alignment editor). Organisms are colour-coded based on kingdom: green text, Plantae; blue text, Excavata; pink text, Animalia; orange text, Fungi. Numbers above the alignments correspond to the *At*PEX5 residue numbers.

Excitingly, three *At*PEX5-C variants (D505H, T536W, and N601A) all pulled down non-PTS1 sequences (**Table 10, group 4**). D505H pulled down dansyl-YQSFY and dansyl-YQSYY, and T536W pulled down dansyl-YQSKA and dansyl-YQSRA. The two residues were also reasoned to be far enough away from one another in the binding site that mutation of one should not have an adverse effect on the other. Wild-type *At*PEX5-C did not pull down dansyl-YQSFY, dansyl-YQSYY, dansyl-YQSKA, or dansyl-YQSRA. In an effort to enhance the pull-down of these non-PTS1 sequences, the mutations D505H and T536W were combined, resulting in the protein variant D505H-T536W. Pull-down-LC-MS screen results from D505H, T536W, and D505H-T536W are shown in **Figure 58**.

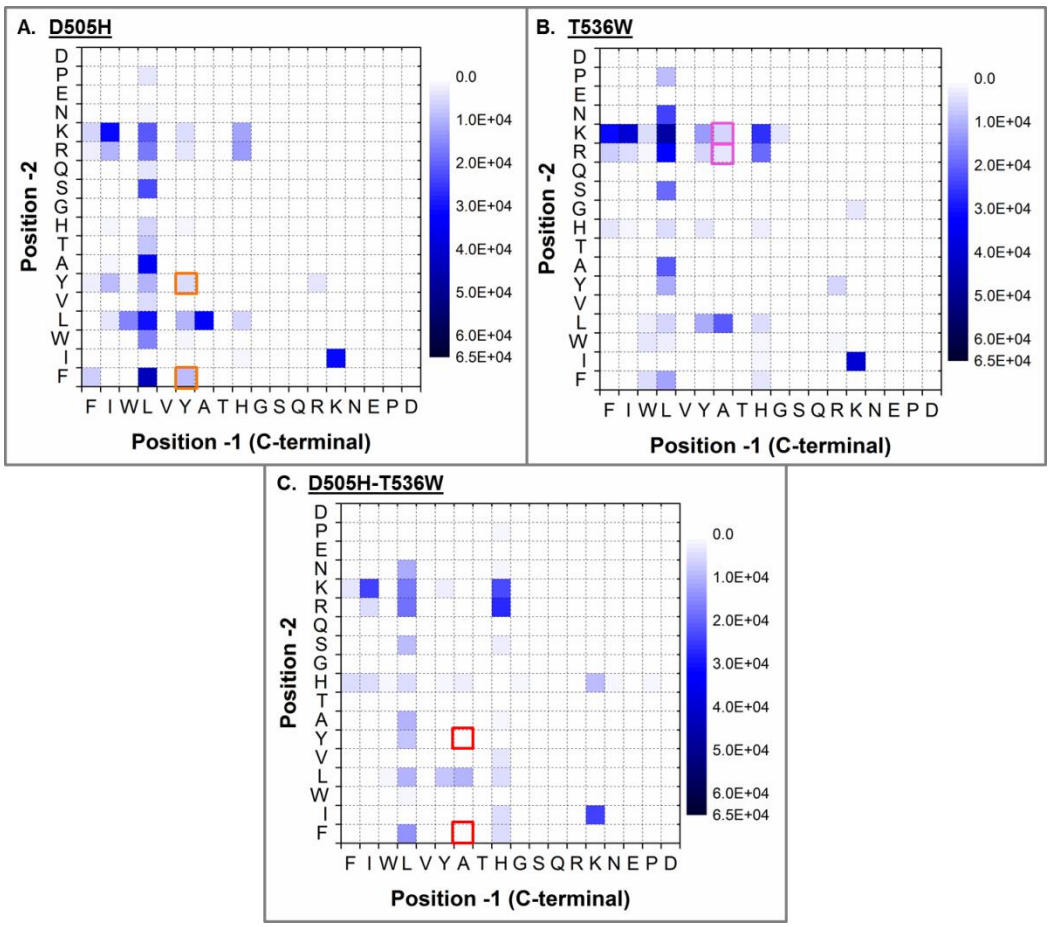


Figure 58| Pull-down-LC-MS screen results for three *At*PEX5-C variants, two single variants (D505H and T536W) and a variant combining the two mutations. A) D505H, non-PTS1 peptide sequences dansyl-YQSFY and dansyl-YQSYY are pulled down. B) T536W, non-PTS1 peptide sequences dansyl-YQSKA and dansyl-YQSRA are pulled down. C) D505H-T536W, similar peptide-binding profile to wild-type *At*PEX5-C. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS. Peptides of interest are highlighted by square outlines.

The combination of mutations D505H and T536W did not enhance the pull-down of peptides with a C-terminal alanine, as was hoped for (**Figure 58, C**). D505H and N601A both showed pull-down of the sequences dansyl-YQSFY and dansyl-YQSYY, which was not seen for wild-type *At*PEX5-C. As both D505H and N601A pulled down the same non-PTS1 sequences, the two mutations in the protein were combined to create D505H-N601A in the hope that pull-down of dansyl-YQSFY and dansyl-YQSYY would be enhanced. Pull-down-LC-MS screen results from D505H, N601A, and D505H-N601A are shown in **Figure 59**.

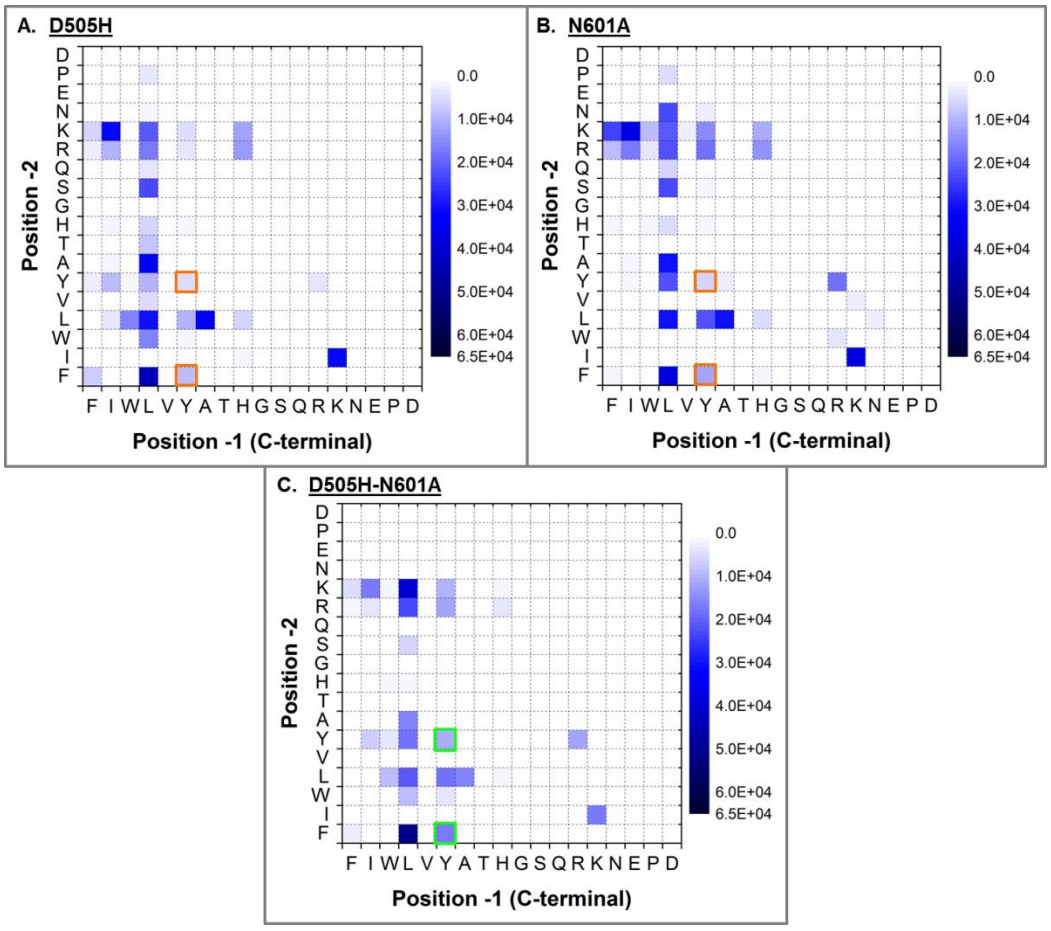


Figure 59| Pull-down-LC-MS screen results for three *At*PEX5-C variants, two single mutants (D505H and N601A) and a variant combining the two mutations. A) D505H, non-PTS1 peptide sequences dansyl-YQSFY and dansyl-YQSYY are pulled down. B) N601A, the same non-PTS1 peptide sequences dansyl-YQSFY and dansyl-YQSYY are pulled down. C) D505H-N601A, non-PTS1 peptide sequences dansyl-YQSFY and dansyl-YQSYY appear to be pulled down more strongly than with either of the single variant *At*PEX5-C proteins. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS. Peptides of interest are highlighted by square outlines.

The combination of mutations D505H and N601A did appear to enhance the pull-down of peptides dansyl-YQSFY and dansyl-YQSYY (**Figure 59, C**). A structural model was

subsequently produced, from the I-TASSER-predicted model of *At*PEX5-C (Chapter 1, **Figure 9**) with the peptide from crystal structure 1FCH (Gatto et al., 2000) in its PTS1-binding site. The structural homology model was produced in an attempt to rationalise the observation that the double variant D505H-N601A exhibited enhanced pull-down of aromatic peptides. Results of pull-down-LC-MS screens for wild-type *At*PEX5-C and for *At*PEX5-C D505H-N601A are shown in **Figure 60**, along with structural modelling of the interactions that were predicted to occur.

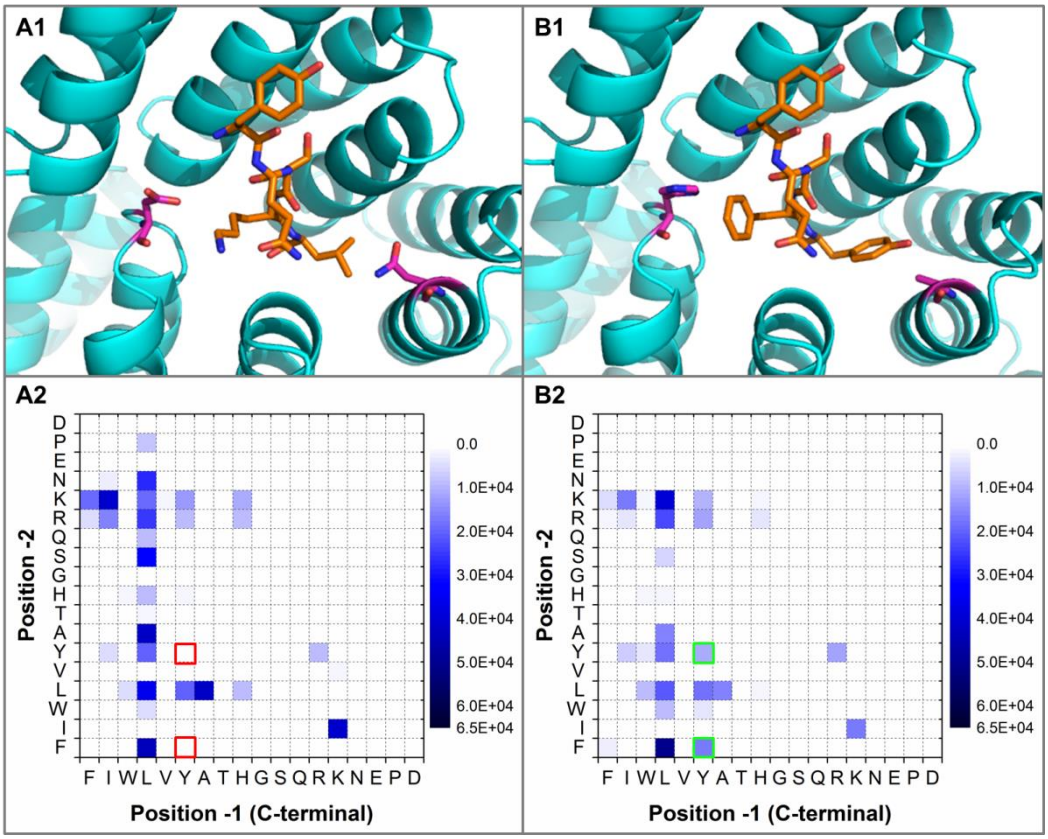


Figure 60| Structural models and pull-down-LC-MS screen results for wild-type *At*PEX5-C and an *At*PEX5-C variant with mutations combined (D505H-N601A) as a result of the initial screening. A1) Structural model of YQSKL (1FCH) bound to wild-type *At*PEX5-C. A2) Pull-down-LC-MS screen result for wild-type *At*PEX5-C. Dansyl-YQSFY and dansyl-YQSYY are not pulled down. B1) Structural model of YQSFY bound to *At*PEX5-C D505H-N601A. B2) Pull-down-LC-MS screen result for D505H-N601A. Dansyl-YQSFY and dansyl-YQSYY are pulled down. (Zhang, 2008; Roy et al., 2010; Yang et al., 2015; Yang and Zhang, 2015 (I-TASSER)) (Gatto et al., 2000 (1FCH peptide)). Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS. Peptides of interest are highlighted by square outlines.

It appeared that the interactions alluded to in the pull-down screen were theoretically possible, as predicted through structural modelling (**Figure 60**). The D505H mutation, through the removal of a negative charge and the introduction of an aromatic residue,

could allow π-stacking to occur with an aromatic residue at position -2 of the binding peptide. The N601A mutation, through enlargement of the PTS1-binding pocket of *At*PEX5-C, could allow a larger hydrophobic residue such as tyrosine to be accommodated at the C-terminus of the binding peptide.

It had been established throughout this study that an *At*PEX5-C variant with a double mutation, D505H-N601A, displayed apparent enhanced binding to peptide sequences dansyl-YQSFY and dansyl-YQSYY. The next stage was to validate the results obtained by pull-down-LC-MS, using fluorescence anisotropy, and to attempt further optimise the binding of *At*PEX5-C D505H-N601A to YQSFY and YQSYY. We also now had a pull-down-LC-MS screen at our disposal to further optimise binding.

4.3 Quantitative binding analysis using fluorescence anisotropy to assess the reliability of results obtained by pull-down-LC-MS screening

Increased binding to two aromatic peptides, dansyl-YQSFY and dansyl-YQSYY, had been found when mutations D505H and N601A were applied to *At*PEX5-C (section 4.2). In order to test whether this binding was genuine, YQSFY and YQSYY were synthesised with an N-terminal lissamine tag, as was performed for previous peptides (Chapter 2, section 2.3; Chapter 6 (Experimental) section 6.13.4). Characterisation of these peptides is shown in Chapter 6 (Experimental) section 6.18. It was then possible to carry out fluorescence anisotropy with the *At*PEX5-C variants D505H, N601A, and D505H-N601A along with peptides synthesised, in order to compare binding properties of these proteins with wild-type *At*PEX5-C.

Fluorescence anisotropy carried out with lissamine-YQSFY revealed that this peptide appeared to show some level of self-association (high anisotropy values at low protein concentration; data not shown), which made the calculation of dissociation constant (K_d) problematic, and which could present issues within an *in vivo* import system. For these reasons, the peptide YQSFY was abandoned as a candidate PTS1* sequence and lissamine-YQSYY was carried forward for testing with protein variants of interest (**Figure 61**).

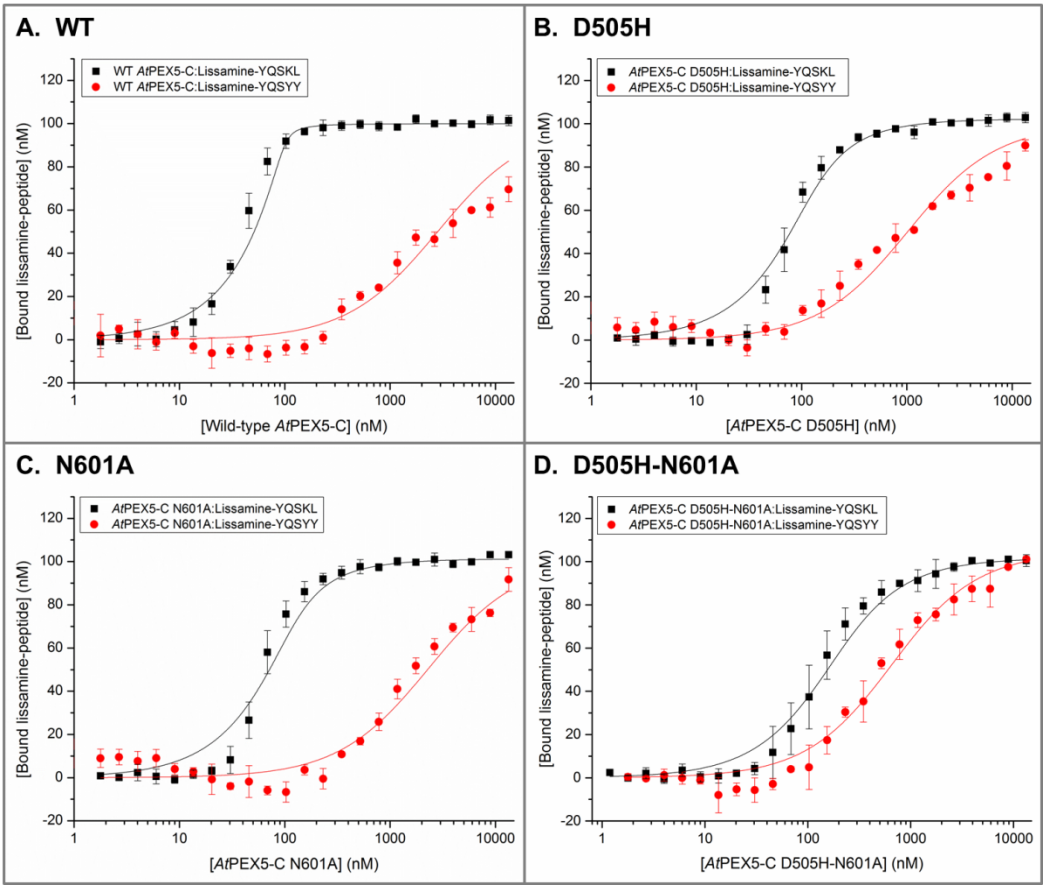


Figure 61| Fluorescence anisotropy results, assessing the binding of lissamine-YQSKL and lissamine-YQSYY (peptides fixed at 100 nM final concentration) to wild-type *At*PEX5-C, and to three variants of *At*PEX5-C (concentration of proteins: 0–13.3 µM final concentration). A) wild-type *At*PEX5-C with lissamine-YQSKL and lissamine-YQSYY. B) *At*PEX5-C D505H with lissamine-YQSKL and lissamine-YQSYY. C) *At*PEX5-C N601A with lissamine-YQSKL and lissamine-YQSYY. D) *At*PEX5-C D505H-N601A with lissamine-YQSKL and lissamine-YQSYY. A quadratic equation was used to fit the curves using OriginPro 9.1, assuming a one-to-one model of binding.

The results from the quantitative binding assays (**Figure 61**) correlated well with the pull down-LC-MS screen results for variant *At*PEX5-C proteins D505H, N601A, and D505H-N601A. The combination of *At*PEX5-C mutations that had previously resulted in enhancement of YQSYY binding as seen in the pull-down-LC-MS screen, did genuinely enhance this binding property. Dissociation constants determined from the fluorescence anisotropy experiments, performed to validate the results of the pull-down-LC-MS screen, are shown in **Table 11**.

At*PEX5-C variant*	*Fluorescent peptide*	*Dissociation constant (K_d)(nM)*	*Standard error (nM)*
Wild-type	*Lissamine-YQSKL*	*1.1*	*0.6*
	Lissamine-YQSYY	*2700*	*200*
D505H	*Lissamine-YQSKL*	*30*	*7.0*
	Lissamine-YQSYY	*960*	*80*
N601A	*Lissamine-YQSKL*	*22*	*8*
	Lissamine-YQSYY	*2300*	*400*
D505H-N601A	*Lissamine-YQSKL*	*108*	*15*
	Lissamine-YQSYY	*603*	*70*

Table 11| Dissociation constants (determined using fluorescence anisotropy) for wild-type *At*PEX5-C and for variants of interest, when tested with lissamine-YQSKL and lissamine-YQSYY. Differences in binding affinities between lissamine-YQSKL and lissamine-YQSYY (the increase seen for lissamine-YQSKL binding over lissamine-YQSYY binding): wild-type *At*PEX5-C ≈ 2454-fold, *At*PEX5-C D505H ≈ 32-fold, *At*PEX5-C N601A ≈ 105-fold, *At*PEX5-C D505H-N601A ≈ 5.6-fold.

Binding of *At*PEX5-C to YQSYY, when compared to YQSKL, had been significantly improved through mutation of two residues in the PTS1-binding site, and this had been found using the pull-down-LC-MS screen developed in this study. When tested quantitatively by fluorescence anisotropy, a ~2454-fold difference in binding affinities had been successfully reduced to a ~5.6-fold difference in binding affinity between lissamine-YQSKL and lissamine-YQSYY, by mutation of the two residues in the PTS1-binding site (D505 and N601). The pull-down-LC-MS screen developed in this study had successfully aided in the finding and enhancement of a non-natural interaction, and the use of this screen as a method of finding new interactions had now been validated through the finding that the *At*PEX5-C variant D505H-N601A did bind to YQSYY much more strongly than wild-type *At*PEX5-C bound to this peptide.

Unfortunately, the *At*PEX5-C variant D505H-N601A was still binding to lissamine-YQSKL with greater affinity than lissamine-YQSYY. To overcome this problem, the approach decided upon was to further mutate the protein, using information about binding tolerance obtained through the developed pull-down-LC-MS screen (section 4.2).

4.4 Further mutation of the preliminary *At*PEX5-C* and subsequent analysis of peptide binding

It had been possible, through pull-down-LC-MS with a small protein library and subsequent combination of mutations of interest, to decrease binding affinity to lissamine-YQSKL by ~98-fold while increasing the binding affinity to lissamine-YQSYY by ~4.5-fold, relative to wild-type *At*PEX5-C. This alteration of binding affinities still resulted in a variant of *At*PEX5-C that bound to lissamine-YQSKL with a ~5.6-fold stronger affinity than it bound to the non-natural PTS1 peptide lissamine-YQSYY. Further mutation of *At*PEX5-C was required, in an attempt to close this gap in binding affinities.

It was reasoned that the histidine at position 505, in the *At*PEX5-C variant D505H-N601A, was allowing possible ring interactions to occur with an aromatic side chain of the binding peptide. It could therefore be possible that phenylalanine would work more favourably in position 505. For this reason, *At*PEX5-C D505F-N601A was produced. After performing the pull-down-LC-MS screening (section 4.2), it had been possible to deduce that certain residues were non-essential for the peptide binding of the PTS1-pocket of *At*PEX5-C as PTS1 binding still occurred after mutation of these residues. One of these residues, D507, is in close proximity to D505, and mutation of D507 to alanine had little effect on peptide binding profile (section 4.2, and Appendix E). This meant that the side chain of D507 was not essential for peptide binding and so it could be altered. The favourable position of residue 507, in close proximity to residue 505, also made this amino acid an obvious choice for mutation with aromatic residues. It was therefore decided to produce the *At*PEX5-C variants D505H-D507H-N601A, and D505F-D507F-N601A.

The three *At*PEX5-C variants (D505F-N601A, D505H-D507H-N601A, and D505F-D507F-N601A) were expressed, purified using cobalt affinity chromatography, and buffer-exchanged for analysis. Mass spectrometry was carried out for characterisation of proteins (Appendix C). The *At*PEX5-C variants were then screened using pull-down-LC-MS (**Figure 62**), and quantitative binding data were obtained after testing with lissamine-YQSKL and lissamine-YQSYY by fluorescence anisotropy (**Figure 63**).

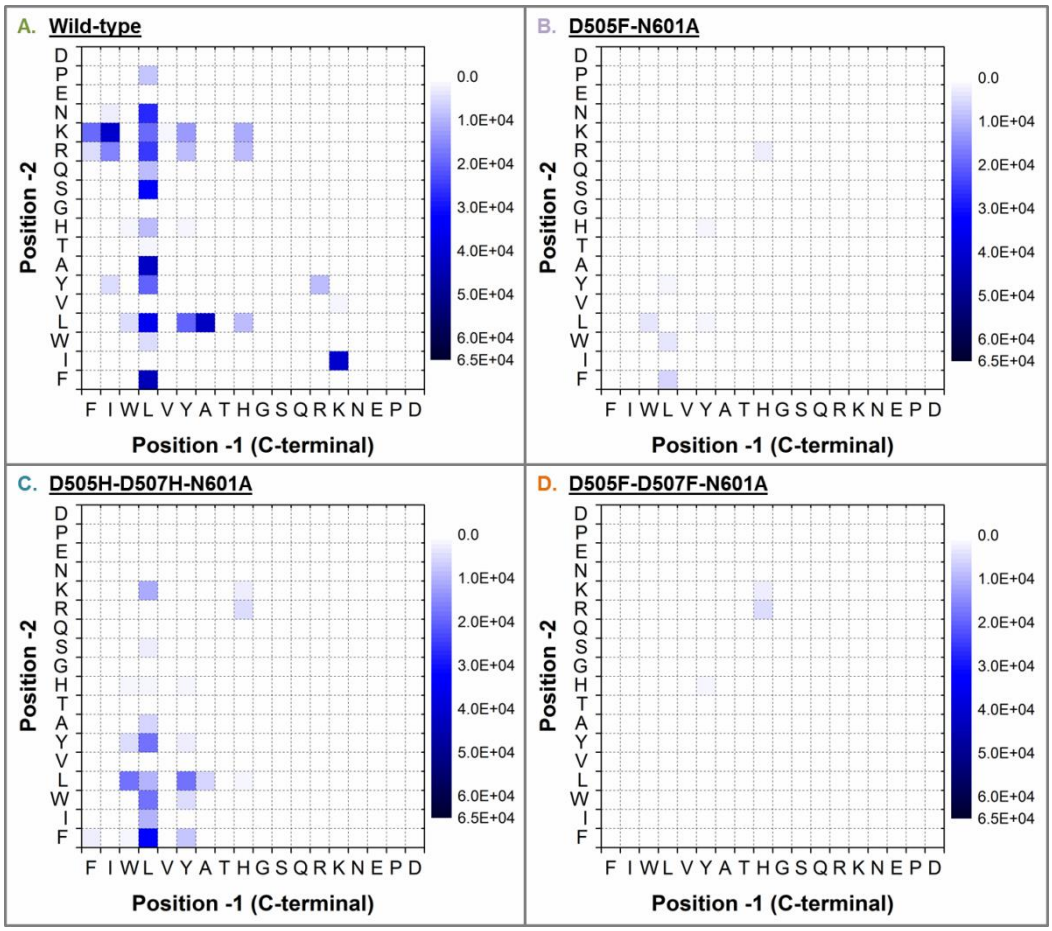


Figure 62| Pull-down-LC-MS screen results for three new variants of *At*PEX5-C, compared to that of wild-type *At*PEX5-C. A) Wild-type *At*PEX5-C. B) D505F-N601A, minimal evidence of peptide pull-down is observed and peptides identified are seen in wild-type *At*PEX5-C pull-down-LC-MS. C) D505H-D507H-N601A, the peptide-binding profile appears to be changed from that of wild-type AtPEX5-C. D) D505F-D507F-N601A, minimal evidence of any peptide pull-down is observed. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

It is apparent from **Figure 62** that the peptide-binding profile of *At*PEX5-C is significantly altered as a result of the mutations made. The *At*PEX5-C variant of interest as a result of these screens appeared to be *At*PEX5-C D505H-D507H-N601A. Where residues have been substituted with phenylalanine, the resulting peptide-binding profile is very similar to that of a background intensity of peptides pulled down in the absence of protein (section 4.2: **Table 10**) and so these mutants are likely to be inactive.

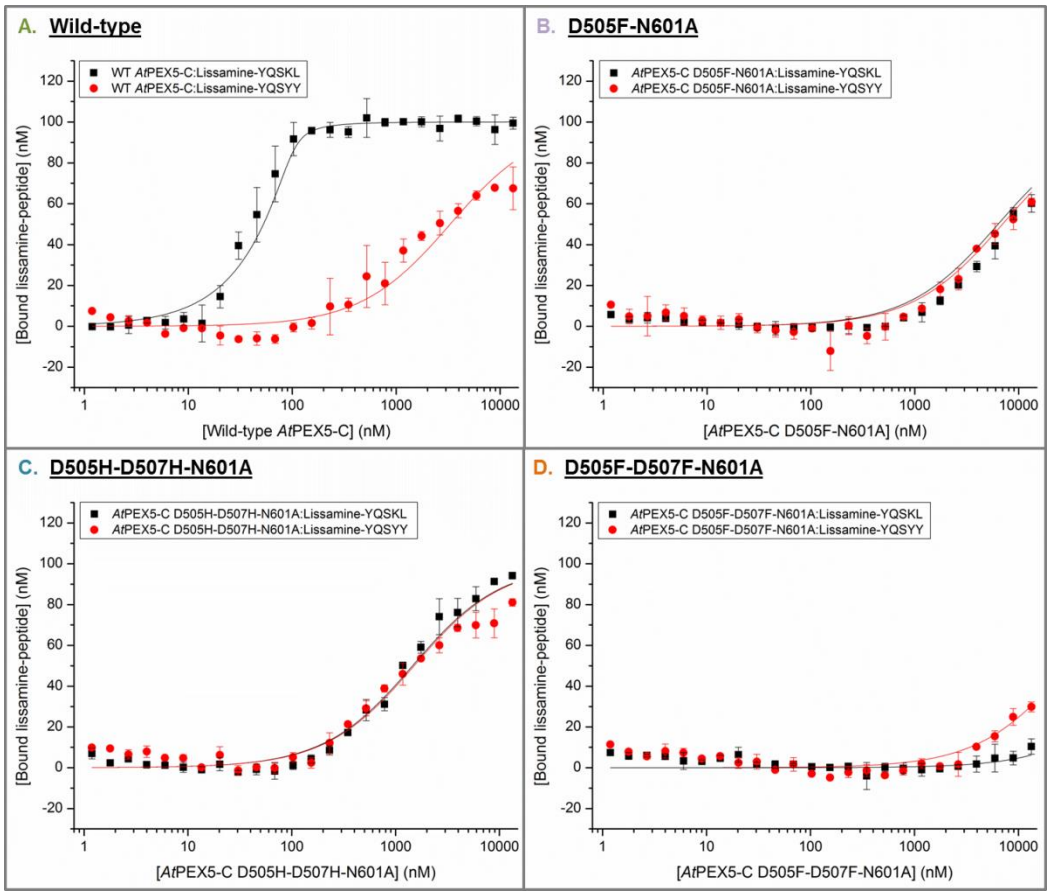


Figure 63| Fluorescence anisotropy results after rational mutation of *At*PEX5-C, assessing the binding of lissamine-YQSKL and lissamine-YQSYY (peptides fixed at 100 nM final concentration) to wild-type *At*PEX5-C, and to three variants of *At*PEX5-C (concentration of proteins: 0–13.3 µM final concentration). A) wild-type *At*PEX5-C with lissamine-YQSKL and lissamine-YQSYY. B) *At*PEX5-C D505F-N601A with lissamine-YQSKL and lissamine-YQSYY. C) *At*PEX5-C D505H-D507H-N601A with lissamine-YQSKL and lissamine-YQSYY. D) *At*PEX5-C D505F-D507F-N601A with lissamine-YQSKL and lissamine-YQSYY. A quadratic equation was used to fit the curves using OriginPro 9.1, assuming a one-to-one model of binding.

Mutation of D507 did result in a change in peptide binding profile of *At*PEX5-C. For the *At*PEX5-C variants D505F-N601A and D505H-D507H-N601A, lissamine-YQSKL binding and lissamine-YQSYY binding appear to be almost identical in affinity. For the *At*PEX5-C variant D505F-D507F-N601A, the binding affinity for lissamine-YQSYY is stronger than that for YQSKL; however, the binding affinities are very weak. For all of these variants, their binding affinity to lissamine-YQSYY is weaker than the binding affinity of YQSKV to wild-type *At*PEX5-C, which has previously been shown not to import into moss peroxisomes *in vivo* (Chapter 2, **Figure 20** (R. Paudyal)). Even for the variant with the strongest binding affinities to lissamine-YQSKL and lissamine-YQSYY (D505H-D507H-N601A), these binding affinities are 1400 ± 100 nM for lissamine-YQSKL and 1500 ± 600 nM for lissamine-YQSYY, whereas the binding

affinity of lissamine-YQSKV to wild-type *At*PEX5-C was stronger than both of these at 1300 ± 100 nM.

One possibility was that a histidine or phenylalanine at position 507 of *At*PEX5-C increased steric bulk and partially hindered the binding of any peptides. Valine and threonine were therefore trialled in position 507 to analyse whether this would change the peptide binding profile in the desired way, without reducing overall binding affinity beyond the approximate level needed for peroxisomal protein import. The resulting *At*PEX5-C variants decided upon were D505H-D507V-N601A and D505H-D507T-N601A. These proteins were expressed, purified, buffer-exchanged, and characterised by mass spectrometry (Appendix C) before analysis. The results of pull-down-LC-MS and fluorescence anisotropy with these two *At*PEX5-C variants are displayed in **Figure 64**.

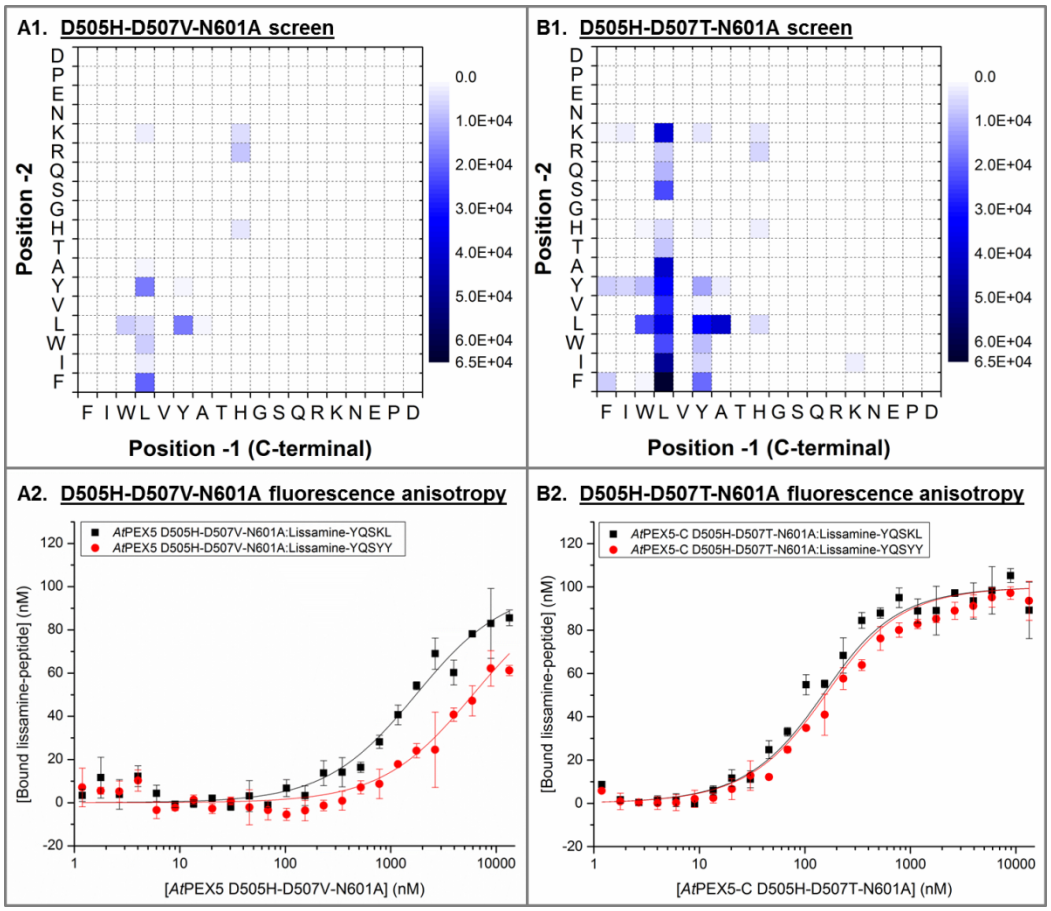


Figure 64| Pull-down-LC-MS results and fluorescence anisotropy data (100 nM fixed peptide final concentration, with concentration of proteins ranging from 0 to 13.3 µM final concentration) for two new *At*PEX5-C variants. A1) D505H-D507V-N601A pull-down-LC-MS. B1) D505H-D507T-N601A pull-down-LC-MS, showing strong pull-down of dansyl-YQSYY. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS. A2) D505H-D507V-N601A fluorescence anisotropy with lissamine-YQSKL and lissamine-YQSYY. B2) D505H-D507T-N601A fluorescence anisotropy with lissamine-YQSKL and lissamine-YQSYY. A quadratic equation was used to fit the curves, assuming a one-to-one model of binding.

Of the two new *At*PEX5-C variants (**Figure 64**), D505H-D507T-N601A appeared to have the optimal peptide binding profile for this study as binding affinities were similar with both lissamine-YQSKL and lissamine-YQSYY. This meant that *At*PEX5-C D505H-D507T-N601A may allow the import of the non-PTS1 sequence YQSYY, albeit with continued import of very strong PTS1 sequences. A strategy was in place for the improving of the strength of the current preliminary PTS1* sequence YQSYY, using upstream residues (section 4.5), so *At*PEX5-C D505H-D507T-N601A was termed *At*PEX5-C* from this point onwards.

The resulting mutations which had been made in *At*PEX5-C* are shown on a structural model in **Figure 65**, with properties of all residues of interest highlighted by the colour

of the residues and the corresponding surface. Circular dichroism, for comparison with wild-type *At*PEX5-C, and the mass spectrum with deconvoluted mass data of the *At*PEX5-C* protein are also shown in **Figure 65**.

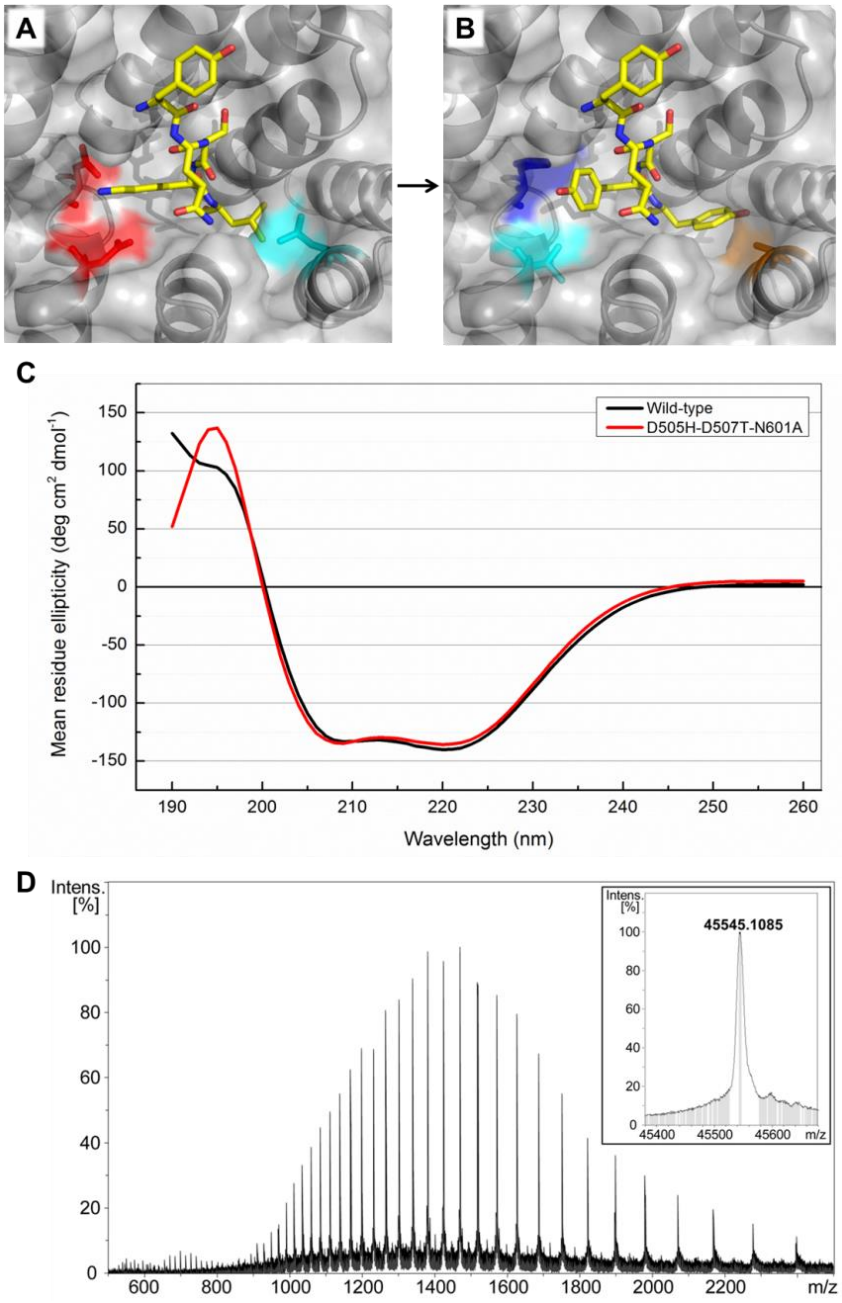


Figure 65| Structural models of wild-type *At*PEX5-C in complex with YQSKL (1FCH) (A) and *At*PEX5-C* in complex with YQSYY (B), circular dichroism of wild-type *At*PEX5-C overlaid with that of *At*PEX5-C* (C), and mass spectrum with deconvoluted mass data (inset) for *At*PEX5-C* (D). For structural models (A) and (B), residues coloured red = acidic, those coloured blue = basic, those coloured cyan = polar, and those coloured orange = hydrophobic. (Zhang, 2008; Roy et al., 2010; Yang et al., 2015; Yang and Zhang, 2015 (I-TASSER)) (Gatto et al., 2000 (1FCH peptide)). C) *At*PEX5-C* (D505H-D507T-N601A) has a similar circular dichroism trace and, therefore, secondary structure to wild-type *At*PEX5-C. D) Expected mass for *At*PEX5-C* (D505H-D507T-N601A): 45,545.3 Da. Observed mass for *At*PEX5-C*: 45,545.1 Da.

It was clear, from **Figure 65, C**, that the mutations applied to create *At*PEX5-C* had not given rise to misfolding of the protein at a secondary structure level. Both circular dichroism traces, that of wild-type *At*PEX5-C and of *At*PEX5-C*, displayed troughs at 208 nm and 222 nm which showed that both proteins adopt an α-helical structure. Dissociation constants (K_d values), determined by fluorescence anisotropy for wild-type *At*PEX5-C, *At*PEX5-C D505H-N601A, and *At*PEX5-C* (D505H-D507T-N601A) are shown in **Table 12** to highlight the gradual improvement seen in relative binding affinities of the protein for lissamine-YQSKL and lissamine-YQSYY through mutation of the protein.

***At*PEX5-C variant**	***Fluorescent peptide***	***Dissociation constant (K_d)(nM)***	***Standard error (nM)***
Wild-type	*Lissamine-YQSKL*	*1.1*	*0.6*
	Lissamine-YQSYY	*2700*	*200*
D505H-N601A	*Lissamine-YQSKL*	*108*	*15*
	Lissamine-YQSYY	*603*	*70*
D505H-D507T-N601A	*Lissamine-YQSKL*	*97*	*20*
	Lissamine-YQSYY	*110*	*50*

Table 12| Dissociation constants (determined using fluorescence anisotropy) for wild-type *At*PEX5-C and for new variants of interest, when tested with lissamine-YQSKL and lissamine-YQSYY. Differences in binding affinities between lissamine-YQSKL and lissamine-YQSYY (the increase seen for lissamine-YQSKL binding over lissamine-YQSYY binding): wild-type *At*PEX5-C ≈ 2454-fold, *At*PEX5-C D505H-N601A ≈ 5.6-fold, *At*PEX5-C D505H-D507T-N601A ≈ 1.1-fold.

Dissociation constants (**Table 12**) confirm that a semi-rational approach had been successful in improving the binding affinity of *At*PEX5-C for lissamine-YQSYY by ~25-fold and in decreasing its binding affinity for lissamine-YQSKL by ~88-fold. The resulting *At*PEX5-C variant (termed *At*PEX5-C*) showed a binding affinity for lissamine-YQSKL only ~1.1-fold higher than its binding affinity for lissamine-YQSYY. The next stage of the research was to test these proteins with longer candidate PTS1* peptides, to attempt to strengthen the binding affinity of *At*PEX5-C* for the preliminary PTS1* (YQSYY) by ensuring that a suitable upstream peptide sequence was determined before any *in vivo* work was carried out.

4.5 The addition of upstream residues to the preliminary PTS1* with the aim of altering binding affinity

It is becoming increasingly apparent that the PTS1 sequence can no longer be defined as a simple tripeptide (Mullen et al., 1997; Brocard and Hartig, 2006; Lingner et al., 2011; Chowdhary et al., 2012). The PTS1 has most recently been re-defined as a 14-amino acid C-terminal sequence, with each of these residues affecting the affinity of the overall sequence for the PTS1-binding site of PEX5 (Reumann et al., 2012). Up to this point, the PTS1 sequences being studied were pentapeptides; however, upstream residues had to be considered as these could have a significant effect on the binding affinity (Brocard and Hartig, 2006; Reumann et al., 2012; Chowdhary et al., 2012). The PWM scoring matrix, which can be used to assign a plant peroxisomal import prediction score to a peptide (PredPlantPTS1 (Lingner et al., 2011; Reumann et al., 2012)), was used to predict a range of sequences that could be added upstream of the sequence YQSYY in order to potentially improve binding affinity to *At*PEX5-C*.

The PWM scoring matrix is shown in **Figure 66** to illustrate how a total PWM score of a peptide is calculated. In this example, a route though the matrix is illustrated in **Figure 66** to show the PWM score of an example 14-aa peptide sequence, with the pentapeptide YQSYY at the C-terminus. To highlight the importance of upstream residues in the ability of a sequence to target proteins to the peroxisome, **Figure 66** shows how a short peptide sequence defined as a non-PTS1 (YQSYY) can be transformed into a sequence defined as a PTS1, as determined by its PWM score, merely through the addition of upstream residues.

	Position														
Res.	**-14**	**-13**	**-12**	**-11**	**-10**	**-9**	**-8**	**-7**	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**	**Res.**
A	-0.1	-0.1	-0.1	0.01	-0	-0.1	-0.1	-0.1	-0.1	-0	-0	0.34	-0.2	-0.2	**A**
R	-0.1	-0.1	0.01	-0.1	-0.1	-0.1	-0.1	-0.1	0.03	-0.1	0.03	-0.2	0.46	-0.2	**R**
N	-0.1	-0.1	-0.1	-0.1	-0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.2	0.01	-0.2	**N**
D	-0.1	-0.1	-0.1	0.02	-0	-0	-0.1	-0.1	-0.1	-0.2	-0.2	-0.2	-0.2	-0.1	**D**
C	-0.1	-0.1	-0.2	-0.1	-0.2	-0.2	-0.1	-0	-0.2	-0.1	-0.2	0.12	-0.1	-0.2	**C**
Q	-0.1	-0.1	-0.1	-0	-0.1	-0.1	0	-0	-0	-0.1	-0	-0.1	-0.1	-0.2	**Q**
E	-0.1	-0.1	-0.1	-0.1	-0.1	-0	-0.2	0	-0.1	-0.1	-0.2	-0.2	-0.1	-0.1	**E**
G	-0.1	-0.1	-0.1	-0	-0.1	-0.1	-0	-0.1	-0.1	-0.1	-0.2	-0.2	-0.2	-0.1	**G**
H	-0.1	-0	-0.1	-0.1	-0	0.01	-0	-0	-0.1	-0.1	0.03	-0.1	-0.1	-0.2	**H**
I	-0	-0.1	-0.1	-0.1	-0.1	-0.1	-0	-0.1	-0	0	-0.1	-0.2	-0.1	0.33	**I**
L	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0	-0	-0.1	-0.2	-0.1	0.66	**L**
K	-0.1	-0.1	-0.1	-0	-0.1	0	-0.1	-0	-0	-0.1	0	-0.1	0.44	-0.2	**K**
M	-0.1	-0	0	-0.1	-0.1	-0.1	-0.1	-0.1	0.07	-0	-0.1	-0.1	-0.1	0.64	**M**
F	-0.1	-0.2	-0.1	-0.1	-0.1	-0	-0.1	-0.2	-0.1	-0	-0.1	-0	-0.2	-0.1	**F**
P	-0.1	-0.1	-0.1	-0.1	0.03	-0.1	-0.1	0.02	-0.1	0.03	0	0.13	-0.2	-0.2	**P**
S	-0.1	-0.1	-0.1	-0	-0.1	-0.1	0	-0.1	-0	-0.1	-0.1	0.48	-0.1	-0.2	**S**
T	-0.1	-0.1	-0	-0	0	-0.1	-0.1	-0.1	-0	-0.1	-0.1	-0.1	-0.2	-0.2	**T**
W	0.15	0.15	0	0.01	-0.1	-0.1	-0.2	-0.1	-0.3	-0.2	-0.1	-0.3	-0.2	-0.2	**W**
Y	0.01	-0	-0.1	-0.3	-0.1	0.02	-0.1	-0.1	0	-0.1	0.01	-0.1	-0.1	-0.2	**Y**
V	-0.1	-0	-0	-0.1	0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.2	-0.1	**V**
	-14	**-13**	**-12**	**-11**	**-10**	**-9**	**-8**	**-7**	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**	
Total PWM score of YQSYY = **0.07**															
Total PWM score of WWRDPYSPMYQSYY = **0.54**															

Figure 66| The PWM scoring matrix (Reumann et al., 2012; Skoulding et al., 2015, supplementary information), with a route through the matrix shown for the calculation of the PWM score of the peptide WWRDPYSPMYQSYY. The PWM score of a peptide is calculated by the sum of all numbers assigned to the individual residues occurring at each position in the peptide. In this example, the addition of the nine residues, shown in blue, to YQSYY (shown in green) means that 0.47 is added to the total PWM score. This redefines the sequence from a non-PTS1 sequence to a PTS1 sequence, as determined by PredPlantPTS1 (Lingner et al., 2011; Reumann et al., 2012).

In the example shown in **Figure 66**, YQSYY has a PWM score of 0.07, whereas WWRDPYSPMYQSYY has a PWM score of 0.54. The peroxisomal import threshold, as defined by the Reumann research group and by Skoulding and colleagues, is 0.15–0.412 (PredPlantPTS1) (Reumann et al., 2012; Skoulding et al., 2015). A C-terminal tyrosine is also likely to be underweighted in the PWM scoring matrix (discussed in Chapter 3, section 3.5) so the real PTS1 prediction score of WWRDPYSPMYQSYY could potentially be higher than 0.54. As a result of the addition of particular upstream residues to YQSYY, the total sequence score is pushed above the threshold for peroxisomal protein import. This illustrates the importance of upstream residues for the

peroxisomal targeting ability of a PTS1 sequence. It was reasoned that the YQSYY sequence could be enhanced as a stronger importing sequence if the appropriate upstream residues were added, and that the resulting sequence may be able to out-compete many of the natural PTS1 sequences for binding to *At*PEX5-C*.

The eventual application of the orthogonal PEX5*:PTS1* import system was in *P. patens* (*Physcomitrella patens*; moss) cells, so it was important to know the types of PTS1 sequence commonly seen in this organism. A visiting postdoctoral researcher, Dr Heba Ebeed, conducted a study of *P. patens* homologues to PTS1 proteins from *A. thaliana* (*Arabidopsis thaliana*; thale cress). PTS1 proteins from *A. thaliana* had been studied much more extensively than those from *P. patens*, and they were the basis of the creation of the PredPlantPTS1 prediction software (Lingner et al., 2011; Reumann et al., 2012). H. Ebeed collated all major, minor and rare PTS1 sequences from *A. thaliana* using AraPerox (Reumann et al., 2004) and produced a record of *P. patens* homologues to each of these *A. thaliana* PTS1 proteins (Appendix F). The 14 C-terminal residues from each *P. patens* homologue were also included in the database produced by H. Ebeed. These C-terminal sequences were used as the predicted *P. patens* PTS1 sequences in this study.

H. Ebeed's results, showing *P. patens* homologues for *A. thaliana* PTS1 proteins, were used to investigate what the 'competition' would be for the orthogonal sequence in the *P. patens* cell. The resulting *P. patens* predicted PTS1 sequences were ordered according to their PWM score (PredPlantPTS1) (Lingner et al., 2011; Reumann et al., 2012). The highest scoring sequence was termed 'MossHigh', as this was the predicted strongest competing PTS1 sequence in moss cells. 'MossHigh' comes from the 14 C-terminal residues of an oxidoreductase in moss. The lowest scoring sequence was termed 'MossLow'; however, 'MossLow' is the homologue of a rare PTS1 protein in *A. thaliana* and comes from the 14 C-terminal residues of a phosphatase in moss. Therefore, it was not known whether the MossLow sequence would act as the lower boundary of moss PTS1 sequences, or as a non-PTS1 control *in vivo*. The PWM score of MossHigh was 1.39 and that of MossLow was 0.41 (**Table 13**).

Six candidate PTS1* sequences were designed (PTS1*a-f), all terminating in YQSYY-CO_2H, with differing upstream residues (**Table 13**). This provided a range of PTS1* sequences that could be tested *in vivo* in order to identify the optimal sequence that would function in the alternate import pathway. Upstream residues for these

sequences were designed using Excel and VBA (code written by Dr Stuart Warriner) to create a 'PWM optimiser' worksheet which would generate a range of upstream sequences to fit a target overall PWM score.

Peptide name	***Peptide sequence***	***PWM score***
MossHigh	*SHIQTEAERLYSKL*	*1.39*
MossLow	*IIAAVDASYNSSTL*	*0.41*
*PTS1*a*	*WIAGDNSQHYQSYY*	*-0.2*
*PTS1*b*	*WHGAAESKFYQSYY*	*0*
*PTS1*c*	*WHWQVHSEIYQSYY*	*0.16*
*PTS1*d*	*WWATVHSQRYQSYY*	*0.28*
*PTS1*e*	*WWWDVHQHRYQSYY*	*0.4*
*PTS1*f*	*WWRDPYSPMYQSYY*	*0.54*
YQSKL	*YQSKL*	*1.45*
YQSYY	*YQSYY*	*0.07*

Table 13| Peptides chosen for in vivo testing, along with sequences and positional weight matrix (PWM) scores for each peptide. PWM scores were assigned using the PWM determined for PredPlantPTS1 (Lingner et al., 2011; Reumann et al., 2012).

The MossHigh and MossLow sequences were synthesised and coupled to lissamine in order to test the binding affinity of these sequences with wild-type *At*PEX5-C, and with *At*PEX5-C D505H-D507T-N601A (*At*PEX5-C*). PTS1*a and PTS1*f were also synthesised and coupled to lissamine to test the predicted strongest and weakest PTS1* sequences. Characterisation of these four lissamine-labelled peptides can be seen in Chapter 6 (Experimental) section 6.18.14 (lissamine-MossHigh), 6.18.15 (lissamine-MossLow), 6.18.16 (lissamine-PTS1*a), and 6.18.17 (lissamine-PTS1*f).

It was predicted that the peptide sequences would rank in order of PWM score in their affinities for wild-type *At*PEX5-C; however, it was not known what would happen when these sequences were tested with *At*PEX5-C*. In addition, it was unclear if the very strong upstream residues in PTS1*f would lead to increased binding affinity of YQSYY to wild-type *At*PEX5-C. Fluorescence anisotropy results for wild-type *At*PEX5-C and *At*PEX5-C*, with the lissamine-labelled peptides MossHigh, MossLow, PTS1*a, and PTS1*f are shown in **Figure 67**.

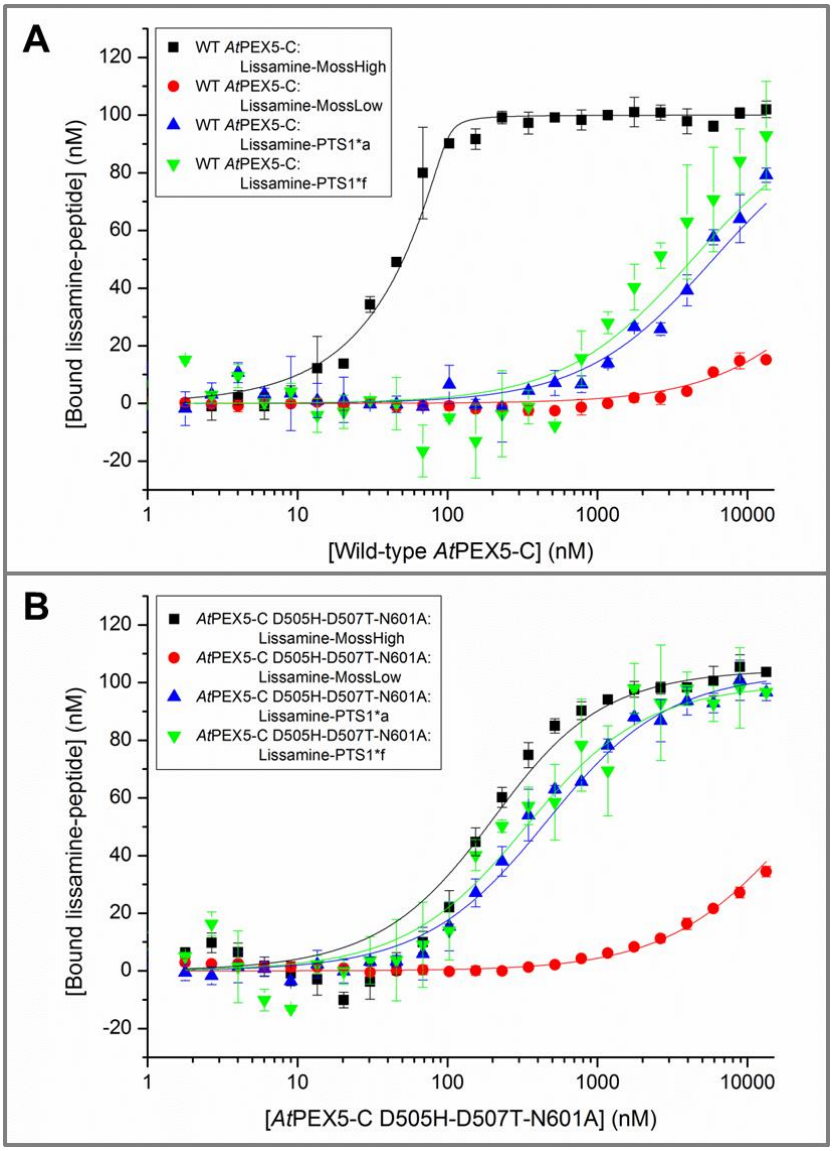


Figure 67| Fluorescence anisotropy testing with 14-aa peptides representing the strongest and weakest PTS1 sequences in moss, and peptide sequence YQSYY with two different upstream sequences. Peptide concentration was fixed at 100 nM final concentration, and proteins were titrated from 0 to 13.3 µM (final concentration). A) Wild-type *At*PEX5-C. B) *At*PEX5-C D505H-D507T-N601A (*At*PEX5-C*). MossHigh, the strongest (predicted) PTS1 sequence in *Physcomitrella patens* (*P. patens*); MossLow, the weakest (predicted) PTS1 sequence in *P. patens*; PTS1*a, peptide YQSYY with a weak upstream sequence; PTS1*f, peptide YQSYY with the strongest possible upstream sequence for peroxisomal targeting. A quadratic equation was used to fit the curves using OriginPro 9.1, assuming a one-to-one model of binding.

It is apparent from **Figure 67** that the fluorescence anisotropy results obtained were not as expected for lissamine-PTS1*a, lissamine-PTS1*f, or lissamine-MossLow, considering the PWM scores of the peptides tested. Lissamine-MossHigh did behave as expected, with a dissociation constant similar to lissamine-YQSKL; however, lissamine-MossLow exhibited an extremely low binding affinity for proteins tested, with a particularly low binding affinity for wild-type *At*PEX5-C. Lissamine-PTS1*f also

exhibited a weaker binding affinity than expected as, with a PWM score of 0.54, it was predicted that the binding affinity of this peptide with wild-type *At*PEX5-C would be stronger than the predicted import threshold (Maynard and Berg, 2007; Reumann et al., 2012; Skoulding et al., 2015): the dissociation constant of wild-type *At*PEX5-C:lissamine-PTS1*f was predicted to be lower than that of wild-type *At*PEX5-C:lissamine-YQSKV (1300 ± 100 nM) (Chapter 2, section 2.4). This could, however, be beneficial as the chances of peroxisomal import of PTS1*f with wild-type PEX5 could be much lower than predicted. As PTS1*f does not have as strong an affinity with wild-type *At*PEX5-C as the PWM score predicts, it may have weak enough targeting efficiency to have no interference with the natural import system *in vivo*. Binding affinities, as determined by fluorescence anisotropy, are summarised in **Table 14**.

***At*PEX5-C variant**	***Fluorescent peptide***	***Dissociation constant (K_d)(nM)***	***Standard error (nM)***	***PWM score***
Wild-type	*Lissamine-MossHigh*	*0.7*	*1.4*	*1.39*
	Lissamine-MossLow	*58,000*	*4000*	*0.41*
	*Lissamine-PTS1*a*	*5900*	*700*	*-0.2*
	*Lissamine-PTS1*f*	*4300*	*1800*	*0.54*
D505H-D507T-N601A	*Lissamine-MossHigh*	*150*	*30*	*N/D*
	Lissamine-MossLow	*22,000*	*1000*	*N/D*
	*Lissamine-PTS1*a*	*400*	*50*	*N/D*
	*Lissamine-PTS1*f*	*300*	*100*	*N/D*

Table 14| Binding affinities of the 14-aa peptides MossHigh, MossLow, PTS1*a, and PTS1*f, as determined by fluorescence anisotropy. PWM scores, showing the predicted likelihood of the peptide sequence directing the peroxisomal import of a protein *via* the PTS1-mediated import pathway, are also shown for comparison.

Following the testing of 14-aa peptides MossHigh, MossLow, PTS1*a, and PTS1*f by fluorescence anisotropy, it was decided that all peptide sequences (MossHigh, MossLow, and PTS1*a-f) would be tested *in vivo* by R. Paudyal (University of Leeds).

4.6 *In vivo* testing of *Phypa*PEX5-N–*At*PEX5-C and *Phypa*PEX5-N–*At*PEX5-C* with PTS1 and PTS1*

Testing of the AtPEX5-C*:PTS1* pair *in vivo* was carried out by R. Paudyal (postdoctoral researcher in the A. Cuming laboratory, University of Leeds) and characterisation of results was enabled by S. Warriner. DNA of wild-type *At*PEX5-C and *At*PEX5-C* in separate pET-28b vectors was given to R. Paudyal for insertion of this (the C-terminus of the *A. thaliana* PEX5 receptor) downstream of the N-terminus of the *P. patens* PEX5 receptor. The resulting vector contained a hybrid receptor (*Phypa*PEX5-N–*At*PEX5-C or *Phypa*PEX5-N–*At*PEX5-C*). PTS1* peptide sequences were also sent to R. Paudyal so that primers could be designed for the insertion of each of these sequences just downstream of green fluorescent protein (GFP) (at the C-terminus) (**Figure 68**). R. Paudyal also inserted the MossHigh or MossLow sequences in the same way, just downstream of red fluorescent protein (RFP) (at the C-terminus) (**Figure 68**). Results in this section (4.6) are reported for completeness of the study.

It was decided that the peroxisomal import of the PTS1* sequences would be tested alongside the MossHigh or MossLow sequences, to simulate the real competition that the orthogonal system would be subject to *in vivo*. In order to achieve this, the gene coding for either the *Phypa*PEX5-N–*At*PEX5-C or the *Phypa*PEX5-N–*At*PEX5-C* receptor was inserted in a vector which also carried the gene coding for RFP-MossHigh or RFP-MossLow, under the control of a separate promoter (the receptor was under the control of an actin promoter, and RFP-MossHigh/MossLow was under the control of a double 35S promoter). The gene coding for either GFP-PTS1*a, GFP-PTS1*b, GFP-PTS1*c, GFP-PTS1*d, GFP-PTS1*e, or GFP-PTS1*f was inserted into a separate vector (under the control of a double 35S promoter). Particle bombardment was then used to introduce the DNA into cells, 38–48 h prior to imaging.

The set-up of the experiment, along with a cartoon of idealised results, is illustrated in **Figure 68**. The expectation, if the interaction had been designed correctly, was that peroxisomes would fluoresce red and the cytosol would fluoresce green in the wild-type import system. In the mutated import system, the expectation was that peroxisomes would fluoresce green and the cytosol would fluoresce red, although some level of red peroxisomal fluorescence was expected due to the presence of native levels of the wild-type PEX5 receptor in moss cells. If both sequences were

targeted to the peroxisome fully, peroxisomes should appear to fluoresce yellow after merging the red and green channels after imaging.

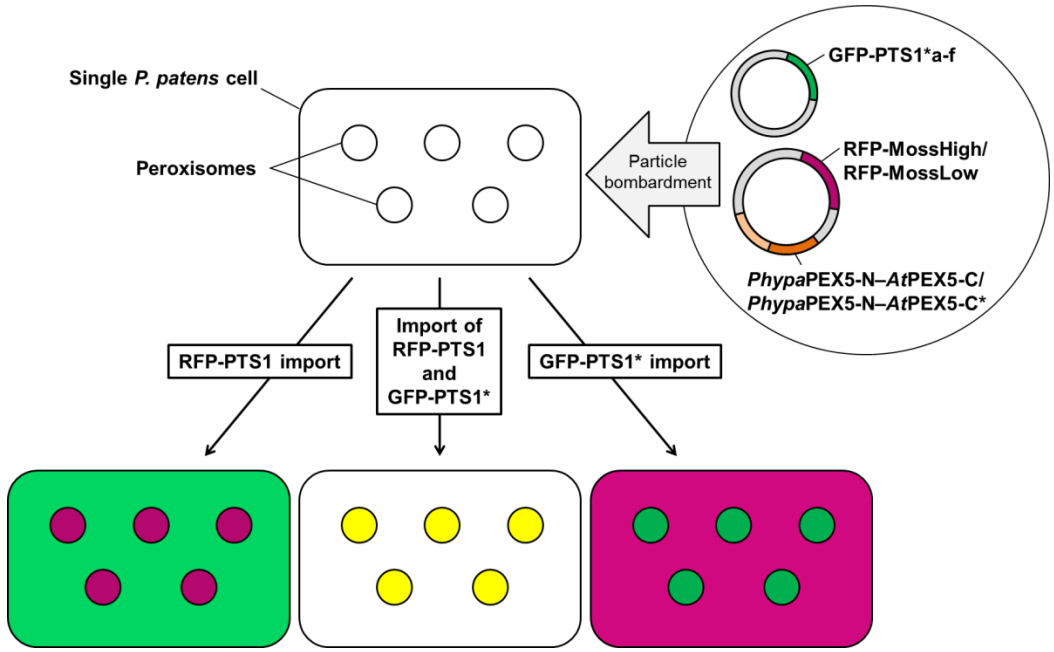


Figure 68| The *in vivo* targeting concept. Particle bombardment can be used to introduce DNA into cells. Imaging is than carried out 38–48 h after particle bombardment. In the wild-type import system, we expect peroxisomes to fluoresce red and the cytosol to fluoresce green. In the mutated import system we expect peroxisomes to fluoresce green and the cytosol to fluoresce red. If both RFP-PTS1 (MossHigh/MossLow) and GFP-PTS1* (PTS1*a-f) are imported, we expect to observe yellow-fluorescing peroxisomes as this shows co-localisation of the two fluorescent proteins. MossHigh and MossLow represent the potential strongest and weakest predicted PTS1 sequences in *Physcomitrella patens* (*P. patens*). PTS1, peroxisomal targeting signal 1; PTS1*, mutated peroxisomal targeting signal 1; RFP, red fluorescent protein; GFP, green fluorescent protein.

Particle bombardment of the vectors, to induce transient expression within the moss cells, was carried out and images captured after 38–48 h. As an initial control experiment, the vector containing the hybrid *Phypa*PEX5-N–*At*PEX5-C receptor and RFP-MossHigh was introduced into moss cells by bombardment, without the introduction of the GFP-PTS1*a-f vector. This was to test whether the hybrid receptor would import RFP, albeit against a background of endogenous levels of *Phypa*PEX5 expression. In this experiment, we expected to see red fluorescing peroxisomes against a background of no other fluorescence. Results from this initial test are shown in **Figure 69**.

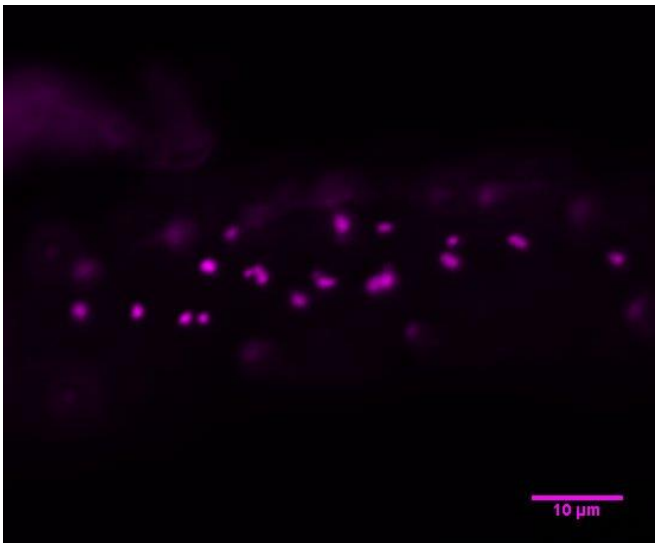


Figure 69| Imaging of a *P. patens* cell, 38–48 h after particle bombardment with a vector containing the *Phypa*PEX5-N–*At*PEX5-C receptor and RFP-MossHigh. Image courtesy of Rupesh Paudyal.

Peroxisomes are apparent in **Figure 69** by the presence of small bright 'dots' in the *P. patens* cell when imaging was carried out by R. Paudyal in the RFP channel. It is clear from **Figure 69** that import of RFP-MossHigh into peroxisomes is functional under the experimental conditions used, so the study could continue. PTS1*c was not successfully inserted into the GFP vector by R. Paudyal, at the C-terminus of GFP (so this is not present in the following results).

To test the import system with a more representative PTS1 sequence, a *P. patens* homologue of another example PTS1 sequence with a lower PWM score (1.01) than MossHigh (1.39) was determined from H. Ebeed's work (GETIVVAGGMKSRL). This sequence was termed 'MossMid' and was sent to R. Paudyal for cloning into the RFP/receptor plasmid, at the C-terminus of RFP. MossMid was still a strong representative PTS1 sequence, as the average PWM score for a *Physcomitrella patens* PTS1 corresponding to a 'major' *Arabidopsis thaliana* PTS1 is 0.880 (median = 0.890). (from H. Ebeed's work).

Following particle bombardment for the full experiment with a range of PTS1* sequences, 7,997 images were captured by R. Paudyal 38–48 h after bombardment and these were classified into one of 7 categories (-1 = unsure which cell to classify; 0 = no fluorescence; 1 = cytosolic; 2 = >75% cytosolic; 3 = mixed; 4 = >75% peroxisomal; 5 = peroxisomal) independently by 7 people. In order to carry out a non-biased classification, all images captured by R. Paudyal were saved in grayscale and

were uploaded onto a server created by S. Warriner, to be presented in a random order (so that information about which receptor was being overexpressed in each cell was not known). When classification began, a random image from the total image set was presented for classification. Upon completing classification of the first image, the next image would be randomly selected from the image set and presented for classification. This continued until all images were classified. Images classified as either -1 or 0, in either the RFP or GFP channel, by more than four people were excluded from analysis, then the average score and standard error for each experiment was taken. Example images from the classification set are shown for *Phypa*PEX5-N–*At*PEX5-C.RFP-MossHigh/MossLow + GFP-PTS1*a/f (**Figure 70**), along with the average classification score assigned by the 7 people.

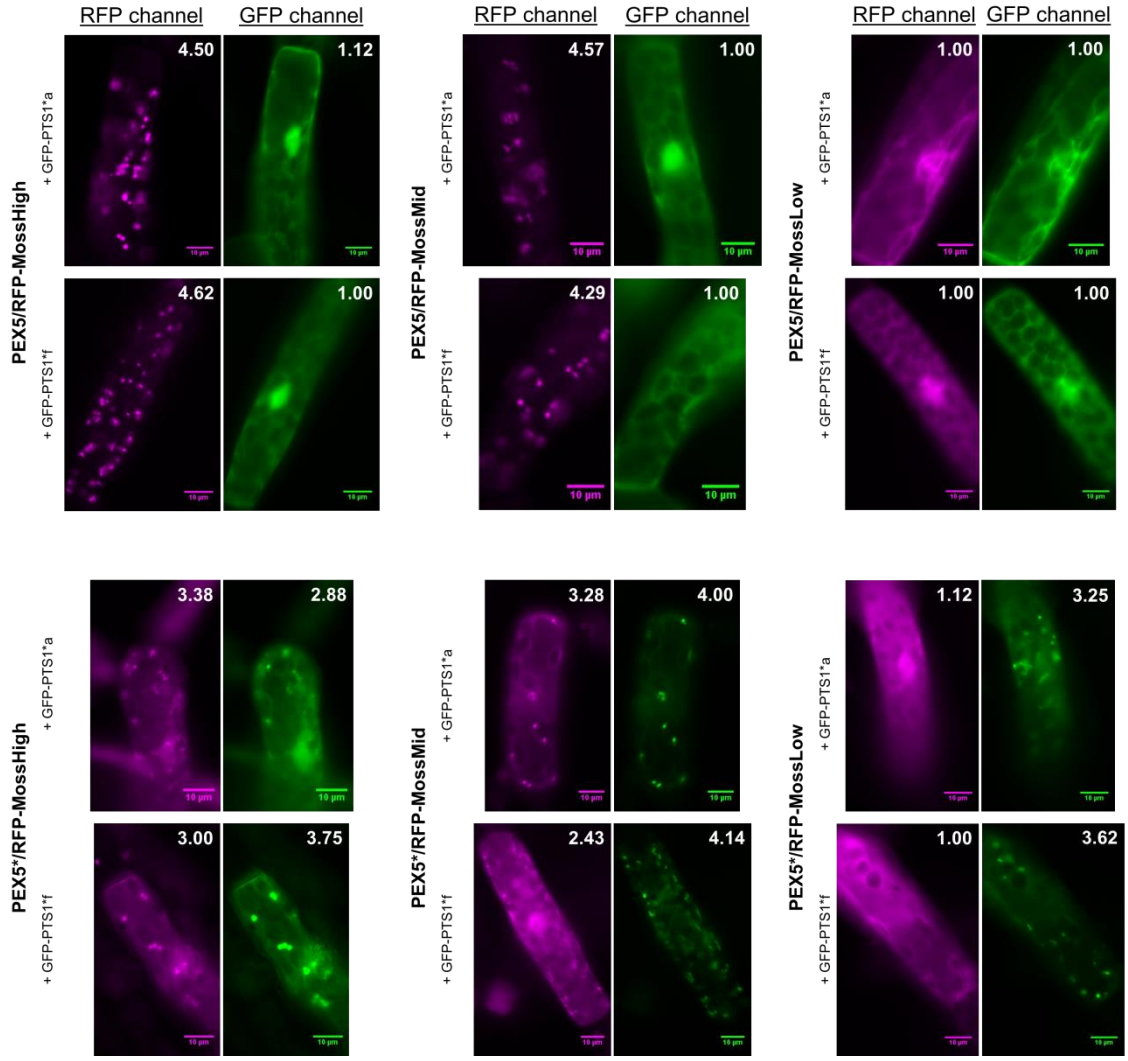


Figure 70| Cell (*Physcomitrella patens*) images captured by R. Paudyal, following particle bombardment with a vector containing either the *Phypa*PEX5-N–*At*PEX5-C or *Phypa*PEX5-N–*At*PEX5-C* receptor and RFP-MossHigh/MossLow, and a vector containing GFP-PTS1*a/f. With *Phypa*PEX5-N–*At*PEX5-C, RFP-MossHigh is imported preferentially over even the strongest PTS1* sequence (GFP-PTS1*f). No import of either RFP-MossLow or GFP-PTS1*a/f is seen with *Phypa*PEX5-N–*At*PEX5-C. *Phypa*PEX5-N–*At*PEX5-C* newly facilitates the import of GFP-PTS1*a/f; however, there is definite co-localisation of RFP-MossHigh when PTS1*a is used as the PTS1* sequence. GFP-PTS1*f appears to out-compete much of the import of RFP-MossHigh, and so *At*PEX5-C* seems to have a preference for importing GFP-PTS1*f over importing RFP-MossHigh. The RFP-MossLow negative control shows that PTS1* import can be fully achieved by *Phypa*PEX5-N–*At*PEX5-C* when there is no competition from over-expressed PTS1 sequences. Images courtesy of Rupesh Paudyal. Average classification scores, for each image, are shown in the top right-hand corner of each image. All scale bars shown represent 10 µm.

A definite change in import behaviour occurred as a result of three mutations in the *At*PEX5-C portion of the receptor. Interestingly, RFP-MossLow was not imported into peroxisomes by either *Phypa*PEX5-N–*At*PEX5-C or *Phypa*PEX5-N–*At*PEX5-C*; however, RFP-MossLow acted as an efficient negative control for peroxisomal protein

import with either of the receptor proteins. The results of the analysis after classification of all images by all 7 people are shown in **Figure 71**.

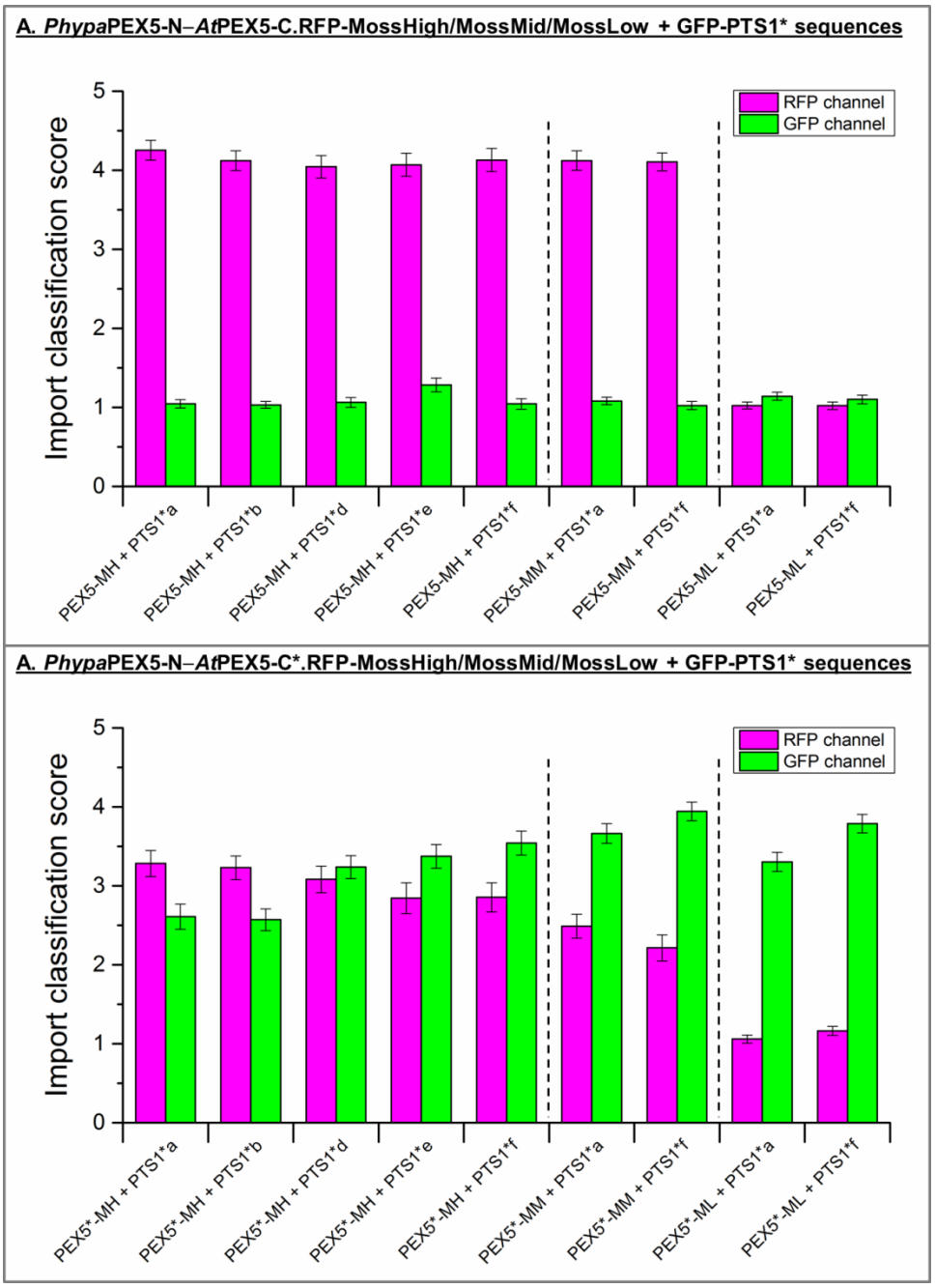


Figure 71| Analysis following the classification of all cell (*Physcomitrella patens*) images captured by R. Paudyal, following particle bombardment with a vector containing either the *Hypa*PEX5-N–*At*PEX5-C or *Hypa*PEX5-N–*At*PEX5-C* receptor and RFP-MossHigh/RFP-MossMid/RFP-MossLow, and a vector containing GFP-PTS1*a/b/d/e/f. A higher import classification score means increased peroxisomal localisation of the fluorescent protein. 1 = fully cytosolic, 5 = fully peroxisomal. The switching of *At*PEX5-C for *At*PEX5-C* downstream of *Hypa*PEX5-N drastically increased the peroxisomal localisation of GFP. MH, MossHigh; MM, MossMid; ML, MossLow.

The analysis of image classification shows that a change has occurred in the nature of the import of both RFP-MossHigh/RFP-MossMid and all GFP-PTS1* sequences tested, following mutation of just three residues in *At*PEX5-C (two in TPR2 and one in TPR5). The import of GFP-PTS1*a/f was selective, as RFP-MossLow was not imported into the peroxisome upon overexpression of *Phypa*PEX5-N–*At*PEX5-C*. The results of the *in vivo* testing have confirmed that it was possible to evolve the import of a protein containing a non-PTS1 sequence into the peroxisome, through mutation of three residues in the PTS1-binding domain of PEX5. When the hybrid receptor *Phypa*PEX5-N–*At*PEX5-C* was overexpressed in the *P. patens* cell (as a result of particle bombardment), non-native PTS1 sequences at the C-terminus of GFP (GFP-PTS1*a/b/d/e/f) mediated the import of GFP into the peroxisome.

4.7 Summary

The pull-down-LC-MS screen developed in this study was used to screen a small library of expressed and purified *At*PEX5-C variants against a larger library of synthesised peptides. Some apparent binding of non-PTS1 sequences was occurring with three of the *At*PEX5-C variants tested, and two mutations in the protein were combined in an attempt to enhance binding to specific non-PTS1 sequences. Binding was enhanced towards aromatic peptides for one of the *At*PEX5-C variants with combined mutations. This appeared to be enhanced as a result of the combination of mutations, and this *At*PEX5-C variant was taken forward for further analysis and optimisation. The final *At*PEX5-C* protein was decided upon through a semi-rational approach, in which a further mutation was introduced into the previous best protein variant for binding to aromatic peptides. This additional mutation was applied at a residue in *At*PEX5-C which had been found, through the pull-down-LC-MS screening in this study, not to be essential for peptide binding within the PTS1-binding site of *At*PEX5-C.

The resulting triple variant of *At*PEX5-C (D505H-D507T-N601A) was termed *At*PEX5-C* and this variant had a very similar affinity for lissamine-YQSKL (the representative PTS1 sequence in this study) and lissamine-YQSYY (the preliminary PTS1* sequence identified in the pull-down-LC-MS screen), by fluorescence anisotropy. This represented a change in selectivity from ~2454-fold in favour of lissamine-YQSKL over lissamine-YQSYY, in the case of wild-type *At*PEX5-C, to ~1.1-fold, in the case of *At*PEX5-C*. The *At*PEX5-C* protein was then tested with longer peptides, which had varying upstream sequences, in an attempt to enhance the

interaction with the preliminary PTS1* sequence YQSYY. These six sequences with additional upstream residues were termed PTS1*a-f and, when compared to wild-type *At*PEX5-C, *At*PEX5-C* was found *in vitro* to bind to lissamine-PTS1*a and lissamine-PTS1*f with significantly higher affinities, and to lissamine-MossHigh, the predicted PTS1 sequence in *P. patens* with the highest PTS1 prediction score (PredPlantPTS1 (Lingner et al., 2011; Reumann et al., 2012)), with a significantly lower binding affinity. Therefore, this protein was taken forward for *in vivo* testing by R. Paudyal (University of Leeds).

R. Paudyal's work included the construction of hybrid receptors (*Phypa*PEX5-N–*At*PEX5-C and *Phypa*PEX5-N–*At*PEX5-C*) and the testing of import of RFP-MossHigh, RFP-MossMid, RFP-MossLow and GFP-PTS1*a/b/d/e/f sequences when genes for the receptors and for the proteins for import were over-expressed following particle bombardment of their DNA at *P. patens* cells (**Figure 68**). The results from R. Paudyal's *in vivo* experiments were very promising, as they showed that the *At*PEX5-C* protein was able to import all PTS1* sequences tested, with decreasing levels of competition from RFP-MossHigh and RFP-MossMid when the PTS1* sequence had a stronger upstream sequence (a higher overall PWM score). The most optimal PTS1* sequence (PTS1*f) resulted in import of GFP while more weakly importing the strongest predicted PTS1 sequence from *P. patens* (RFP-MossHigh). From this point onwards, PTS1*f will be referred to as PTS1*.

The sequence MossMid still directed the peroxisomal import of RFP by wild-type PEX5, yet was out-competed more readily by GFP-PTS1* when the triple variant of the receptor (PEX5*) was expressed in *Physcomitrella patens*. The sequence MossLow served as a specificity control for the receptors, and showed that PEX5* does not import sequences non-specifically. This work has demonstrated that it has been possible to swap the selectivity of the PEX5 receptor from preferentially importing RFP-MossHigh and RFP-MossMid to preferentially importing GFP-PTS1*, by mutation of three residues in the PTS1-binding site.

Chapter 5

General Discussion and Future Perspectives

The aim of this study was to evolve an interaction, involving a PEX5 variant, that would direct non-PTS1-containing proteins to the peroxisome whilst reducing or eliminating natural PTS1-cargo import.

5.1 General discussion

Through a semi-rational approach to the mutation of *At*PEX5-C, alongside screening of protein variants against a library of chemically synthesised peptides, it was possible to isolate the variants D505H and N601A as binding to the peptide YQSYY, where wild-type *At*PEX5-C did not. These two mutations were combined, resulting in a stronger interaction between YQSYY and the double variant D505H-N601A. A further mutation, D507T, was applied to the protein and this resulted in the triple variant D505H-D507T-N601A, which had a binding affinity for YQSYY ~25-fold stronger than wild-type *At*PEX5-C did, and a binding affinity for YQSKL ~88-fold weaker than wild-type *At*PEX5-C did. The addition of enhancing residues upstream of YQSYY resulted in the PTS1* sequence create a new binding interaction between the mutated PEX5 (*At*PEX5-C*) and a novel targeting signal (PTS1*) (**Figure 72**). Pleasingly, *in vivo* the mutated PEX5 receptor (termed PEX5*) was able to direct proteins containing this targeting signal from the cytosol to the peroxisome where native PEX5 could not.

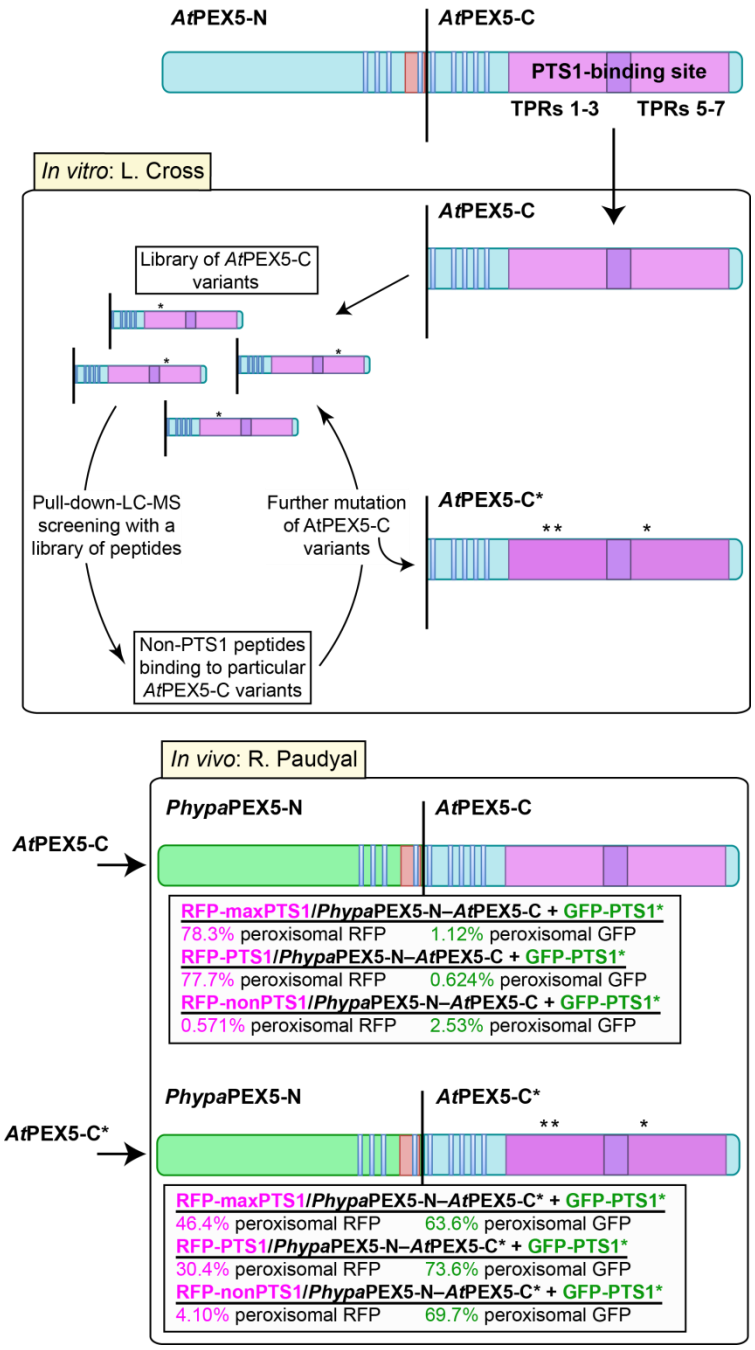


Figure 72| Summary of the research study. The C-terminal domain of *Arabidopsis thaliana* PEX5 was mutated and, following screening for binding to short peptide sequences, manipulated to bind to a non-PTS1 sequence as determined *in vivo* (*in vivo* work carried out by Dr Rupesh Paudyal). The thin blue bars on *At*PEX5 = WxxxF/Y repeats; the pale red box on *At*PEX5 = PEX7 binding site.

A novel screen was developed, in which over 300 peptide sequences (per assay) could be tested for their binding to a purified protein. There was no need for a genotype-phenotype link in this assay, as binding peptides could be identified after pull down using high-resolution mass spectrometry. This method had benefits over colony blotting (the initial screen developed in this work) as it was much faster to perform the

assay, processing of screen results could be almost fully automated so hits were identified quickly, and a full peptide-binding profile of each protein variant was obtained.

Despite the advantages of the screen developed, there were several factors that needed to be considered. As there were many peptide sequences being purified and tested in one mix in this assay, it was not possible to determine the exact concentration of each peptide in the library (as peptides were resuspended based on the average mass of the peptides in the library). The split-and-mix synthesis strategy eliminated bias at the initial synthesis stage but some peptides ionise more effectively, so could have been more abundant in the mass spectrum. This technique, therefore, presents a way to quickly assess protein binding of a range of peptides before performing more detailed binding assays.

Using the pull-down-LC-MS screen, it was possible to confirm which residues in *At*PEX5-C are necessary for PTS1 binding (**Figure 73**). When mutated to alanine, some residues knocked out binding with any peptide, indicating that these residues are crucial for the PTS1-binding site of PEX5 (**Figure 73**, coloured magenta).

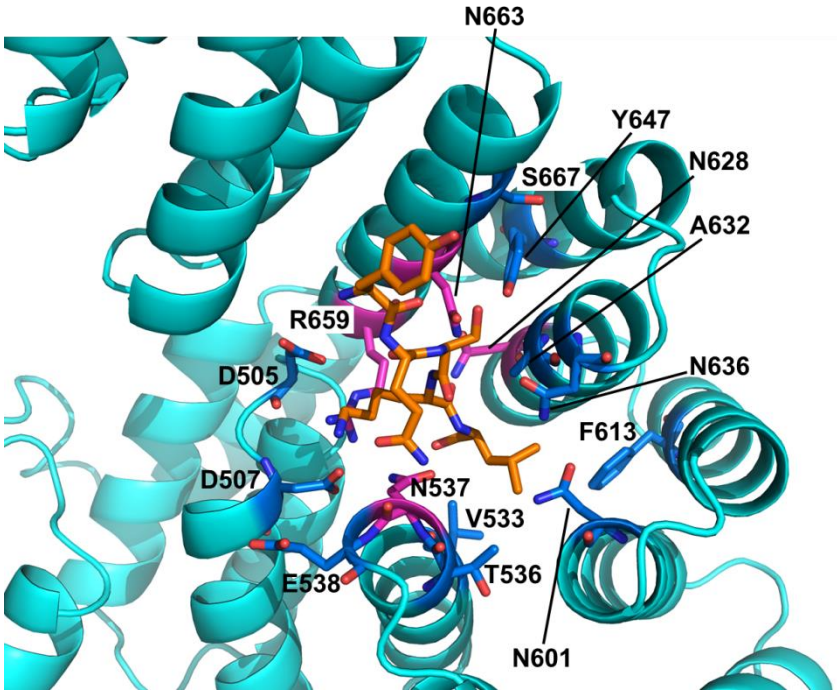


Figure 73| Residues studied in this research. Magenta = crucial for peptide binding, as determined using the pull-down-LC-MS screen. AtPEX5-C model produced using I-TASSER (Zhang, 2008; Roy et al., 2010; Yang et al., 2015; Yang and Zhang, 2015), and the YQSKL peptide is from a HsPEX5-C:PTS1 crystal structure (PDB file 1FCH, Gatto et al., 2000).

Most of the residues crucial for peptide binding, determined in this study through the use of the pull-down-LC-MS screen, were consistent with mutations in *Saccharomyces cerevisiae* PEX5 or *Homo sapiens* PEX5 which knocked out PTS1 binding or import (Klein et al., 2001; Dodt et al., 1995): *At*PEX5 N537 is equivalent to *Sc*PEX5 N393, *At*PEX5 N628 is equivalent to *Hs*PEX5S N489, and *At*PEX5 R659 is equivalent to *Sc*PEX5 R526 (section 1.4.2). The pull-down-LC-MS screen was therefore a fast and sensitive way to obtain a great deal of information on peptide sequences that bind to *At*PEX5-C, and known binding sequences of wild-type *At*PEX5-C (native PTS1 sequences) were successfully pulled down using this method.

A major deliverable of this research was a new targeting signal (PTS1*) unable to direct proteins to the peroxisome in the presence of wild-type PEX5 but, upon expression of PEX5*, enabled the direction of PTS1*-tagged proteins to the peroxisome. The overall aims of this study bear similarities to recently published research in *Saccharomyces cerevisiae* performed by DeLoache and colleagues (DeLoache et al., 2016). However, in the study performed by DeLoache and colleagues, their deliverable was an enhanced PTS1 (‘ePTS1’) that was imported preferentially into the peroxisome over an example PTS1 sequence used. This allowed the direction of a non-peroxisomal two-enzyme metabolic pathway into peroxisomes, where it was able to function (DeLoache et al., 2016); however, natural PTS1 sequences were still able to import into peroxisomes normally, as the PEX5 receptor remained unchanged. In this study, the import of PTS1 sequences was greatly reduced upon expression of the receptor variant PEX5*.

This work suggests that the PWM score (PredPlantPTS1 (Lingner et al., 2011; Reumann et al., 2012) could over-predict when there is a non-PTS1 at the C-terminus. The PWM score threshold for peroxisomal import has been defined as 0.15–0.412 (Lingner et al., 2011; Reumann et al., 2012; Skoulding et al., 2015); however, the sequence PTS1* has a PWM score of 0.54 and shows no import in P. patens. This could be due to the length of the import experiments, as Reumann and colleagues carry out imaging up to 72 h after particle bombardment and, in this study, the maximum time after particle bombardment and before imaging was 48 h.

5.2 Future perspectives

It would be interesting to determine the association and dissociation rates for the newly evolved *At*PEX5-C*:PTS1* interaction, in comparison to that of the

*At*PEX5-C:PTS1 interaction, using surface plasmon resonance (SPR). Work towards this goal had started during this study by coupling biotin to the peptide YQSKL with different linkers, to determine which biotinylated peptide exhibited a binding affinity most similar to unlabelled YQSKL (Appendix F). The most appropriate biotin label was identified as biotin-$(PEG)_2$, so biotin-$(PEG)_2$-YQSYY was also produced (Chapter 6 (Experimental) section 6.18.21); therefore, biotin-$(PEG)_2$-YQSKL or biotin-$(PEG)_2$-YQSYY could be attached to a streptavidin-coated gold chip, to measure the SPR with wild-type *At*PEX5-C or *At*PEX5-C*. It would also be of interest to crystallise wild-type *At*PEX5-C:PTS1 and *At*PEX5-C*:PTS1*, in order to compare shapes of the binding sites. A further N-terminally truncated construct of *At*PEX5-C was created for this purpose, as crystallisation attempts with *At*PEX5-C (*At*PEX5(340-728)) had been unsuccessful. In *At*PEX5(444-728), all of the predicted flexible portion of the protein had been removed, which should have resulted in a more compact protein. Initial crystallisation screens (Chapter 6 (Experimental) section 6.17) using this protein construct were unsuccessful; however, crystallisation of *At*PEX5(444-728) could be possible by performing an increased number of crystallisation screens with more variation.

The pull-down-LC-MS screen developed in this study could also be used to test PEX5 proteins from different species and determine whether there are key differences in the PTS1 sequences that they bind to. For instance in parasitic trypanosomes, peroxisomes are crucial for cell survival and, therefore, so is PEX5. By identifying a PTS1 sequence that binds strongly to trypanosome PEX5 and not to human PEX5, it could be possible to selectively target trypanosome PEX5.

One process in peroxisomal matrix protein import that is yet to be unravelled is cargo unloading into the matrix of the peroxisome. A recent study showed that the addition of a bulky C-terminal tag to mammalian PEX5 (designed to mimic a cargo protein unable to unload from PEX5) resulted in a monoubiquitinated, membrane-bound form of PEX5. It could be suggested that without a free C-terminus PEX5 is unable to participate in cargo unloading and, if export-driven import is the correct model (Chapter 1, section 1.3.5), it is therefore unable to be exported to the cytosol from the peroxisome membrane. There appears to be a conserved sequence, LxKEF, at the C-terminus of PEX5 across plant species (**Figure 74**, insert), similar to the sequence identified at the N-terminus of mammalian PEX5, LVxEF, as being important for the initial interaction of PEX5 with PEX14. It is, therefore, tempting to propose that this C-

terminal sequence could allow the cargo unloading process to occur, and could be a final point-of-contact for PEX5 to PEX14, before PEX5 begins the recycling process.

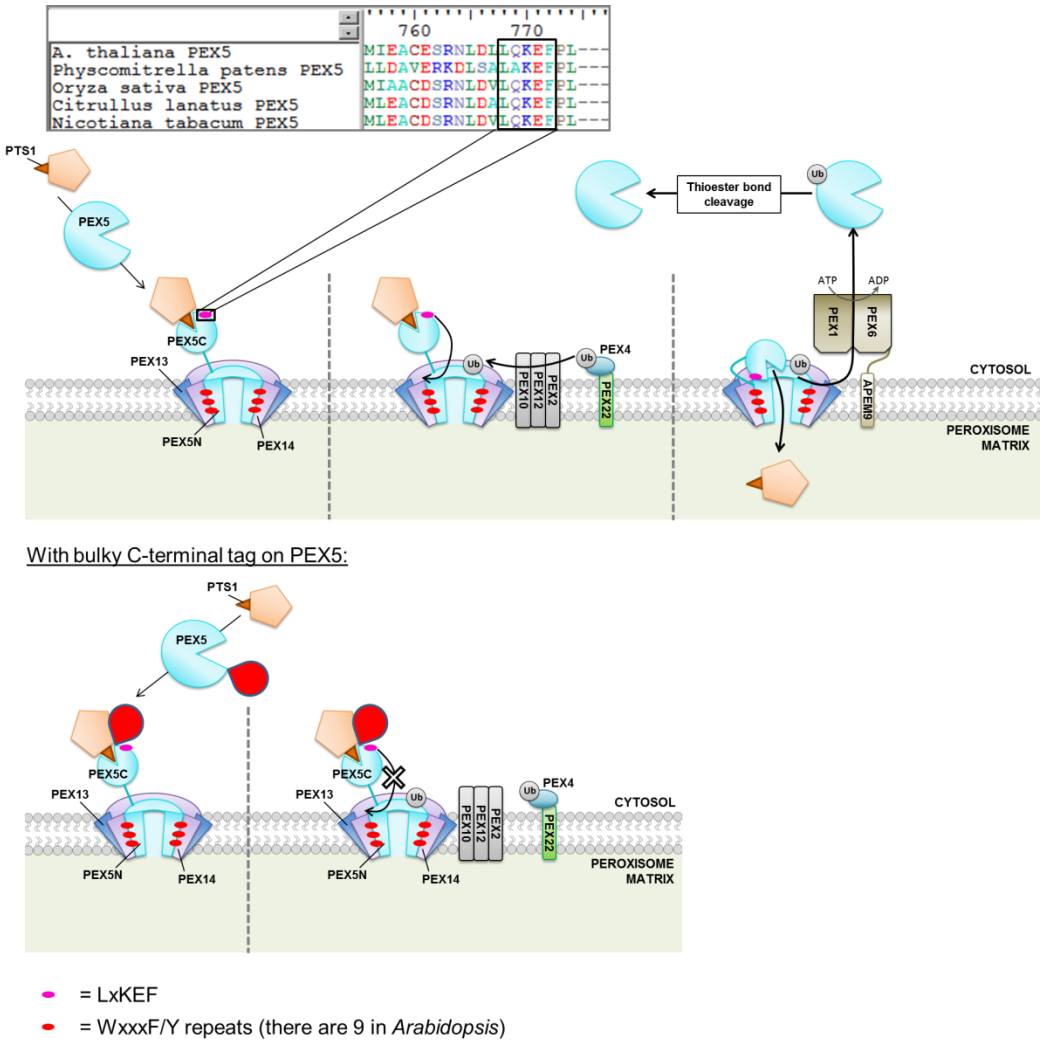


Figure 74| Proposed mechanism for cargo unloading. The conserved C-terminal sequence LxKEF in plant PEX5 could act as a final interacting motif for PEX5 with PEX14, after the WxxxF/Y repeats in PEX5 are interacting with PEX14. This final interacting step could allow for the release of PTS1-cargo into the peroxisomal matrix, as PEX5 is recycled to the cytosol in parallel. With a bulky C-terminal tag on PEX5, this could prevent the interaction of LxKEF with PEX14, preventing cargo release and therefore preventing PEX5 recycling after monoubiquitination (assuming export-driven import). Alignments were performed using BioEdit (sequence alignment editor).

The orthogonal import system developed in this study could be used to test whether the C-terminus of PEX5 is involved in cargo unloading, as we now have a signal sequence (PTS1*) that shows no import with the native PEX5 receptor. Transient expression, therefore, could be used to test this theory (as background native PEX5 levels would not have an effect). If a stop codon was introduced immediately upstream

of the LxKEF sequence, the cellular localisation of GFP-PTS1* would reveal the subsequent effect.

The orthogonal system produced in this study could be fully realised by the creation of an inducible system, in which native PEX5 expression is knocked out alongside upregulation of PEX5* expression. A potential way to do this would be using Cre-Lox recombination, as illustrated in **Figure 75**. This would mean, when PEX5* is being produced, that no background of native PEX5 levels would interfere.

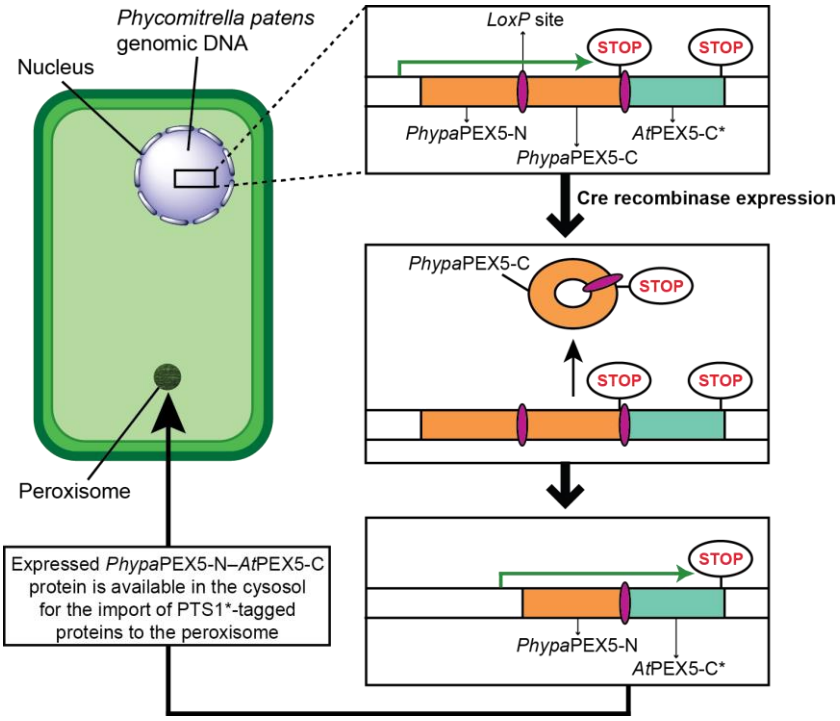


Figure 75| Cre-Lox recombination, for an inducible switch in expression between *Phypa*PEX5 and *Phypa*PEX5-N–*At*PEX5-C*. One *loxP* site could be placed in an exon between the N- and C-terminal domain of *Phypa*PEX5, and the other could be placed in an exon just after the stop codon of *Phypa*PEX5. After the stop codon of *Phypa*PEX5, *At*PEX5-C* could be inserted into the genome. Upon induction of Cre recombinase expression, the *Phypa*PEX5 C-terminal domain would be excised from the genome, leaving the *Phypa*PEX5-N–*At*PEX5-C* protein to be expressed.

A major concern of manipulating PTS1-protein import to by-pass the native pathway, upon switching PEX5 expression for PEX5*, is that cells depend on peroxisomal protein import, and therefore native PEX5, for their survival. Tanner and colleagues produced an artificial peroxisome by encapsulating two antioxidant enzymes within a polymer vesicle containing channel proteins for substrate-product exchange (Tanner et al., 2013). This artificial peroxisome successfully carried out reactive oxygen species

(ROS) detoxification, so it could be possible to develop artificial peroxisomes for use in plant cell vitality while customising the role of pre-existing peroxisomes in the cell.

As this proof-of-principle work was successful, it may be possible to work towards the development of peroxisomes as storage organelles. This would require a stable line of moss cells with inducible PEX5*, so that levels of wild-type will decline after a user-specified switch of PEX5 expression for PEX5* expression. There also needs to be a way to keep the cell alive, which could possibly be achieved through the use of artificial peroxisomes (Tanner et al., 2013). To allow for the storage of as much protein as possible, a PEX11 knockout cell line could be used, which has been shown to contain 'giant peroxisomes' (Kamisugi et al., 2016). This could allow for the storage of therapeutic proteins, for example recombinant vaccines or monoclonal antibodies. Such a storage organelle could allow for longevity of unstable proteins within living cells until desired, at which time we envisage that stored proteins could be isolated from cell extract using the PEX5* receptor (**Figure 76**).

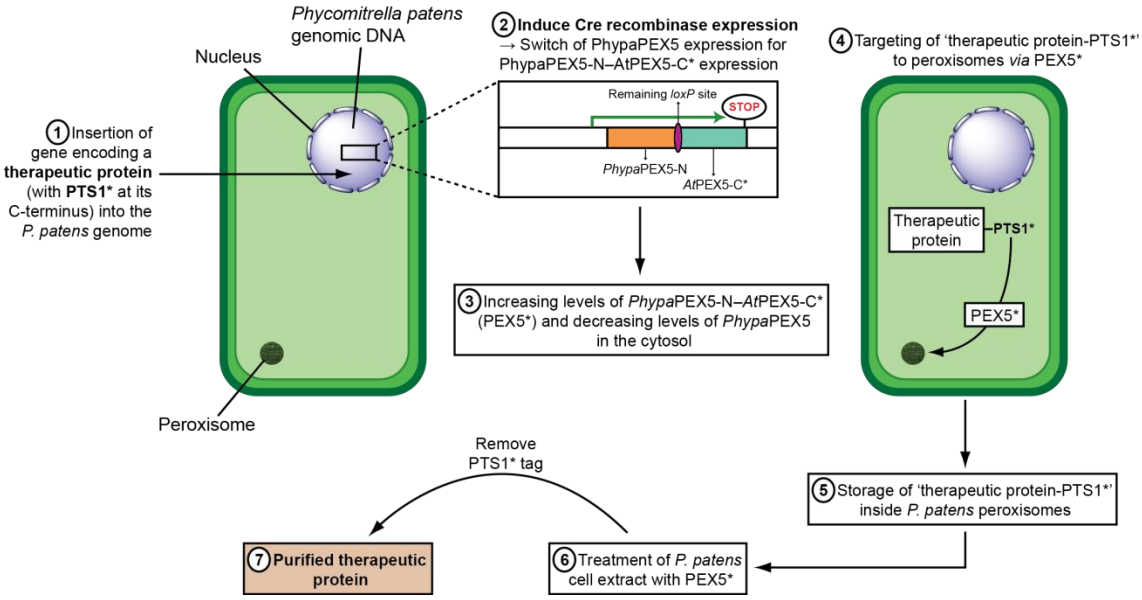


Figure 76| A use for a synthetic peroxisome. The gene of interest is inserted into the *P. patens* genome, and expression of natural *Phypa*PEX5 is switched for *Phypa*PEX5-N–*At*PEX5-C*. Therapeutic protein is then stored inside peroxisomes until needed.

Moss (*Physcomitrella patens*) was used as the host organism in this study as it is commonly used as a bioreactor in industrial applications (Decker and Reski, 2004) and the genome is fully sequenced (Rensing et al., 2008) so genes can be stably transformed into the genome for expression. It will be interesting to determine how versatile peroxisomes can become when the protein content of the matrix is drastically

altered, and if it will be possible to create a system where re-purposed peroxisomes carry out specified import and storage alongside maintenance of cell vitality by artificial peroxisomes.

Chapter 6

Experimental

6.1 Bacterial strains and plasmids

XL10-Gold (Stratagene)

Genotype: Tet^{r}Δ(*mcrA*)*183* Δ(*mcrCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* Hte [F´ *proAB lacI*^{q}*Z*Δ*M15* Tn*10* (Tet^r) Amy Cam^r]

BL21-Gold (DE3) (Agilent Technologies)

Genotype: *E. coli* B F$^-$ *ompT hsdS*(rB$^-$ mB$^-$) *dcm*$^+$ Tet^r *gal* λ(DE3) *endA* Hte

The pET-28b (Kan^r) plasmid, containing the His$_6$-*At*PEX5(340-728) gene, was a gift from T. Lanyon-Hogg (former student of University of Leeds). The His$_6$-*At*PEX5(340-728) gene was originally cloned into pET-28b by S. Gunn (former student of University of Leeds).

6.2 Kanamycin stock solution

Kanamycin monophosphate (Formedium^{TM}), made up in distilled water (50 mg/mL stock solution) and filter-sterilised.

6.3 Bacterial media

All bacterial media was made up in distilled water and autoclaved at 121°C for 20 minutes. For ZY-AI, all components other than ZY were sterile-filtered and added after autoclaving of ZY.

LB (Luria Bertani) media

Tryptone (1% w/v), yeast extract (0.5% w/v), NaCl (1% w/v)

LB-agar plates

Tryptone (1% w/v), yeast extract (0.5% w/v), NaCl (1% w/v), agar (1.5% w/v)

ZY media

Tryptone (1% w/v), yeast extract (0.5% w/v), NaOH (1 mM)

ZY-AI (autoinduction media)

ZY media (928 mL/L), $MgSO_4$ (1 mM), $(NH_4)_2SO_4$ (25 mM), KH_2PO_4 (50 mM), Na_2HPO_4 (50 mM), glycerol (0.5% w/v), glucose (0.05% w/v), α-lactose (0.2% w/v)

'Selection media'

LB, LB-agar, or ZY with kanamycin stock solution added in a 1:1000 dilution; or ZY-AI with kanamycin stock solution added in a 1:500 dilution

6.4 Restriction enzymes

Restriction enzymes were obtained from New England Biolabs (NEB) and conditions for restriction digests were determined using the NEB online 'Double Digest Finder'.

6.5 Site-directed mutagenesis

A QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies) was used according to manufacturer's instructions. Transformation of each DpnI-treated mutagenesis reaction into XL10-Gold cells was performed, followed by DNA purification and confirmation of mutagenesis by sequencing (Beckman Genomics).

6.6 Ligation-independent cloning (for the production of *At*PEX5(444-728))

Ligation-independent cloning (LIC) was used for the deletion of a large portion of the N-terminus of *At*PEX5(340-728) in order to generate the truncated variant *At*PEX5(444-728). Primers were designed with a 16-bp overlap according to a methodology article about LIC, or 'FastCloning' (Li et al., 2011). An 18-cycle PCR amplification was performed using Q5 polymerase, followed by a 1 h DpnI digestion of the PCR product and transformation into XL10-Gold cells.

6.7 Transformations

Cells were divided into 45 µL aliquots in pre-chilled 14 mL BD Falcon tubes. If using XL10-Gold cells, 2 µL 2-mercaptoethanol was added to each aliquot of cells and incubated for 10 minutes on ice with occasional swirling. 1 µL purified DNA or 2 µL mutagenesis reaction was added and incubated with the cells for 30 minutes. Heat shock at 42°C was performed for the length of time specified in **Table 15**. Cells were then incubated on ice for 2 min, and 450 µL of pre-heated LB media (42°C) was added to each aliquot of cells. Transformations were incubated at 37°C for 1 h with shaking and cells were plated onto both LB-agar plates containing 50 µg/mL kanamycin and LB-agar plates with no antibiotic as a positive control. Plates were incubated at 37°C overnight.

Cell strain	**Length of heat shock (seconds)**
XL10-Gold	30
BL21-Gold (DE3)	20

Table 15| Required length of heat shock for competent cells used.

6.8 Plasmid DNA extractions

Plasmid DNA was extracted from cells using a QIAprep Spin Miniprep Kit (QIAGEN) according to manufacturer instructions. DNA was eluted from the spin column using 50 µL nuclease-free water.

6.9 Protease inhibition

Protease inhibitors (COmplete EDTA-free, Roche) were used at a final concentration of 1 tablet per 50 mL buffer. Protease inhibitors were included in buffers used for cell resuspension elution during protein purifications.

6.10 Expression and purification of His_6-*At*PEX5(340-728) and variants

6.10.1 Solutions

Lysis buffer (wash buffer 1)

NaH_2PO_4 (50 mM), NaCl (300 mM), 2-mercaptoethanol (10 mM), glycerol (15% v/v), pH 8.0 with ~50 mL 1M NaOH

Elution buffer

NaH_2PO_4 (50 mM), NaCl (300 mM), 2-mercaptoethanol (10 mM), glycerol (15% v/v), imidazole (200 mM), pH 8.0

Wash buffer 2

97.5% lysis buffer, 2.5% elution buffer to achieve a final imidazole concentration of 5 mM

Phosphate-buffered saline (PBS)

NaCl (8 g/L), KCl (0.2 g/L), Na_2HPO_4 (1.44 g/L), KH_2PO_4 (0.24 g/L), pH 7.4

Cobalt-agarose resin

HisPur™ Cobalt Resin (Fisher Scientific), 500 µL settled cobalt-agarose resin used per 500 mL original autoinduction culture

6.10.2 Autoinduction of His_6-*At*PEX5(340-728) and variants

The pET-28b.His_6-AtPEX5(340-728) DNA, of either wild-type PEX5 or each of the site-directed variants, was transformed into BL21-Gold (DE3) cells (section 6.7) for efficient protein expression. Overnight growth of single colonies in 5 mL selection media (LB) at 37°C was performed, followed by 8 h cultures of 20 µL of overnight culture in 1 mL selection media (LB) at 37°C. Flasks containing 500 mL ZY-AI with kanamycin (1:500) were inoculated with day culture (1:2000) and incubated at 28°C for 18 h. Cells were then pelleted by centrifugation at 4,500 x g (20 min, 4°C) and pellets were stored at -80°C until required.

6.10.3 Cell disruption of BL21-Gold (DE3) *E .coli* cells

Cell pellets from 500 mL autoinduced cell culture were transferred to 150 mL universal containers. Pellets were thawed on ice and resuspended in 20 mL chilled lysis buffer (containing protease inhibitors) per 500 mL original autoinduction culture. Cells were homogenised, transferred to the inlet of a TS Series Benchtop Cell Disruptor (Constant Systems Ltd.) and broken by applying 20 kpsi. Cells were chased through with 20 mL chilled lysis buffer containing protease inhibitors and the 40 mL lysed cell slurry was passed through the cell disruptor a second time.

6.10.4 Purification of His_6-*At*PEX5(340-728) and variants

Induced, lysed cells were centrifuged at 25,000 x g (30 min) and supernatant was loaded onto 500 µL settled Co-NTA resin (Thermo Fisher) per 40 mL supernatant. This was incubated at 4°C for 2 h with constant agitation. The resin was pelleted by centrifugation at 2,000 x g (2 min) and supernatant was decanted, leaving approximately 4 mL on the resin. Lysis buffer (section 6.1.9) was used for resin washing (2 x 7.5 mL (per 500 µL resin) followed by centrifugation at 2,000 x g (2 min) and decanting of supernatant). The resin was then washed in the same way (2 washes) with wash buffer 2. Elution buffer was then added at 5 mL per 500 µL resin, resin was incubated at 4°C for 30 min, and resin was pelleted by centrifugation at 2,000 x g (2 minutes). Eluted His_6-*At*PEX5(340-728) protein was then concentrated and buffer-exchanged into an appropriate buffer (containing protease inhibitors) (section 6.10.6).

6.10.5 Gel filtration of His_6-*At*PEX5(340-728)

Gel filtration was performed on an ÄKTA prime, on which an S75 column (GE Healthcare Life Sciences) was first washed with 600mL sterile de-gassed distilled water (3 mL/min) and equilibration was performed using 300mL lysis buffer, or low salt buffer for crystallisation screens (section 6.17.1) (3 mL/min). Protein was loaded onto the S75 column by syringe and flow rate was slowed to 1 mL/min for 45 minutes. Flow rate was then further slowed to 0.2 mL/min and 2mL fractions were collected over a period of approximately 2 h. Fractions were analysed by denaturing SDS-PAGE and the fractions containing purest looking protein were confirmed. These fractions were then pooled and concentrated to the desired concentration.

6.10.6 Concentration and buffer exchange of His_6-*At*PEX5(340-728)

Concentrators (Millipore, 30 kDa cut-off) were equilibrated with 5 mL distilled water (centrifugation at 4,500 x g, 6 minutes), followed by 5 mL of the appropriate buffer (the current buffer of the protein). Protein solution was loaded onto a concentrator and concentrated to 1 mL. 4 mL of the desired buffer for the protein (containing protease inhibitors) was then added and concentrators centrifuged for 15 min (x 2). Protein solutions were then topped up with 2 x 4 mL desired buffer (with 2 x 15 min centrifugation after each top-up). Finally, protein solutions were concentrated to 1 mL and stored at 4°C for a maximum of one month.

6.10.7 Protein concentration determination

Protein concentration was determined by the Beer-Lambert law using the absorbance at 280 nm, measured using a Kontron Uvikon 930 spectrophotometer. Extinction coefficients were calculated using ExPASy ProtParam.

6.11 SDS-PAGE

Samples were prepared by the addition of 2 x SDS-PAGE loading buffer (1:1) and samples were boiled for denaturation in a PTC-100® Thermal Cycler (MJ Research) at 100°C for 5 min. SDS-PAGE gels were made up of 60–70% resolving gel and 30–40% stacking gel. Gels were prepared in a BIO-RAD gel casting system. Prepared samples were loaded into wells and gels were run at 30 mA if running one gel (60 mA if running more than one) in SDS running buffer. Gels were transferred into Quick Blue stain (20 mL per gel), left to stain for 30 min, transferred to distilled water for 30 min and then imaged using an InGenius gel imager (Syngene).

<u>Stacking gel for SDS-PAGE (2 gels)</u>

30% acrylamide (625 µL/5 mL), 1 M Tris·HCl (625 µL/5 mL, pH 6.9), SDS (10% solution) (50 µL/5 mL), distilled water (3.65 mL/5 mL), ammonium persulfate (25% solution) (50 µL/5 mL), TEMED (5 µL/5 mL)

Resolving gel for SDS-PAGE (2 gels)

30% acrylamide (7.5 mL/15 mL), 1.5 M Tris·HCl (3.75 mL/15 mL, pH 8.8), SDS (10% solution) (150 µL/15 mL), distilled water (3.5 mL/15 mL), Ammonium persulfate (25% solution) (100 µL/15 mL), TEMED (10 µL/15 mL)

2 x SDS-PAGE loading buffer

Tris·HCl (50 mM, pH 6.8), DTT (100 mM), SDS (2% w/v), bromophenol blue (0.1% w/v), glycerol (10% v/v)

SDS-PAGE running buffer (10x)

Tris base (30.2 g/L), glycine (144 g/L), SDS (10 g/L). Used at a 1 x working solution

Quick Blue - Coomassie stain

20 mL Quick Blue (Triple Red) was used per gel

Blue Prestained Protein Standard (Broad Range (11-190 kDa), P7706S, NEB)

5 µL used in the first lane of each SDS-PAGE gel

6.12 Blotting

6.12.1 Antibodies

Primary

- Anti-polyhistidine from mouse (Sigma) used at a 1:3,000 dilution
- Anti-*At*PEX5(340-728) raised in rabbit (Genosphere) used at a 1:10,000 dilution

Secondary

- Anti-mouse IgG from rabbit conjugated to horse radish peroxidase (Sigma) used at a 1:20,000 dilution
- Anti-rabbit IgG from goat conjugated to horse radish peroxidase (Sigma) used at a 1:5,000 dilution

HRP-conjugates

- HRP-YQSKL used at a final concentration of 350 nM
- HRP-YQSEL used at a final concentration of 350 nM

- HRP-YQSKV used at a final concentration of 350 nM
- HRP-YQSEV used at a final concentration of 350 nM

6.12.2 Western blotting

After running an SDS-PAGE gel (section 6.11), the gel was placed in transfer buffer in preparation for transfer using a Mini Trans-Blot® Electrophoretic Transfer Cell (BIO-RAD) and a transfer stack was assembled according to manufacturer instructions. Transfer of proteins from the gel to the nitrocellulose membrane was performed for 1 h at 100V. The nitrocellulose membrane was then placed in blocking buffer for 1 h at room temperature (if using 5% skimmed milk powder in blocking buffer) or at 4°C overnight (if using 3% BSA in blocking buffer). Blocking buffer was removed and the membrane was incubated for 1 h at room temperature in primary antibody solution. The membrane was washed with PBS-T (3 x 10 min washes), followed by a 45 min incubation with secondary antibody solution at room temperature. If using HRP-SKL or PEX14(1-154)-HRP conjugates, these were diluted in blocking buffer at an appropriate dilution and membranes were incubated with these solutions for 1 h immediately after the blocking step. The membrane was washed with PBS-T (3 x 10 min washes) followed by a wash with PBS (10 min). Membranes were then visualised with increasing levels of exposure on a GeneGnome^5 system (Syngene) after incubation (1 min) with ECL Prime Western blotting detection reagent (VWR International).

PBS-T

PBS, Tween-20 (0.05% v/v)

Blocking buffer

5% skimmed dried milk powder or 3% BSA in PBS-T

Antibody solution

Blocking buffer (10 mL per membrane) with antibody at the appropriate dilution

Transfer buffer

Tris base (3.02 g/L), glycine (14.4 g/L), methanol (20% v/v)

6.12.3 Dot blotting

An appropriate amount of the protein to be analysed was dotted onto nitrocellulose membrane. Blocking of the membrane was performed as for western blotting, as were all downstream steps of the procedure.

6.12.4 Colony blotting

Saturation primers were used in a mutagenesis reaction (section 6.5), and mutagenesis reactions were transformed directly into an expression cell line (BL21-Gold (DE3)). Colonies were picked using cocktail sticks and were used to inoculate separate wells of 96-deep-well plates, each containing 1.5 mL selection media (LB with kanamycin (50 mg/L)). Cells were grown for 20 h at 24°C in a microplate incubator shaker, and were then transferred into 384-well plates, with each well containing 70 µL selection media, using the 're-arraying' function of a Genetix QPix colony picker. Following the growing of cells for 20 h at 24°C in a microplate incubator shaker, cells were stamped onto nitrocellulose membrane using the 'gridding' function of a Genetix QPix colony picker. Remaining cells in the 96-deep-well plates were centrifuged at 2,000 x g for 10 min, supernatant was removed, and cells were stored at -80°C for retrospective identification by DNA purification and sequencing.

Nitrocellulose membranes were transferred (colony side up) to LB-agar-kanamycin plates, without introducing air bubbles. Plates were incubated (inverted) for 14 h at 30°C, and nitrocellulose membranes were then transferred onto fresh LB-agar plates containing 50 mg/L kanamycin and 250 µM IPTG. Plates were incubated (inverted) for 4 h at 37°C to induce protein expression. Cell lysis and blotting steps were performed as per the QIA expressionist (The QIAexpressionist, 2001), with the exception of the blocking step, which was carried out overnight (without agitation) at 4°C. Blotting was carried out using HRP-peptide conjugates (HRP-YQSKL/EL/KV/EV) at a final concentration of 350 nM. Nitrocellulose membranes were then visualised, with 30 seconds exposure time, on a GeneGnome^5 system (Syngene) after incubation (1 min) with ECL Prime Western blotting detection reagent (VWR International).

SDS solution

10% (w/v) sodium dodecyl sulfate

Denaturing solution

NaOH (0.5 M), NaCl (1.5 M)

Neutralisation solution

NaOH (1.5 M), Tris base (0.5 M), pH 7.5

2 x SSC

NaCl (8.76 g per 500 mL), trisodium citrate·$2H_2O$ (5.02 g per 500 mL)

TBS

Tris base (10 mM), NaCl (150 mM), pH 7.5

TBS-T

Tris base (20 mM), NaCl (500 mM), Tween-20 (0.05% v/v), pH 7.5

Blocking solution

5% skimmed dried milk powder in TBS-T

6.13 Peptide synthesis

General Information

Solvents were purchased from Fluka, Sigma-Aldrich, or Fisher Scientific. All Fmoc-protected amino acids, pre-loaded 2-chlorotrityl resin (0.6–0.9 mmol/g), HCTU (2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate), Oxyma Pure® (ethyl (hydroxyimino)cyanoacetate) and N-Biotinyl-NH-PEG_2-COOH were purchased from Novabiochem. EDC (N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride) was purchased from Sigma-Alrich. Dansyl chloride was purchased from Sigma-Aldrich and will be referred to as dansyl from this point onwards. 7-Methoxycoumarin-4-acetic acid was purchased from Novabiochem and will be referred to as coumarin from this point onwards. Lissamine™ Rhodamine B sulfonyl chloride (mixed isomers) was purchased from Thermo Fisher Molecular Probes and will be referred to as lissamine from this point onwards.

N-terminal Fmoc-protected amino acids and pre-loaded 2-chlorotrityl resins with the following side chain protection were used: Ala, Asp(OtBu), Glu(OtBu), Phe, Gly,

His(Trt), Ile, Lys(Boc), Leu, Met, Asn(Trt), Pro, Gln(Trt), Arg(Pbf), Ser(tBu), Thr(tBu), Val, Trp(Boc), Tyr(tBu). Manual Fmoc solid phase peptide synthesis was performed using reservoir tubes with a polyethylene frit (Grace Discovery Sciences) and a vacuum tank attached to a water aspirator. Automated Fmoc solid phase peptide synthesis was performed using a CEM Liberty® automated microwave assisted peptide synthesiser, with double coupling (20°C, 60 min) used for each amino acid addition.

6.13.1 General procedure for Fmoc-protected amino acid coupling (manual SPPS)

DMF (2 mL) was added to pre-loaded 2-chlorotrityl chloride resin (1 eq.) and the resin left to swell for 1h. DMF was drained from the reaction vessel. A solution of Fmoc-protected amino acid (5 eq.) in 1 mL DMF, HCTU (5 eq.) in 1 mL DMF, and DIPEA (10 eq.) were added and the reaction mixture agitated for 1h at RT. The solution was filtered from the resin by vacuum filtration and resin was washed with DMF (3 x 2 mL x 2 min), 20% piperidine in DMF (5 x 2 mL x 2 min), and DMF (5 x 2 mL x 2 min). This coupling process was repeated three additional times, resulting in a pentamer. Resin was then washed with DMF (3 x 2 mL x 2 min), DCM (3 x 2 mL x 2 min) and MeOH (3 x 2 mL x 2 min) before drying of the resin *in vacuo*. The crude resin-bound peptide was then stored at -20°C, and then weighed out for use as required. Either direct cleavage and purification of peptides was performed (section 6.13.9), or N-terminal coupling of a fluorescent label (sections 6.13.4, 6.13.5, and 6.13.6) or a biotin moiety (sections 6.13.7 and 6.13.8) was carried out.

6.13.2 Procedure for Fmoc-protected amino acid coupling (automated SPPS)

Automated Fmoc SPPS was performed on a 0.1 mmol scale using a CEM Liberty® automated microwave assisted peptide synthesiser. Pre-loaded 2-chlorotrityl resin was swelled in DMF for at least 1h prior to synthesis, and all amino acid coupling steps were performed using double coupling with no microwave. Following peptide synthesis, resin was washed with DMF (3 x 2 mL x 2 min), DCM (3 x 2 mL x 2 min) and MeOH (3 x 2 mL x 2 min) before drying of the resin *in vacuo*. The crude resin-bound peptide was then stored at -20°C until required.

6.13.3 Split-and-pool amino acid coupling

Peptide sub-libraries were synthesised by first adding 30 mg of each required preloaded 2-chlorotrityl chloride resin into a reaction vessel. The pooled resin was swelled in DMF (5 mL) for 1 h, and then the reaction vessel was drained. DMF washes were performed (3 x 2 mL x 2 min), followed by washes with 20% piperidine in DMF (5 x 2 mL x 2 min) and DMF (5 x 2 mL x 2 min). Resin was washed with DCM (3 x 2 mL x 2 min) and dried *in vacuo* before splitting equally between reservoir tubes for coupling of each of the amino acids at position -2. To each aliquot of resin, a different Fmoc-protected amino acid (5 eq.) in DMF (1 mL) was added, along with HCTU (5 eq.) in DMF (1 mL) and DIPEA (10 eq.). The reaction was agitated for 1 h (at RT). All aliquots of resin were pooled for the deprotection (20% piperidine in DMF) and wash (DMF) steps, and all subsequent amino acids were coupled using standard protocol (see section 6.13.1).

6.13.4 N-terminal lissamine coupling

Lissamine couplings were performed under an inert atmosphere in foil-covered round-bottom flasks. Lissamine™ Rhodamine B sulfonyl chloride (mixed isomers) (3 eq.) in DMF (1 mL) was added to peptide-bound resin (1 e.q.) at 0°C. Dry DIPEA (10 eq.) was added and the solution was stirred overnight. The solution was filtered from the resin, which was then washed with DMF (3 x 2 mL x 2 min), DCM (3 x 2 mL x 2 min) and MeOH (3 x 2 mL x 2 min) before drying of the resin *in vacuo*. Cleavage of peptides from resin was performed (see section 6.13.9), and peptides were purified by mass-directed preparatory HPLC (gradient elution 5–95% MeOH in H_2O). Solvent was removed *in vacuo* using a Büchi rotary evaporator and peptides were lyophilised to give a dark pink solid (yield = 15–20%). Before use in fluorescence anisotropy, peptides were resuspended in sterile H_2O to produce a 500μM solution.

6.13.5 N-terminal coumarin coupling

Peptide-bound resin (1 eq.) was swelled in DMF (2 mL) for 1 h. DMF was drained from the reaction vessel. 7-Methoxycoumarin-4-acetic acid (4 eq.) in DMF (1 mL) was added to EDC (4.4 eq.) and Oxyma Pure® (8 eq.) in DMF (1 mL), and the solution was well-mixed then incubated at RT for 10 min. The coumarin mixture was then added to the peptide-bound resin (1 eq.) and agitated for 4 h. The solution was filtered from the resin, which was then washed with DMF (3 x 2 mL), DCM (3 x 2 mL) and MeOH (3 x 2 mL) before drying of the resin *in vacuo*. This method was modified from a published

protocol (Pennington, 1994). Cleavage of peptides from resin was performed (see section 6.13.9), and peptides were purified by mass-directed preparatory HPLC (gradient elution 5–95% MeOH in H_2O). Solvent was removed *in vacuo* and peptides were lyophilised to give a pale yellow solid (yield = 15–25%).

6.13.6 N-terminal dansyl coupling

Peptide-bound resin (1 eq.) was swelled in DMF (2 mL) for 1 h. DMF was drained from the reaction vessel. DIPEA (6 eq.) was added to dansyl chloride (5 eq.) in DMF (2 mL), and the solution was well-mixed and then incubated at RT for 10 min. The dansyl mixture was then added to the peptide-bound resin (1 eq.) and stirred overnight. The solution was filtered from the resin which was then washed with DMF (3 x 2 mL), DCM (3 x 2 mL) and MeOH (3 x 2 mL) before drying of the resin *in vacuo*. Cleavage of peptides from resin was performed (see section 6.13.9), and peptides were purified by mass-directed preparatory HPLC (gradient elution 5–95% MeOH in H_2O). Solvent was removed *in vacuo* and peptides were lyophilised to give a yellow solid (yield = 86–93%).

6.13.7 N-terminal (+)-biotin coupling

Peptide-bound resin (1 eq.) was swelled in DMF (2 mL) for 1 h. DMF was drained from the reaction vessel. (+)-Biotin (6 eq.) was dissolved in DMF:DMSO (1:1) and HCTU (6 eq.) was added to the biotin solution. DIPEA (10 eq.) was then added to the biotin + HCTU solution, and the resulting activated biotin solution was added to the peptide-bound resin (1 eq.) and stirred overnight. The solution was filtered from the resin which was then washed with DMF:DMSO (1:1) (2 x 2 mL), DMF (2 x 2 mL), DCM (3 x 2 mL) and MeOH (3 x 2 mL) before drying of the resin *in vacuo*. Cleavage of peptides from resin was performed (see section 6.13.9), and peptides were purified by mass-directed preparatory HPLC (gradient elution 5–95% MeOH in H_2O). Solvent was removed *in vacuo* and peptides were lyophilised to give a white solid (yield = 38%).

6.13.8 N-terminal biotin-$(PEG)_2$ coupling

Peptide-bound resin (1 eq.) was swelled in DMF (2 mL) for 1 h. DMF was drained from the reaction vessel. *N*-Biotinyl-NH-PEG_2-COOH·DIPEA (2 eq.) and HCTU (2 eq.) were added to the peptide-bound resin (1 eq.) and stirred for 3 h. The solution was filtered from the resin which was then washed with DMF (3 x 2 mL), DCM (3 x 2 mL) and MeOH (3 x 2 mL) before drying of the resin *in vacuo*. Cleavage of peptides from resin

was performed (see section 6.13.9), and peptides were purified by mass-directed preparatory HPLC (gradient elution 5–95% MeOH in H_2O). Solvent was removed *in vacuo* and peptides were lyophilised to give a white solid (yield = 35–40%).

6.13.9 Cleavage of peptides from the solid resin support

A cleavage cocktail of TFA, H_2O, and TIPS (95:2.5:2.5) was added to dried resin (500 µL per 25 mg resin) and agitated for 1h at RT. The cleavage mixture was filtered drop-wise into cold diethyl ether (1:100) and precipitated peptide was collected by centrifugation. Three diethyl ether washes were performed, all diethyl ether collected and water extraction carried out. Diethyl ether-precipitated peptide was left open to air overnight to allow for evaporation of diethyl ether and was then resuspended in sterile H_2O and pooled with the aqueous fraction. Preparative HPLC was performed if further purification was required. After HPLC, peptides were lyophilised and resuspended in sterile water to the desired concentration.

6.14 Fluorescence anisotropy

6.14.1 Solutions

FA buffer

HEPES (20 mM), NaCl (150 mM), pH 7.5

Plate blocking buffer

Gelatin from porcine skin, type A (Sigma) (0.32 mg/mL) in FA buffer

6.14.2 Fluorescently labelled peptide solutions

After N-terminal lissamine coupling, peptides were purified by HPLC, then lyophilised and re-dissolved to produce a 500 µM solution. Lissamine-labelled peptides were used at a final concentration of 100 nM for protein titration assays, and a final concentration of 30 nM for peptide competition assays.

6.14.3 Protein solutions

After concentration and buffer exchange of proteins into FA buffer, dilution to an appropriate concentration was performed for binding assays.

6.14.4 General assay information

Fluorescence anisotropy studies used an EnVision™ 2103 multilabel plate reader (Perkin Elmer) All required wells of a Black Optiplate™-384 F (PerkinElmer) 384-well plate was blocked with 80 µL plate blocking buffer per well, sealed with self-adhesive plate seals (Fasson®) and stored at 4°C at least 14 hours prior to use.

6.14.5 Protein titration

To determine binding affinity of a protein to a lissamine-labelled peptide, a dilution series of *At*PEX5-C (wild-type or variant) (0.3 nM–10 µM) was measured with 100 nM (final concentration) lissamine-peptide (final volume 40 µL).

Plate blocking buffer (60 µL) was removed from each well, after the ≥14 h blocking incubation. Protein solution (40 µL) was added to wells of rows A–F in column 1, this was agitated by pipetting, and 40 µL was removed and added to wells in column 2. This process was repeated until the protein was diluted along all 24 columns of the plate (leaving column 24 with no protein solution). FA buffer was added to rows A–C of the plate (20 µL per well) to act as the blanks, and peptide solution (20 µL per well) (at 200 nM) was added to rows D–F, to test binding of the peptide to the protein.

6.14.6 Peptide competition assay

To determine the inhibition constant for peptides competing with lissamine-YQSKL for the binding site of *At*PEX5-C, a number of stock solutions were prepared for each competing peptide. Final concentration of competing peptide in the assay ranged from 0 to 1 mM. Final lissamine-YQSKL concentration was 30 nM and final protein concentration was 200 nM in each well.

Only columns 1–14 were used in these assays. Plate blocking buffer (70 µL) was removed from each well, after the ≥14 h blocking incubation. Stock solutions of competing peptides were prepared (14 solutions, from 0–4 mM) and 10 µL of each was added into separate wells of columns 1–14 (rows A–F). 10 µL of lissamine-YQSKL (stock concentration 120 nM) was added into each well of columns 1–14 (rows D–F), and 10 µL of FA buffer was added into each well of columns 1–14 (rows A–C) to act as the blanks. 10 µL protein solution (stock concentration 800 nM) was then added to all wells for the assay.

6.14.7 Reading of fluorescence anisotropy plates

Plates were agitated with linear shaking (900 rpm) using a Envision^{TM} 2103 multilabel plate reader (PerkinElmer) for 15 minutes, and values obtained by reading using a BODIPY TMR dichroic mirror (555 nm) and the following filters:

Excitation: BODIPY TMR FP 531 (Wavelength 531 nm, bandwidth 25 nm)
Emission 1: BODIPY TMR FP P-pol 595 (Wavelength 595 nm, bandwidth 60 nm)
Emission 2: BODIPY TMR FP S-pol 595 (Wavelength 595 nm, bandwidth 60 nm)

Plates were read at a measurement height of 7.4 mm with a g-factor of 1.16. Each well received 30 flashes per measurement. Anisotropy values were obtained by applying the following formula to the blank corrected P-values and S-values:

Anisotropy (r) = 1000*(S-G*P)/(S + 2*G*P)

Amount of peptide bound was then calculated using anisotropy values, and formulas previously published (Skoulding et al., 2015). The mean values and standard deviations for each triplicate were plotted using OriginPro 9.1, with protein concentration plotted along the x axis (logarithmic scale).

6.15 Pull-down-LC-MS screening

6.15.1 Pull-down of binding peptides by Co-NTA purification of *At*PEX5-C protein

Purified protein (final concentration: 12.5 µM) was added to a library of peptides (final concentration: 500 nM each peptide in the library (concentration calculated based on the average molecular weight of the peptides in the sub-library)) in a 500 µL reaction mixture in lysis buffer (wash buffer 1 (section 6.10.1)), and this was incubated at 4°C for 1 h with agitation. This protein-peptide mixture was added to Co-NTA resin (100 µL settled resin per 500 µL reaction) and was incubated at 4°C for 1 h with agitation. Supernatant was removed and 500 µL wash buffer was added to the resin (4 x 500 µL washes) followed by wash buffer containing 5 mM imidazole (3 x 500 µL washes). Elution of the protein was performed by incubating the resin with wash buffer containing 6 M urea (300 µL) at 4°C for 30 min with agitation. Eluate was collected in tapered vials and 10 µL was injected for analysis by ESI-LC-MS.

6.15.2 Mass spectrometry for identification of binding peptides

Electrospray ionisation LC-MS was performed using a Bruker MaXis Impact time of flight mass spectrometer in the positive ion mode. Column guard used: Waters Acquity UPLC Peptide CSH C18 column 130 angstrom, 1.7 µm stationary phase Vanguard precolumn (column dimensions 2.1 mm x 5 mm). Analytical column used: Waters Acquity UPLC Peptide CSH C18 column 130 angstrom, 1.7 µm stationary phase (column dimensions 2.1 mm x 100 mm). A Dionex Ultimate 3000 HPLC was used, with solvents: A) water + 0.1% formic acid; B) acetonitrile + 0.1% formic acid. Timetable (flowrate 0.7 mL/min, linear gradients between points): -1.3 min = 99:1 A:B (pre-equilibration); 0 min = inject; 0.3 min = 99:1 A:B; 1.5 min = 80:20 A:B; 2.5 min = 78:22 A:B; 4.5 min = 70:30 A:B; 5.5 min = 60:40 A:B; 6 min = 5:95 A:B; 7.5 min = 1:99 A:B. Calibration of the mass spectrometry instrument was performed using sodium formate, injected at the end of each run.

6.15.3 Data processing

Following ‘elution’ of peptides upon unfolding of the protein after peptide pull-down, LC-MS was performed and extracted ion chromatograms (EICs) were added for the mass of each peptide in the full library of dansyl-labeled peptides. An Excel sheet was exported containing the area-under-EIC for each EIC added, along with corresponding potential peptide sequence (based on mass) and retention time from the LC column. The retention times and potential peptide sequences were compared to a master sequence-retention time list, and the ‘best match’ peptide sequence was found using this information. Overall drift times (up to ±5 seconds) were applied to the entire list of retention times in the exported data set in order to find the lowest average difference between the sequence-retention time master list and the raw data (**Figure 77**). There were 45 pairs of peptides with identical masses and retention times, so these could not be distinguished from one another; however, the peak (if present) could be assigned as belonging to the pair. One of each of these pairs was included in the sequence-retention time master list (shaded in pale blue) and the other, for each pair, can be found in the table entitled ‘Peptides with identical exact mass and retention times to 45 peptides in the retention time-sequence list’. If a peptide from one of these pairs was found in a data set after pull-down-LC-MS, the corresponding peptide in the pair was automatically added to the output matrix, as we could not be certain which peptide had actually been pulled down in these cases.

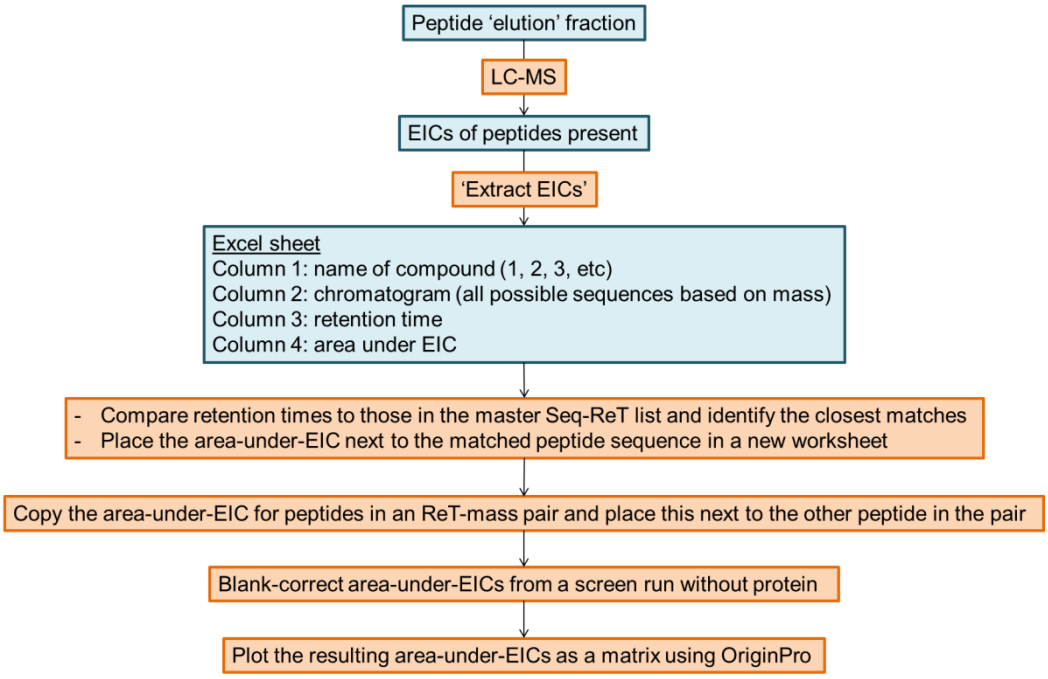


Figure 77| Data processing pipeline after the capture of peptides using purified protein associated with cobalt resin, and 'elution' of peptides by unfolding of the protein with 6 M urea. LC-MS, liquid chromatography-mass spectrometry; EIC, extracted ion chromatogram; Seq-ReT, sequence and retention time; ReT-mass, retention time and mass.

6.16 Circular dichroism

6.16.1 Buffer exchange of His_6-*At*PEX5(340-728)

Buffer exchange was performed as previously described (section 6.10.6). The protein was buffer-exchanged into 10 mM sodium phosphate before circular dichroism (CD) measurements.

6.16.2 Measurement of circular dichroism

CD was measured by loading 250 µL protein (at 0.2 mg/mL in 10 mM sodium phosphate, pH 7.5) into a 0.1 mm pathlength cuvette. CD measurements were taken in 1 nm increments from 180 nm to 260 nm at 21°C. HT voltage was also measured to ensure that this did not exceed 500V, in which case the CD would be an unreliable measurement.

6.17 Crystallisation screens

6.17.1 Gel filtration, concentration and buffer exchange of His_6-*At*PEX5(340-728)

This was performed as previously described (sections 6.10.5 and 6.10.6) after protein purification. The buffer used for setting up crystal screens was a low salt buffer, to try to reduce the chance of salt crystals forming.

Low salt buffer

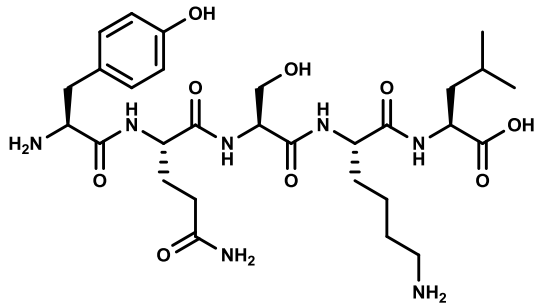
NaCl (50 mM), HEPES (5 mM), pH 7.5, filtered and de-gassed

6.17.2 Setting up of crystal screens

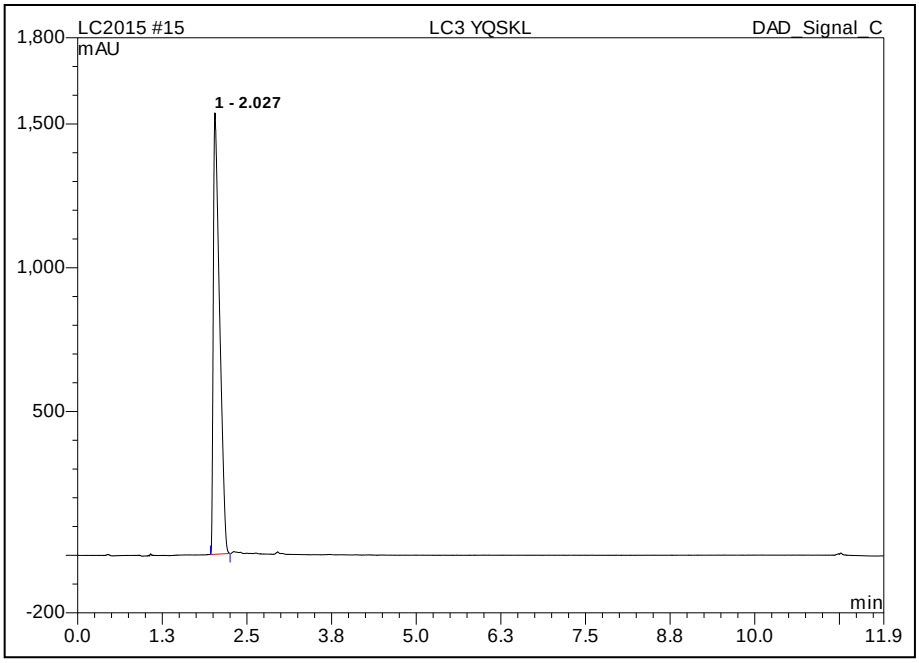
Five 96-well plate screens were prepared: Hampton 1 and 2, Index 1 and 2, Wizard 1, 2, 3 and 4, and SaltRx 1 and 2. 80 µL of each screen was placed into wells of separate 96-well plates. A Douglas Crystallisation Robot was used to load 1 µL His_6-*At*PEX5(340-728) onto each platform with 1 µL of screen. 1 µL low salt buffer was also loaded onto a platform alongside the protein, along with 1 µL of screen solution. After all screens were prepared, lines of the Douglas robot were washed with 70% ethanol, and screen plates were covered with clear adhesive seal. Plates were analysed by light microscopy to record any immediate amorphous precipitate, and plates were then stored at 18°C. Plates were analysed by light microscopy after 7, 9, 11, 13, 15 and 18 days in order to record any precipitate, microcrystals or ‘plate’ structures formed.

6.18 Synthetic peptide analytical data

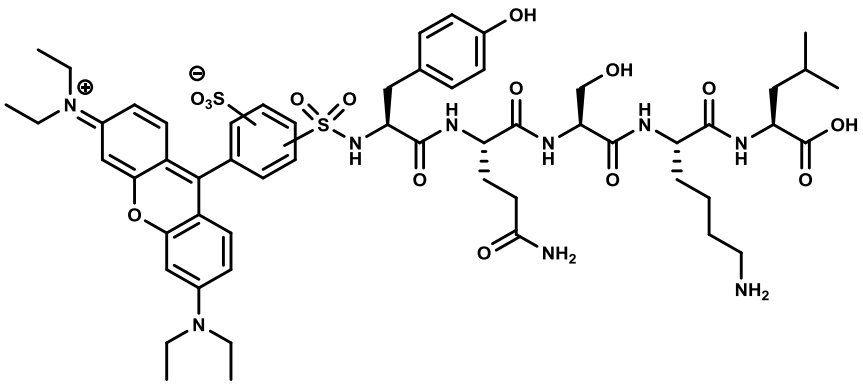
6.18.1 H_2N-YQSKL-CO_2H



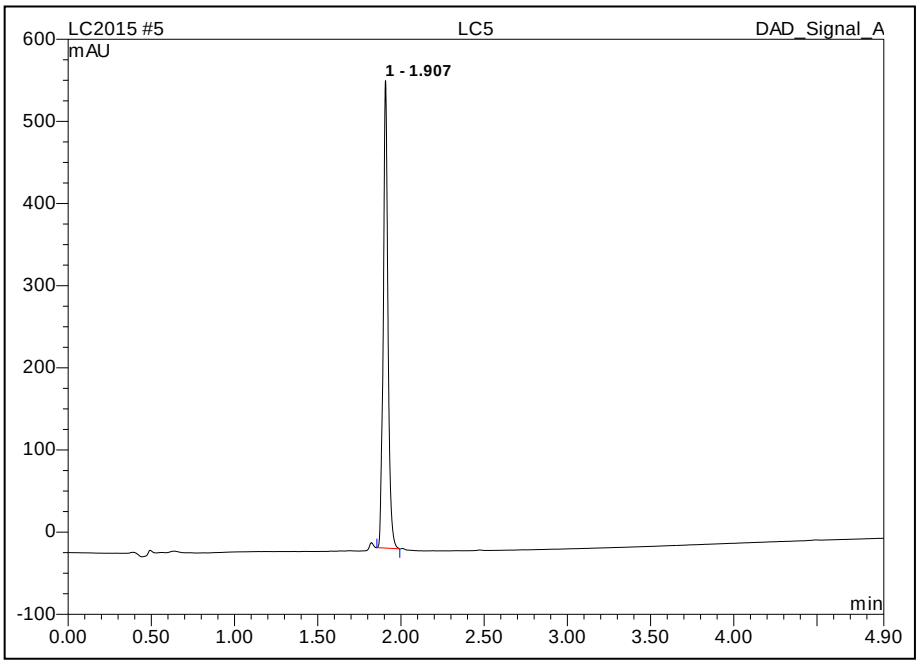
Ascentis Peptide (ES-C18, 100 x 2.1 mm, 2.7 µm particle size); Flow rate 0.5 mL min^{-1}; Gradient elution t = 0 min, MeCN:H_2O with 0.1 % TFA, 5:95; t = 10.0 min, 95:5; t = 10.5 min, 5:95; t = 12.0 min, 5:95; Rt 2.0 min, *m/z* (ES) found $[M+H]^+$ 638.3512; $C_{29}H_{48}N_7O_9$ requires 638.3508.



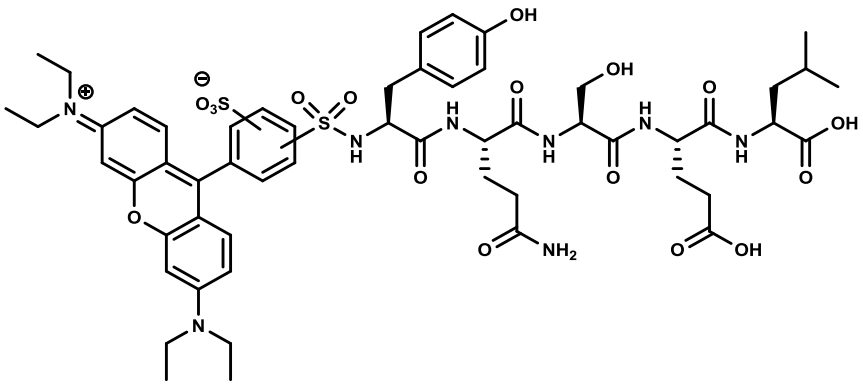
6.18.2 Lissamine-YQSKL-CO_2H



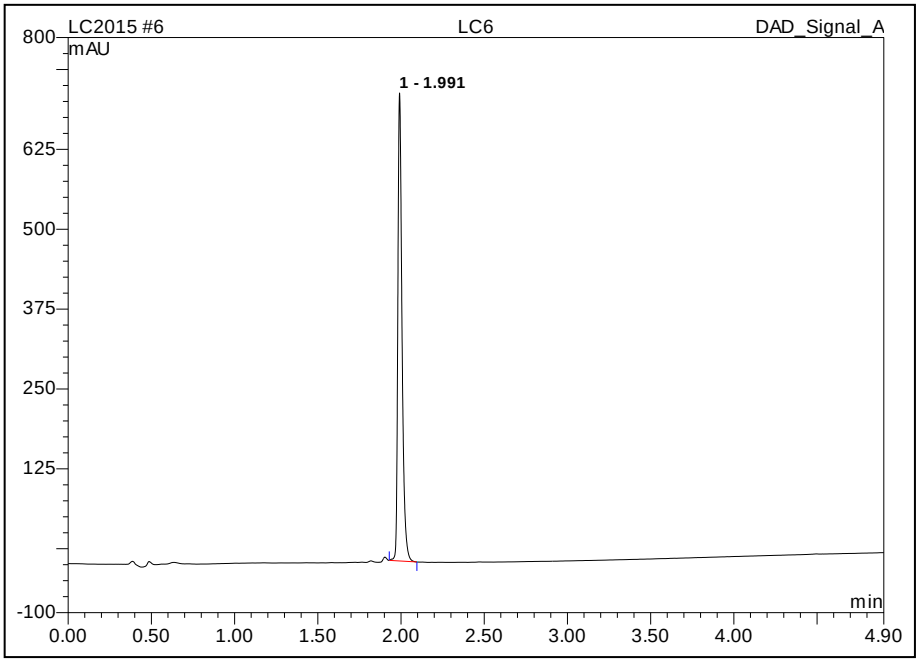
Ascentis Express (C18, 50 x 2.1 mm, 2.7 µm particle size); Flow rate 0.5 mL min^{-1}; Gradient elution t = 0 min, MeCN:H_2O with 0.1 % TFA, 5:95; t = 5.0 min, 95:5; Rt 1.9 min, *m/z* (ES) found $[M+H]^+$ 1178.4906; $C_{56}H_{76}N_9O_{15}S_2$ requires 1178.4903.



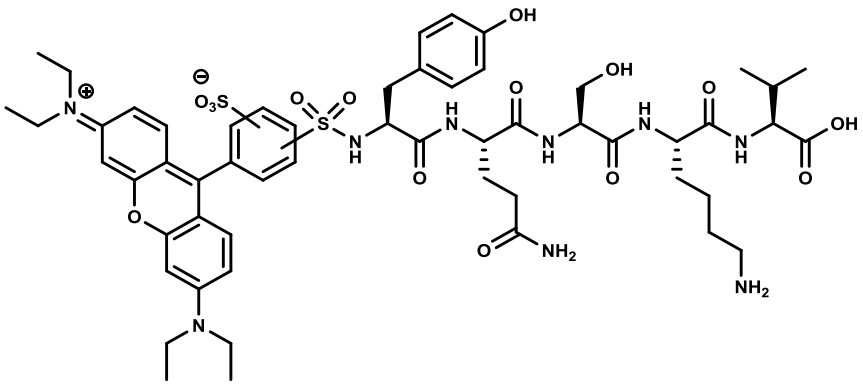
6.18.3 Lissamine-YQSEL-CO_2H



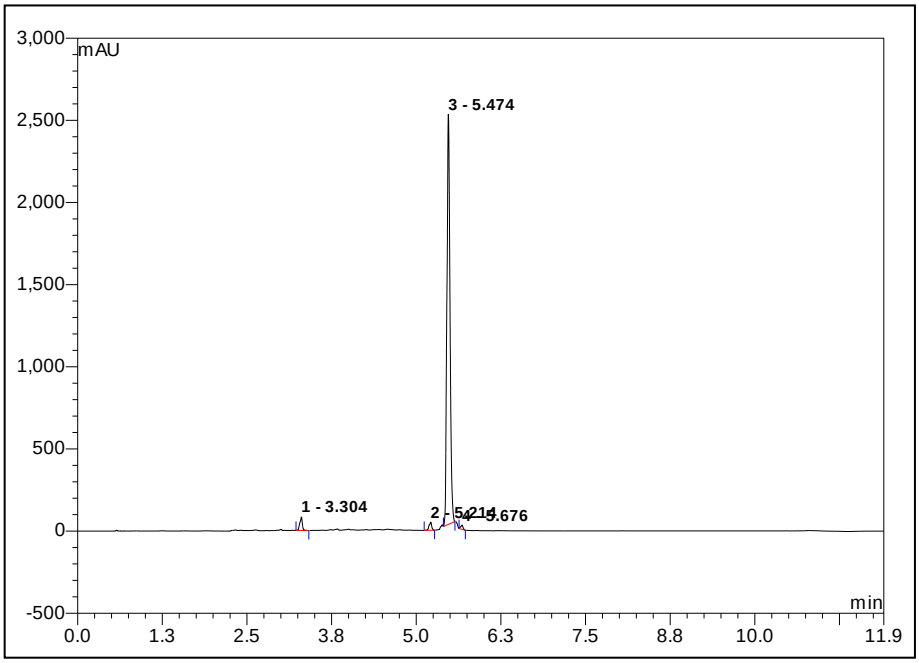
Ascentis Express (C18, 50 x 2.1 mm, 2.7 µm particle size); Flow rate 0.5 mL min^{-1}; Gradient elution t = 0 min, MeCN:H_2O with 0.1 % TFA, 5:95; t = 5.0 min, 95:5; Rt 2.0 min, *m/z* (ES) found $[M+H]^+$ 1179.4397; $C_{55}H_{71}N_8O_{17}S_2$ requires 1179.4379.



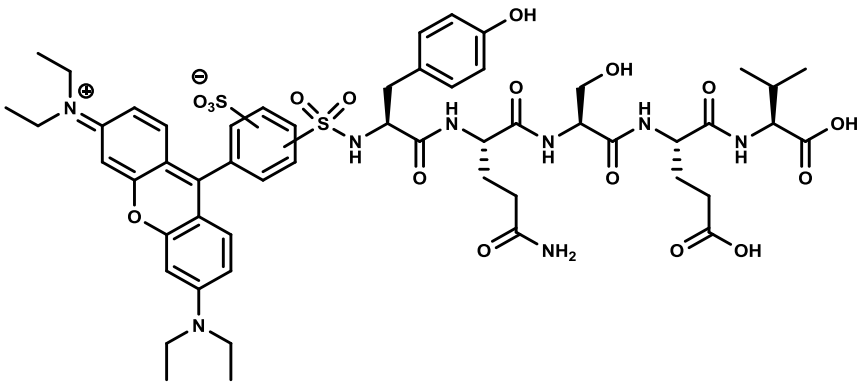
6.18.4 Lissamine-YQSKV-CO_2H



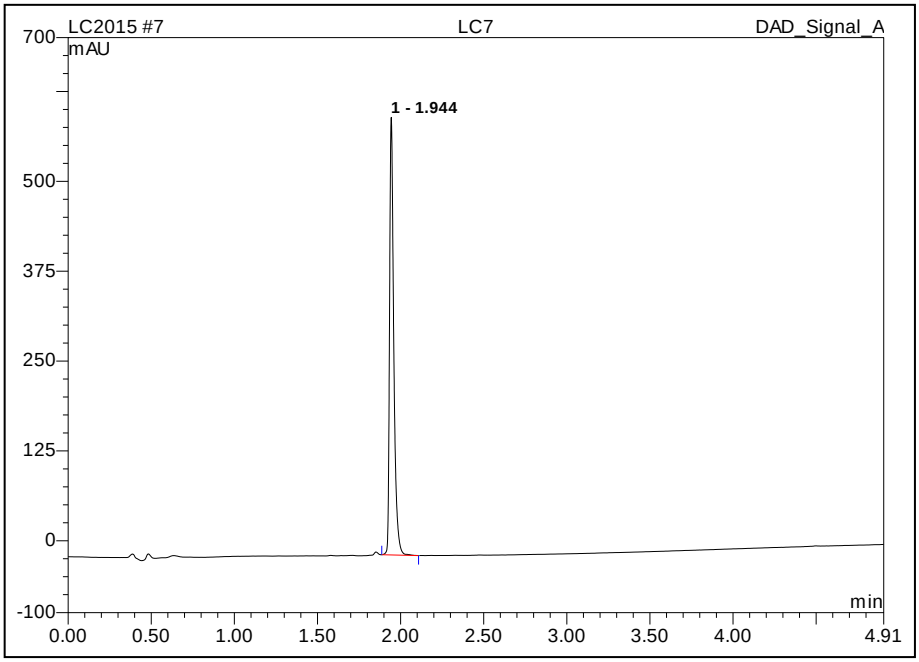
Ascentis Peptide (ES-C18, 100 x 2.1 mm, 2.7 µm particle size); Flow rate 0.5 mL min^{-1}; Gradient elution t = 0 min, MeCN:H_2O with 0.1 % TFA, 5:95; t = 10.0 min, 95:5; t = 10.5 min, 5:95; t = 12.0 min, 5:95; Rt 5.5 min, *m/z* (ES) found $[M+H]^+$ 1164.4752; $C_{55}H_{74}N_9O_{15}S_2$ requires 1164.4746.



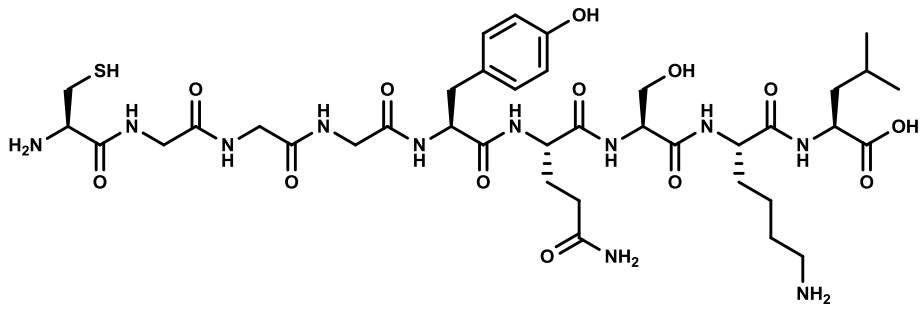
6.18.5 Lissamine-YQSEV-CO_2H



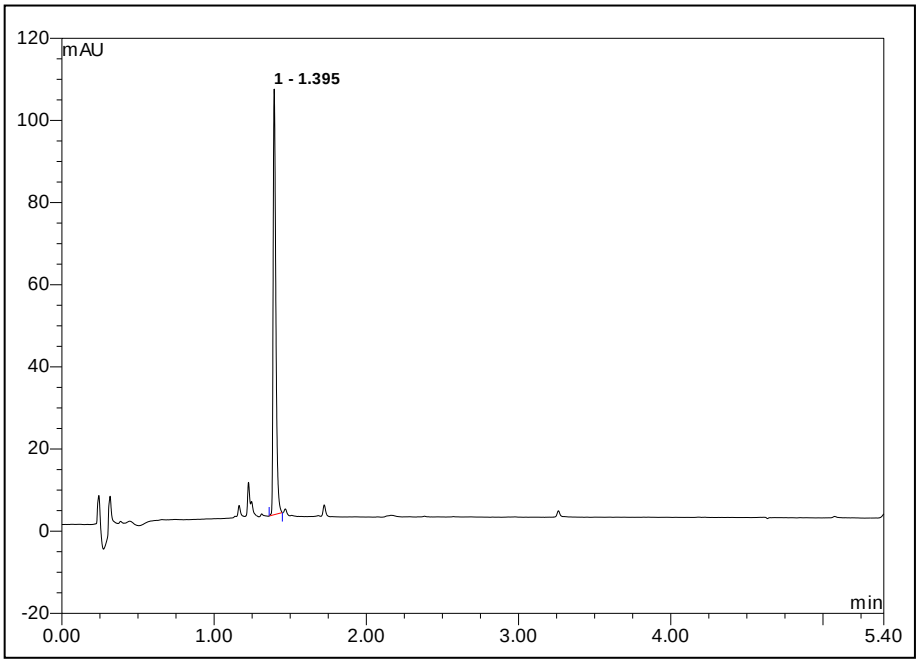
Ascentis Express (C18, 50 x 2.1 mm, 2.7 µm particle size); Flow rate 0.5 mL min^{-1}; Gradient elution t = 0 min, MeCN:H_2O with 0.1 % TFA, 5:95; t = 5.0 min, 95:5; Rt 1.9 min, *m/z* (ES) found $[M+H]^+$ 1165.4216; $C_{54}H_{69}N_8O_{17}S_2$ requires 1165.4223.



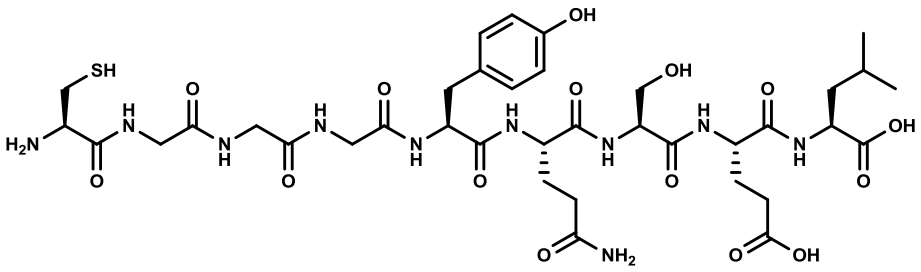
6.18.6 H_2N-CGGGYQSKL-CO_2H



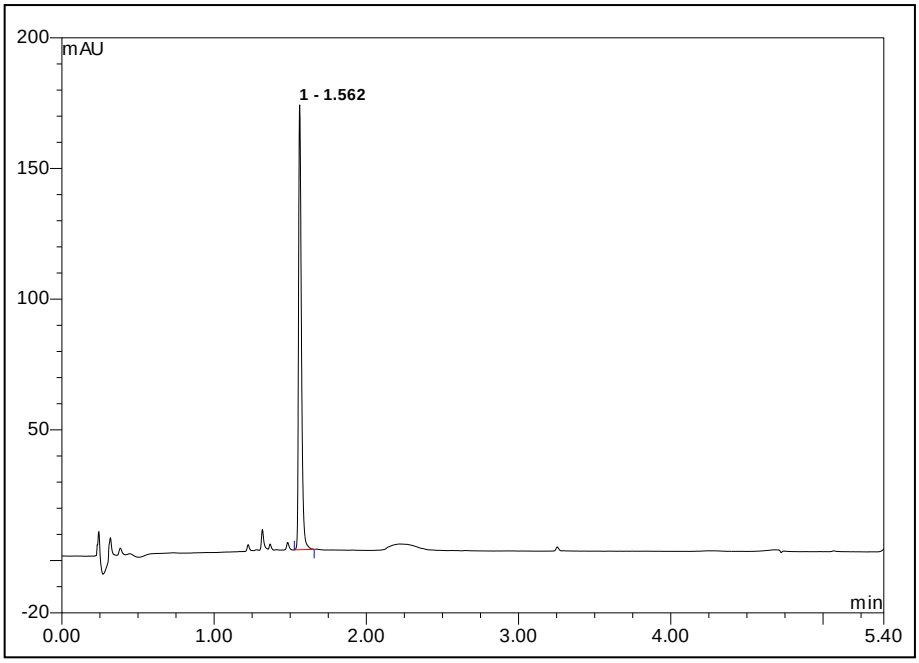
Ascentis Express (C18, 50 x 2.1 mm, 2.7 µm particle size); Flow rate 0.5 mL min^{-1}; Gradient elution t = 0 min, MeCN:H_2O with 0.1 % TFA, 5:95; t = 5.0 min, 95:5; t = 5.4 min, 95:5; Rt 1.4 min, *m/z* (ES) found $[M+H]^+$ 912.4229; $C_{38}H_{62}N_{11}O_{13}S$ requires 912.4244.



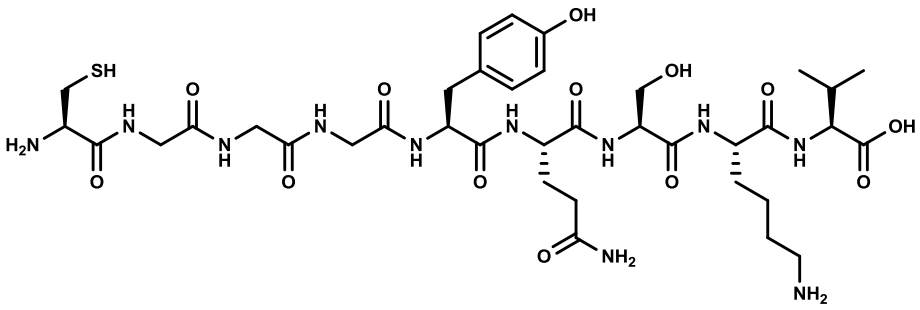
6.18.7 H_2N-CGGGYQSEL-CO_2H



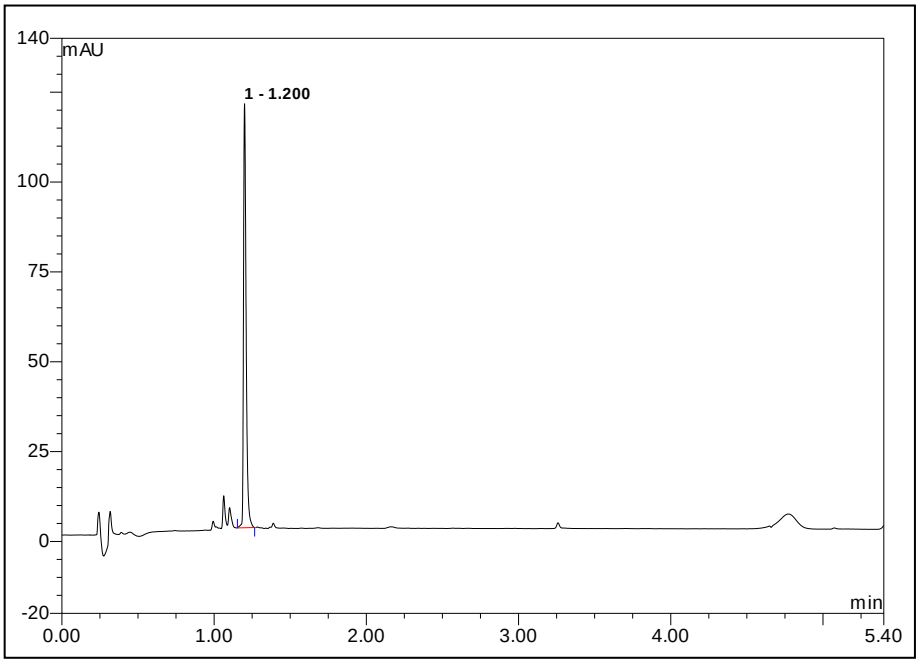
Ascentis Express (C18, 50 x 2.1 mm, 2.7 µm particle size); Flow rate 0.5 mL min^{-1}; Gradient elution t = 0 min, MeCN:H_2O with 0.1 % TFA, 5:95; t = 5.0 min, 95:5; t = 5.4 min, 95:5; Rt 1.6 min, *m/z* (ES) found $[M+H]^+$ 913.3700; $C_{37}H_{57}N_{10}O_{15}S$ requires 913.3720.



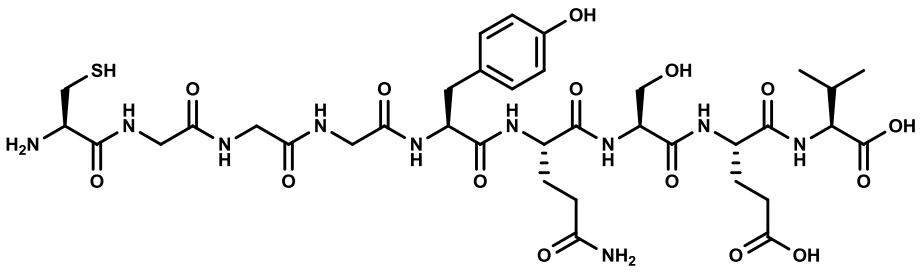
6.18.8 H_2N-CGGGYQSKV-CO_2H



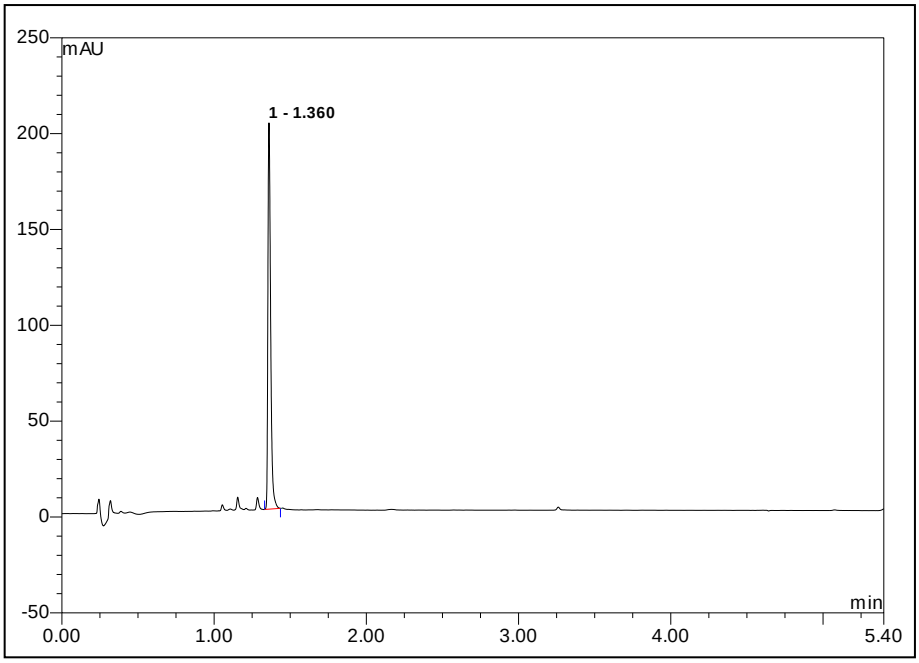
Ascentis Express (C18, 50 x 2.1 mm, 2.7 µm particle size); Flow rate 0.5 mL min^{-1}; Gradient elution t = 0 min, MeCN:H_2O with 0.1 % TFA, 5:95; t = 5.0 min, 95:5; t = 5.4 min, 95:5; Rt 1.2 min, *m/z* (ES) found $[M+H]^+$ 898.4088; $C_{37}H_{60}N_{11}O_{13}S$ requires 898.4087.



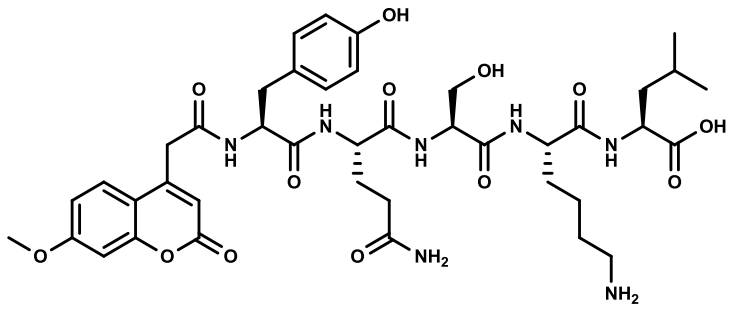
6.18.9 H_2N-CGGGYQSEV-CO_2H



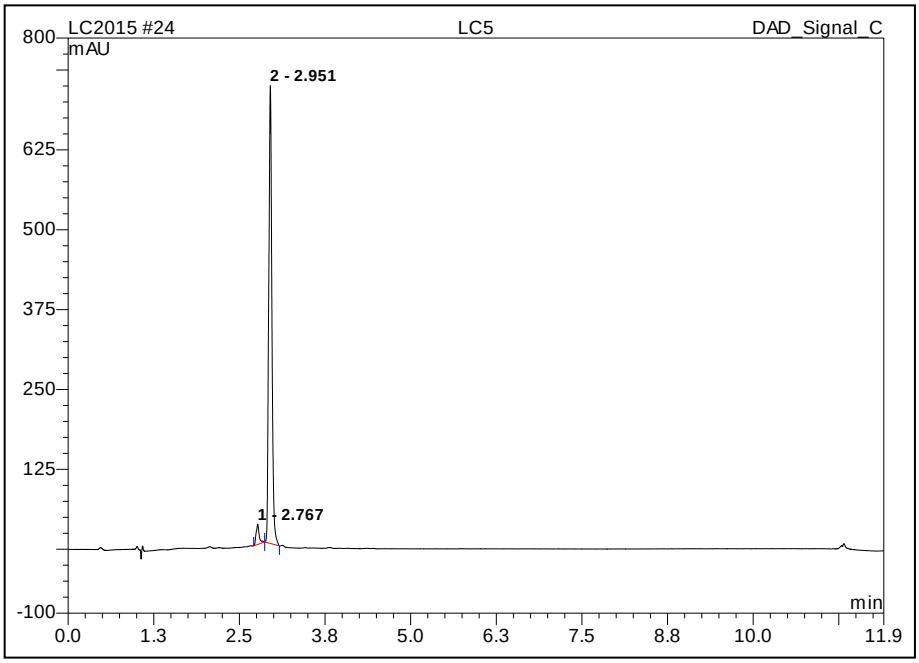
Ascentis Express (C18, 50 x 2.1 mm, 2.7 µm particle size); Flow rate 0.5 mL min^{-1}; Gradient elution t = 0 min, MeCN:H_2O with 0.1 % TFA, 5:95; t = 5.0 min, 95:5; t = 5.4 min, 95:5; Rt 1.4 min, *m/z* (ES) found $[M+H]^+$ 899.3538; $C_{36}H_{55}N_{10}O_{15}S$ requires 899.3564.



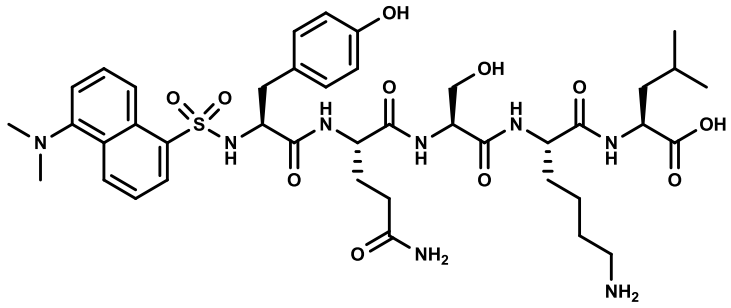
6.18.10 Coumarin-YQSKL-CO_2H



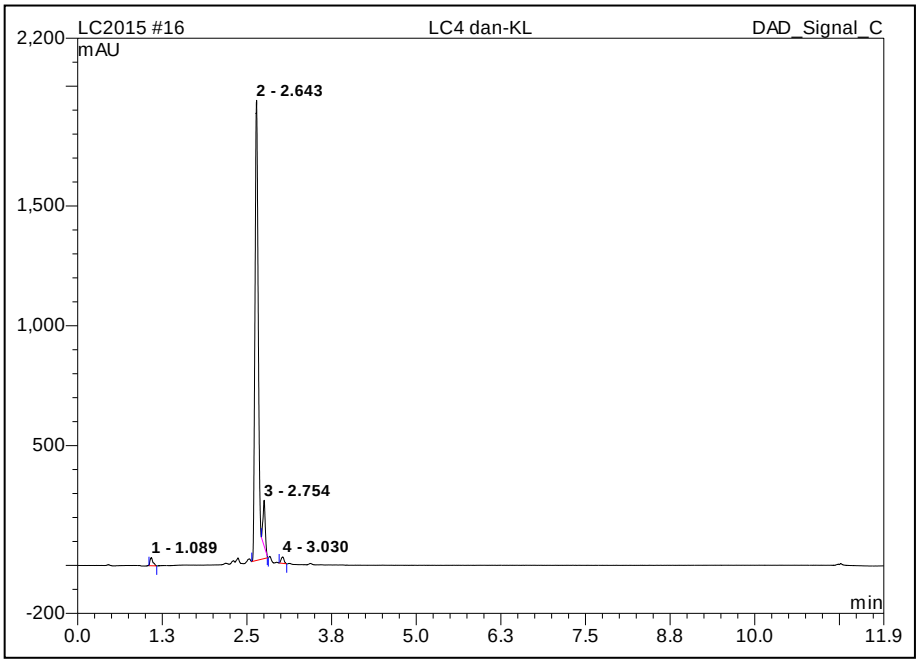
Ascentis Peptide (ES-C18, 100 x 2.1 mm, 2.7 µm particle size); Flow rate 0.5 mL min^{-1}; Gradient elution t = 0 min, MeCN:H_2O with 0.1 % TFA, 5:95; t = 10.0 min, 95:5; t = 10.5 min, 5:95; t = 12.0 min, 5:95; Rt 3.0 min, *m/z* (ES) found $[M+H]^+$ 854.3925; $C_{41}H_{56}N_7O_{13}$ requires 854.3937.



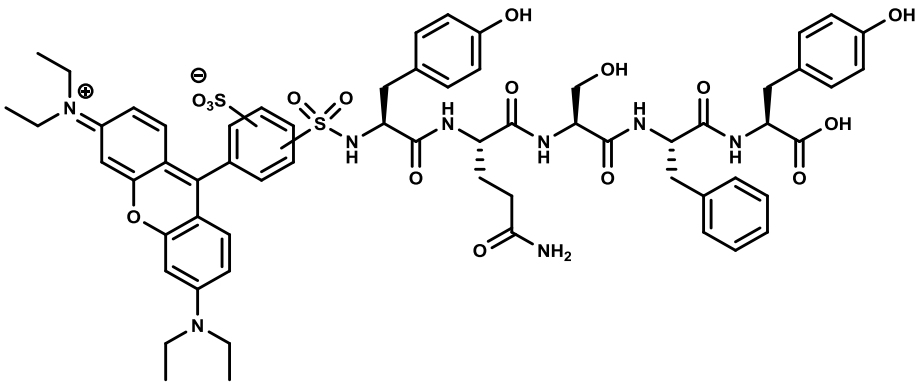
6.18.11 Dansyl-YQSKL-CO_2H



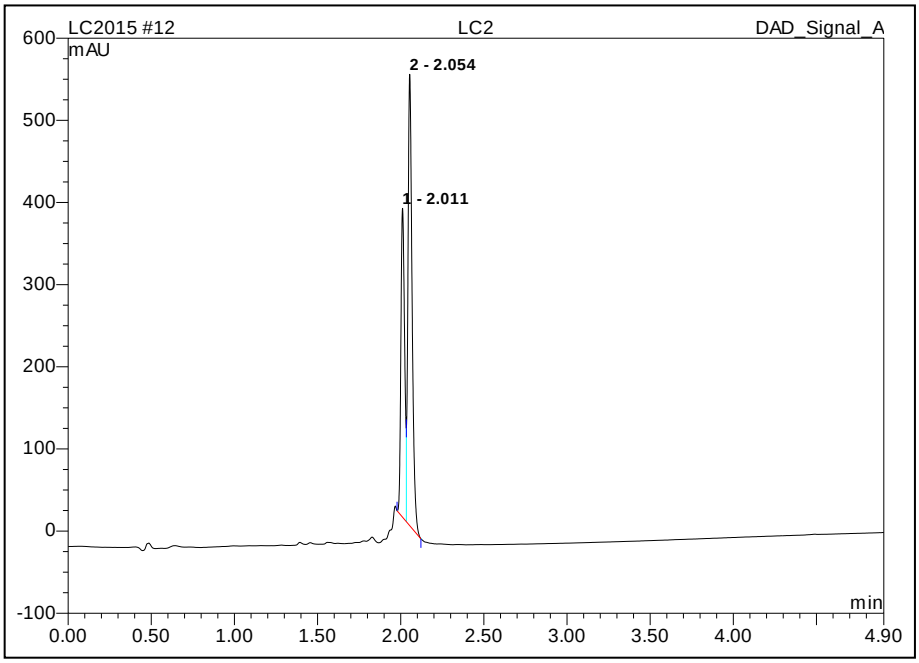
Ascentis Peptide (ES-C18, 100 x 2.1 mm, 2.7 µm particle size); Flow rate 0.5 mL min^{-1}; Gradient elution t = 0 min, MeCN:H_2O with 0.1 % TFA, 5:95; t = 10.0 min, 95:5; t = 10.5 min, 5:95; t = 12.0 min, 5:95; Rt 2.6 min, *m/z* (ES) found $[M+H]^+$ 871.4023; $C_{41}H_{59}N_8O_{11}S$ requires 871.4019.



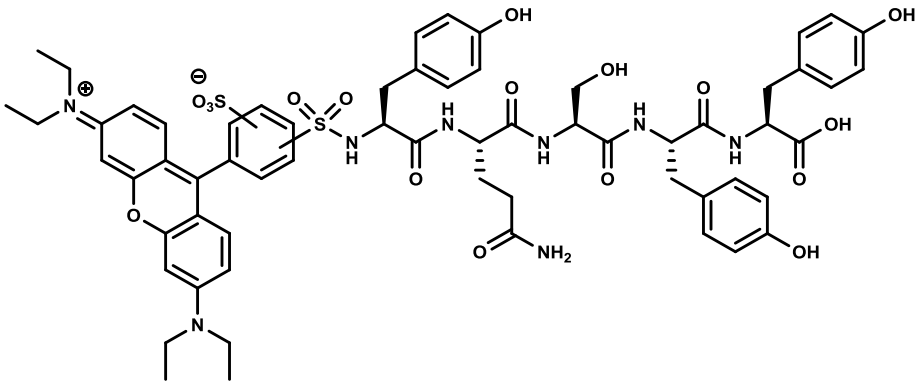
6.18.12 Lissamine-YQSFY-CO_2H



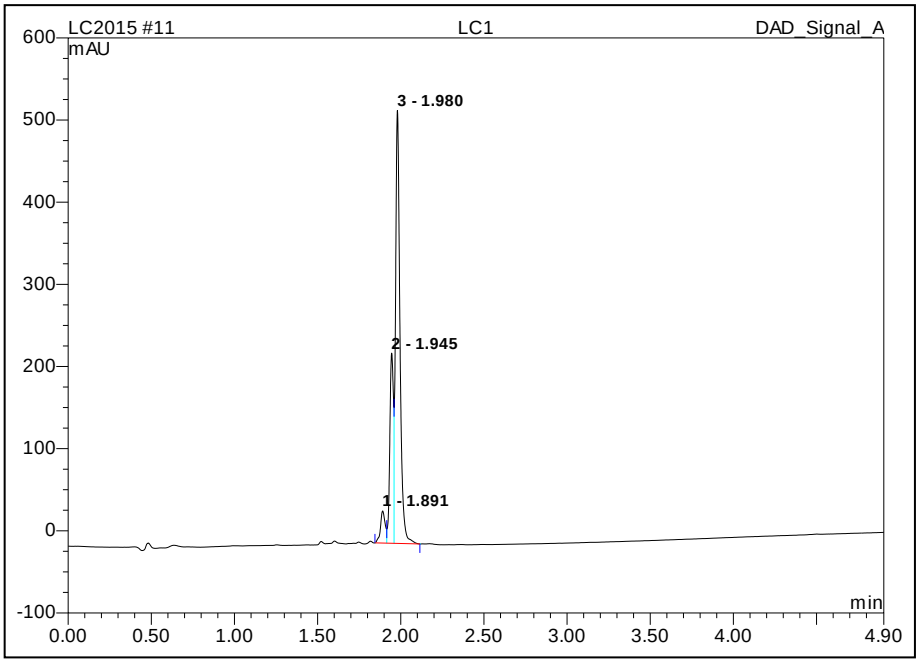
Ascentis Express (C18, 50 x 2.1 mm, 2.7 µm particle size); Flow rate 0.5 mL min^{-1}; Gradient elution t = 0 min, MeCN:H_2O with 0.1 % TFA, 5:95; t = 5.0 min, 95:5; Rt 2.0 and 2.1 min (mixed isomers of Lissamine Rhodamine B sulfonyl chloride used), *m/z* (ES) found $[M+H]^+$ 1247.4433; $C_{62}H_{71}N_8O_{16}S_2$ requires 1247.4424.



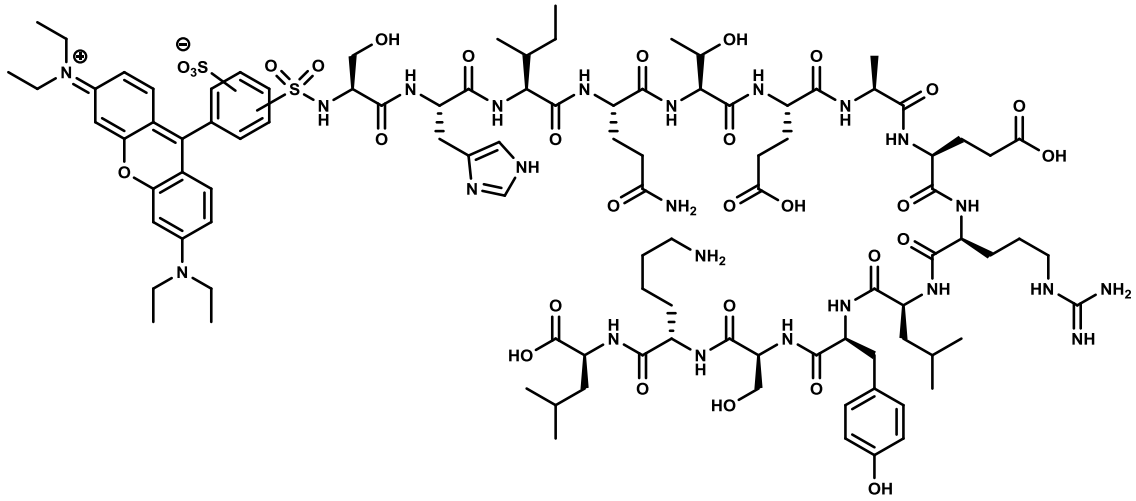
6.18.13 Lissamine-YQSYY-CO_2H



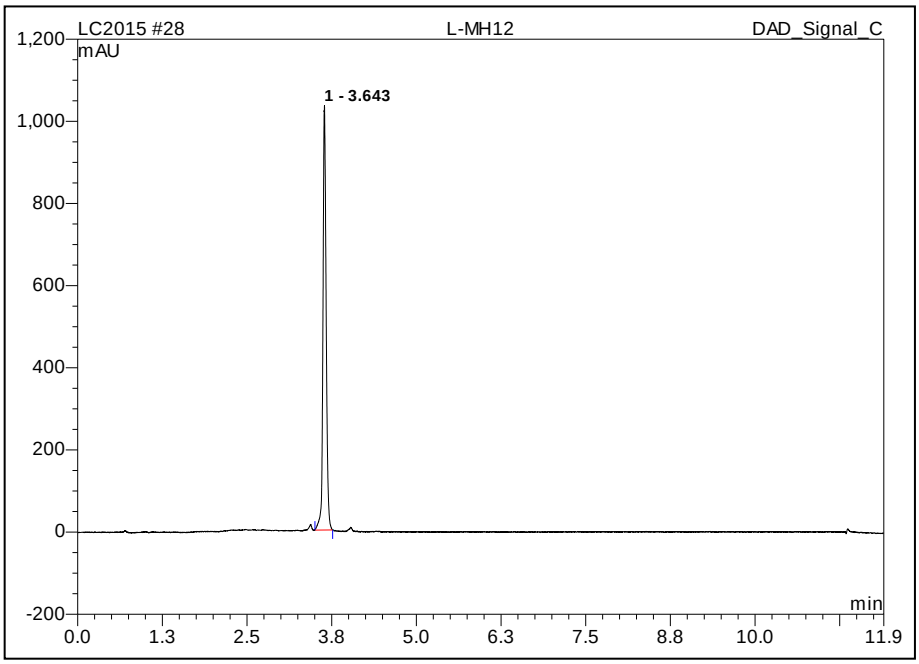
Ascentis Express (C18, 50 x 2.1 mm, 2.7 µm particle size); Flow rate 0.5 mL min^{-1}; Gradient elution t = 0 min, MeCN:H_2O with 0.1 % TFA, 5:95; t = 5.0 min, 95:5; Rt 1.9 and 2.0 min (mixed isomers of Lissamine Rhodamine B sulfonyl chloride used), *m/z* (ES) found $[M+H]^+$ 1263.4381; $C_{62}H_{71}N_8O_{17}S_2$ requires 1264.4373.



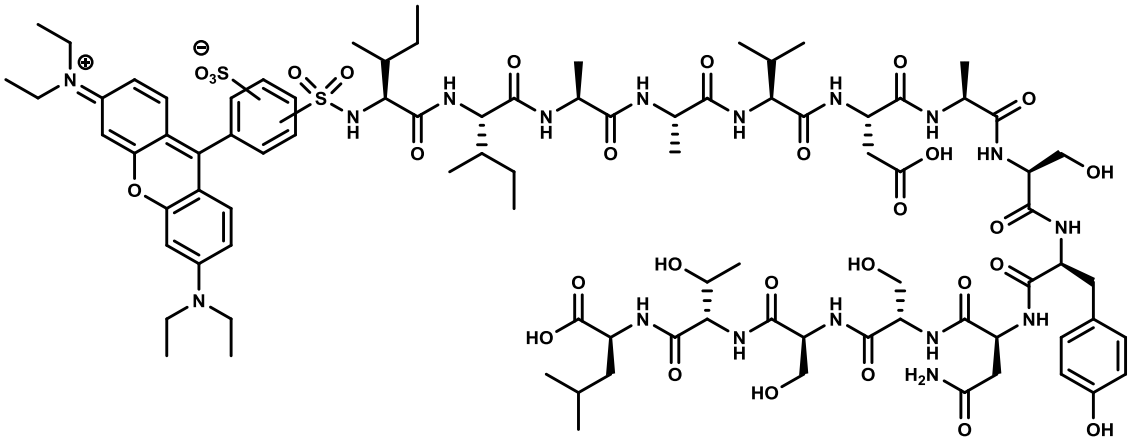
6.18.14 Lissamine-SHIQTEAERLYSKL-CO_2H



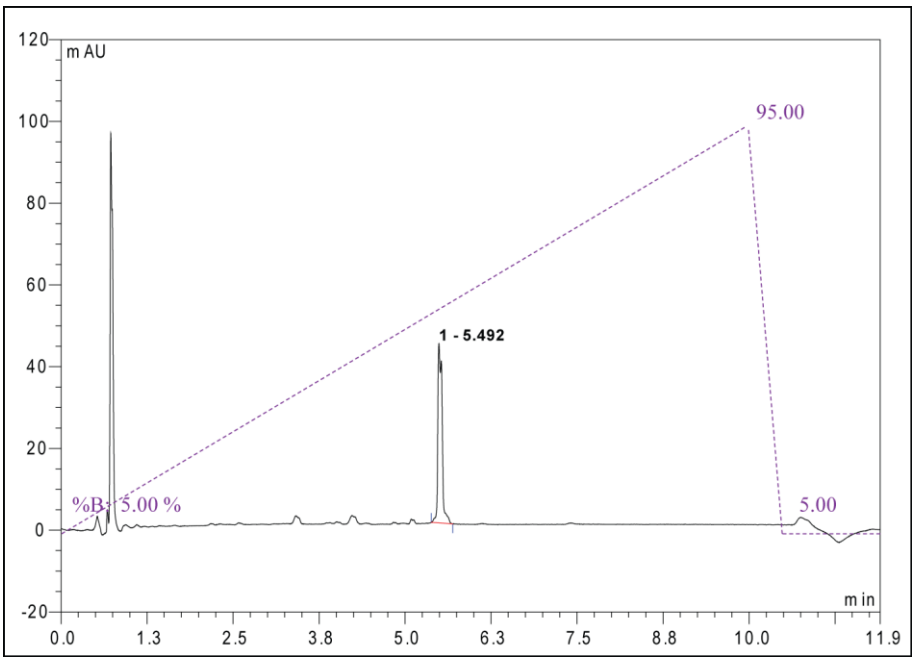
Ascentis Peptide (ES-C18, 100 x 2.1 mm, 2.7 µm particle size); Flow rate 0.5 mL min^{-1}; Gradient elution t = 0 min, MeCN:H_2O with 0.1 % TFA, 5:95; t = 10.0 min, 95:5; t = 10.5 min, 5:95; t = 12.0 min, 5:95; Rt 3.6 min, *m/z* (ES) found $[M+2H]^{2+}$ 1108.5170; $C_{100}H_{148}N_{23}O_{30}S_2$ requires 1108.5181.



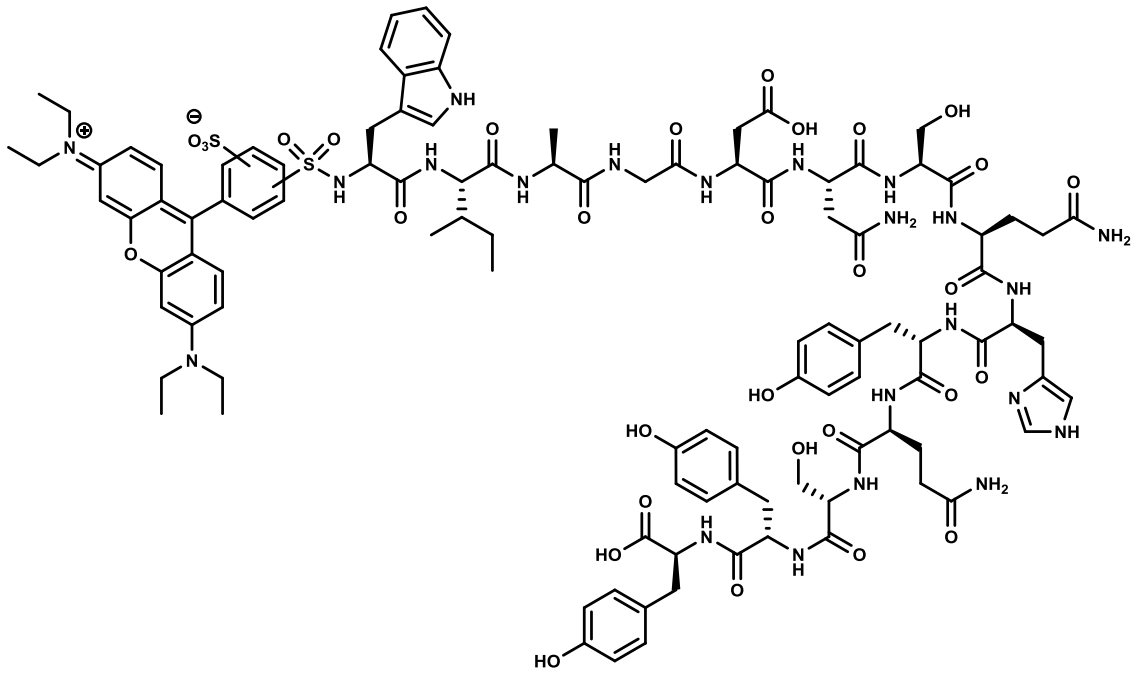
6.18.15 Lissamine-IIAAVDASYNSSTL-CO_2H



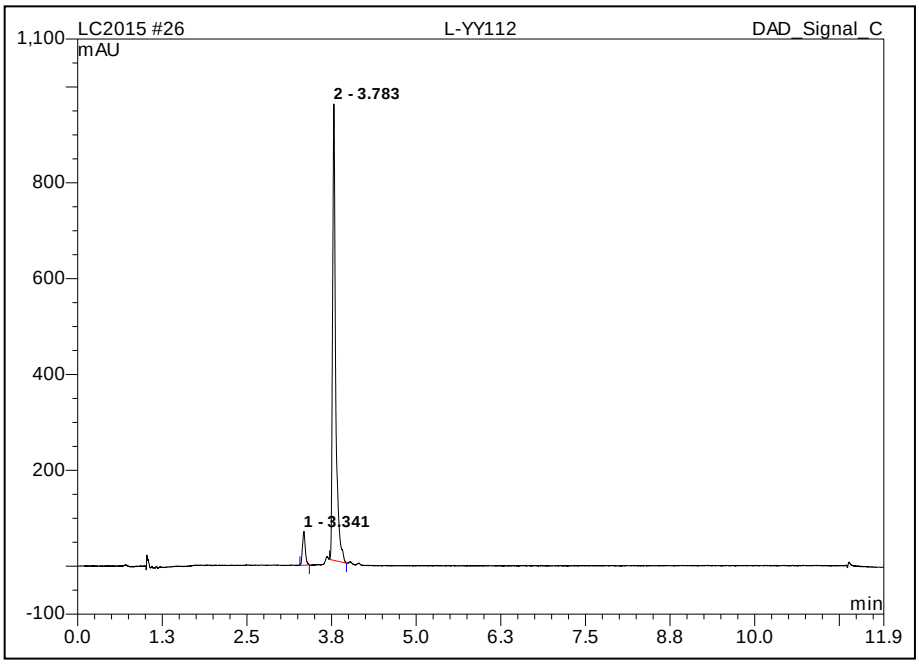
Ascentis Peptide (ES-C18, 100 x 2.1 mm, 2.7 µm particle size); Flow rate 0.5 mL min^{-1}; Gradient elution t = 0 min, MeCN:H_2O with 0.1 % TFA, 5:95; t = 10.0 min, 95:5; t = 10.5 min, 5:95; t = 12.0 min, 5:95; Rt 5.5 min, *m/z* (ES) found $[M+2H]^{2+}$ 983.4400; $C_{89}H_{130}N_{17}O_{29}S_2$ requires 983.4410.



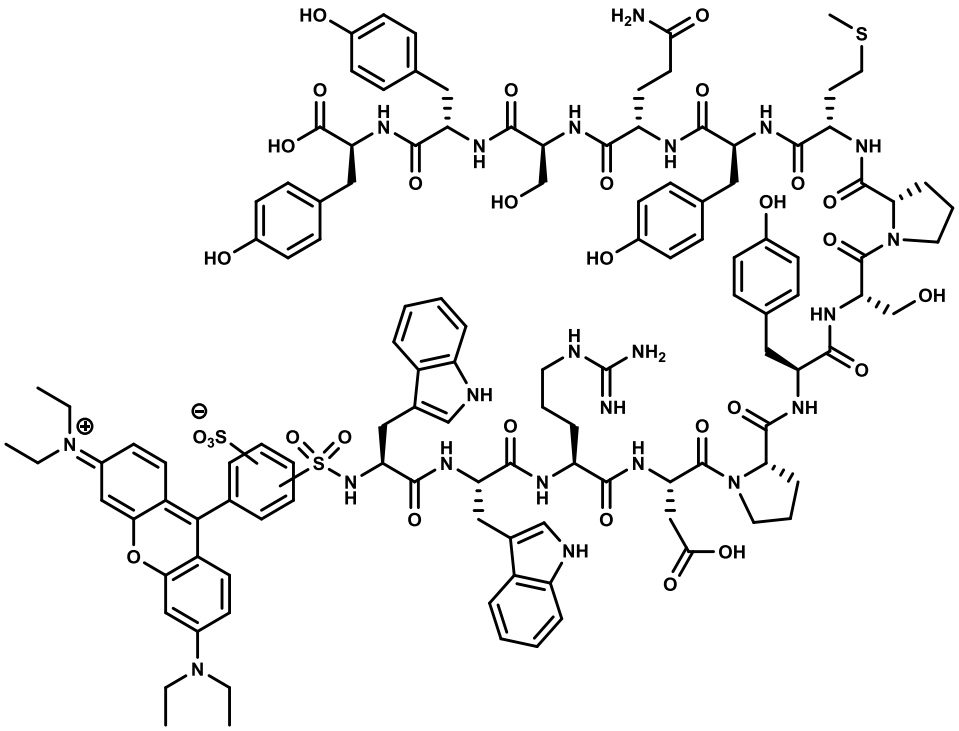
6.18.16 Lissamine-WIAGDNSQHYQSYY-CO_2H



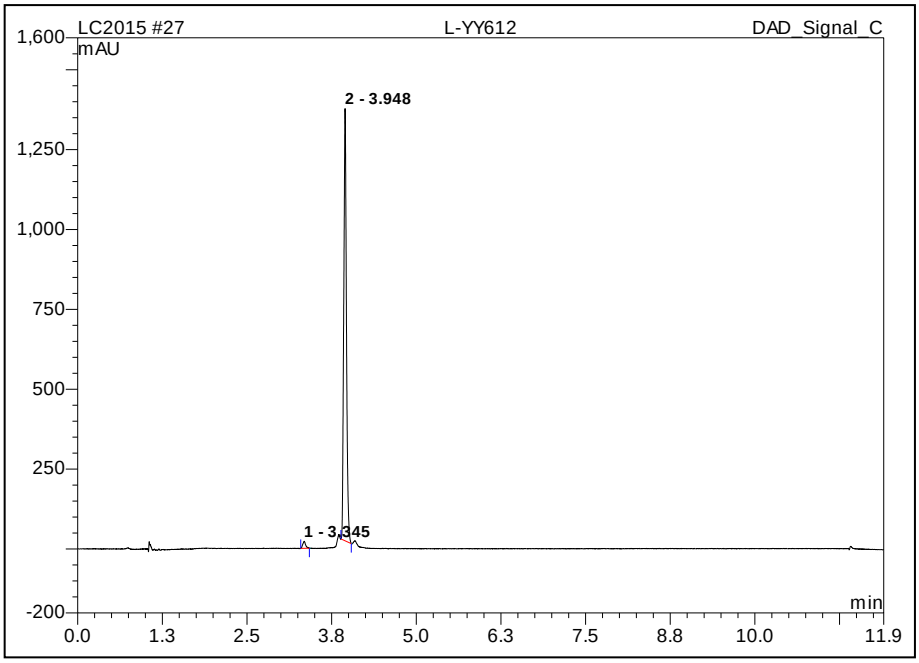
Ascentis Peptide (ES-C18, 100 x 2.1 mm, 2.7 µm particle size); Flow rate 0.5 mL min^{-1}; Gradient elution t = 0 min, MeCN:H_2O with 0.1 % TFA, 5:95; t = 10.0 min, 95:5; t = 10.5 min, 5:95; t = 12.0 min, 5:95; Rt 3.8 min, *m/z* (ES) found $[M+2H]^{2+}$ 1136.9460; $C_{106}H_{131}N_{22}O_{31}S_2$ requires 1136.9475.



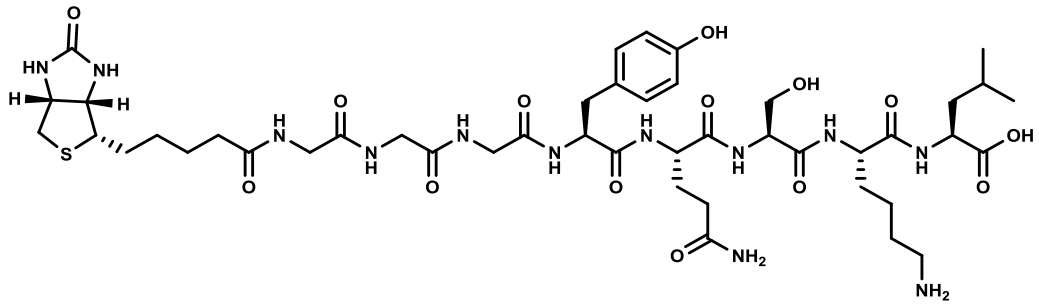
6.18.17 Lissamine-WWRDPYSPMYQSYY-CO_2H



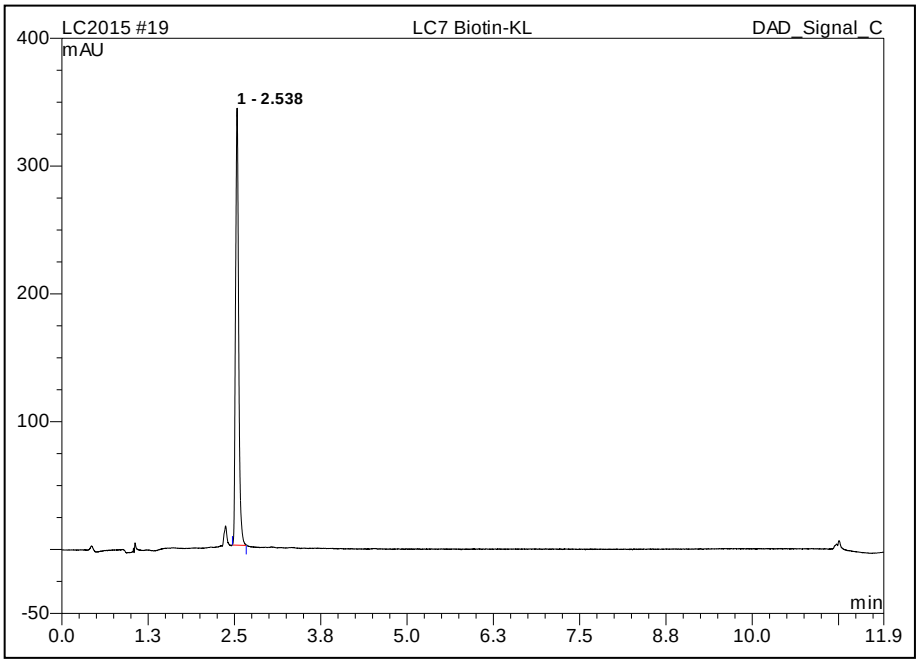
Ascentis Peptide (ES-C18, 100 x 2.1 mm, 2.7 µm particle size); Flow rate 0.5 mL min^{-1}; Gradient elution t = 0 min, MeCN:H_2O with 0.1 % TFA, 5:95; t = 10.0 min, 95:5; t = 10.5 min, 5:95; t = 12.0 min, 5:95; Rt 3.9 min, *m/z* (ES) found $[M+2H]^{2+}$ 1241.9890; $C_{121}H_{145}N_{22}O_{32}S_2$ requires 1241.9908.



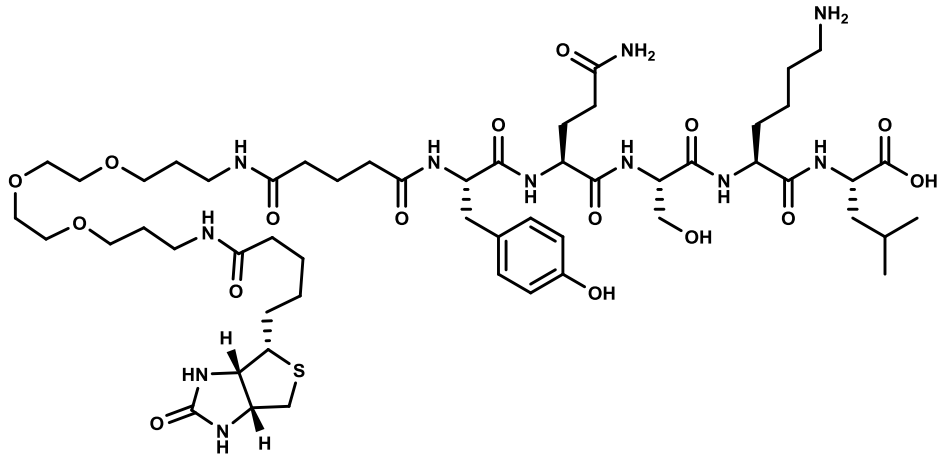
6.18.18 Biotin-GGGYQSKL-CO_2H



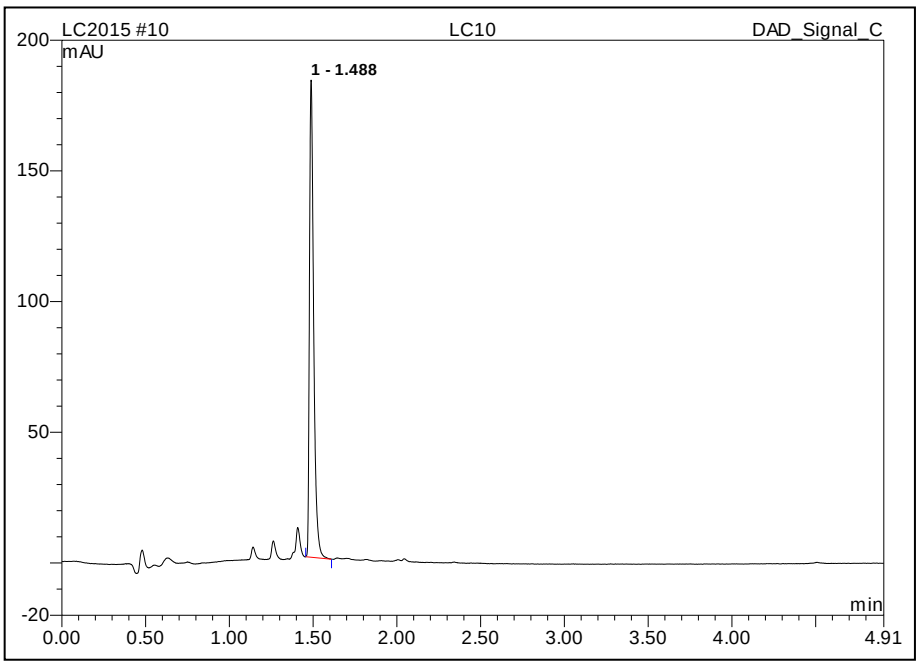
Ascentis Peptide (ES-C18, 100 x 2.1 mm, 2.7 µm particle size); Flow rate 0.5 mL min^{-1}; Gradient elution t = 0 min, MeCN:H_2O with 0.1 % TFA, 5:95; t = 10.0 min, 95:5; t = 10.5 min, 5:95; t = 12.0 min, 5:95; Rt 2.5 min, *m/z* (ES) found $[M+H]^+$ 1035.4936; $C_{45}H_{71}N_{12}O_{14}S$ requires 1035.4928.



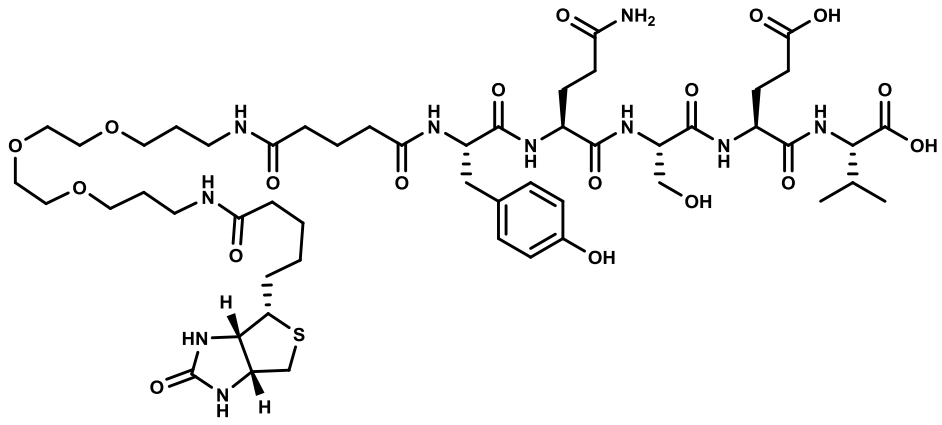
6.18.19 Biotin-(PEG)$_2$-YQSKL-CO$_2$H



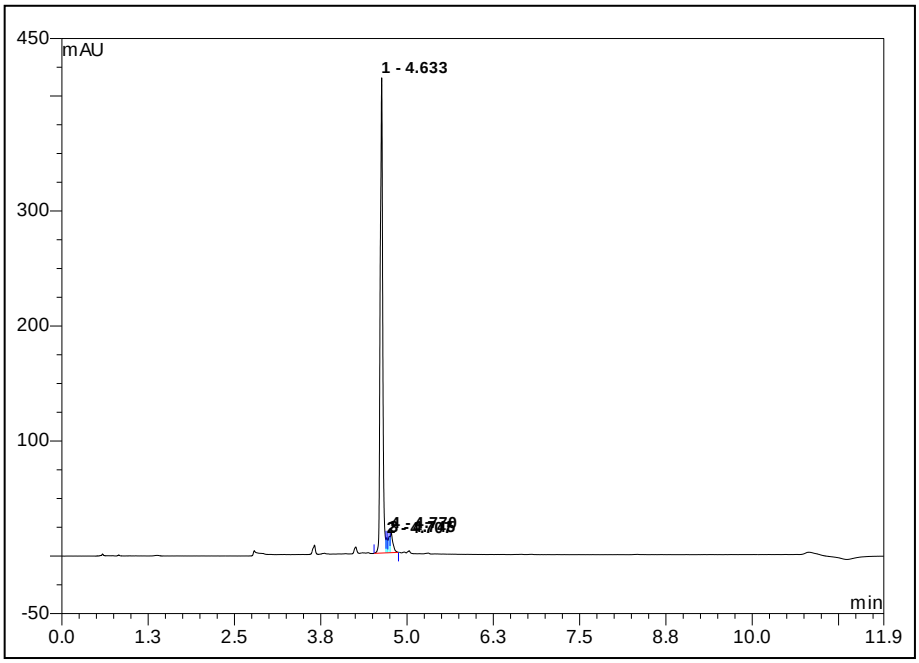
Ascentis Express (C18, 50 x 2.1 mm, 2.7 µm particle size); Flow rate 0.5 mL min^{-1}; Gradient elution t = 0 min, MeCN:H_2O with 0.1 % TFA, 5:95; t = 5.0 min, 95:5; Rt 1.5 min, *m/z* (ES) found $[M+H]^+$ 1180.6291; $C_{54}H_{90}N_{11}O_{16}S$ requires 1180.6243.



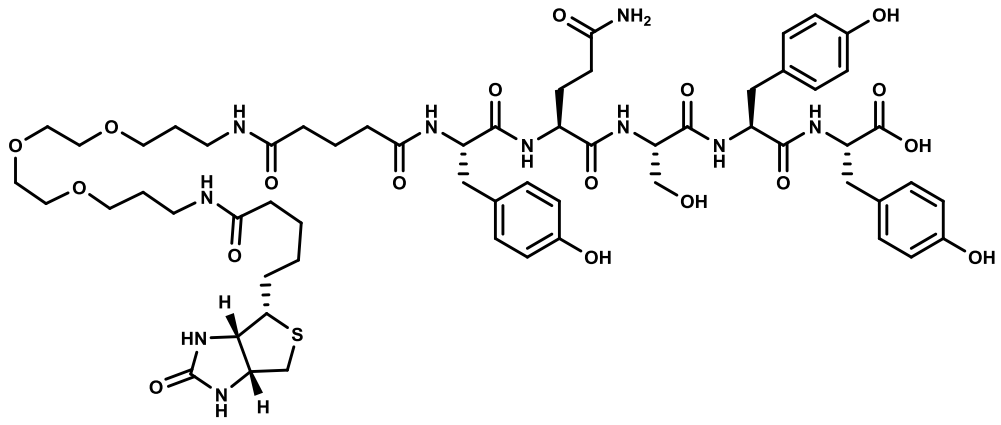
6.18.20 Biotin-(PEG)$_2$-YQSEV-CO$_2$H



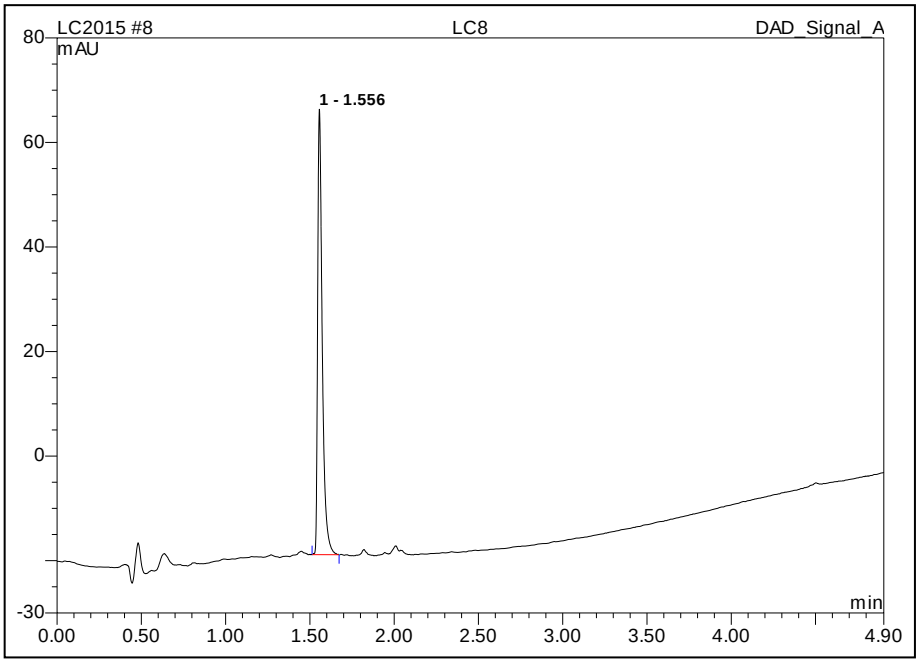
Ascentis Peptide (ES-C18, 100 x 2.1 mm, 2.7 µm particle size); Flow rate 0.5 mL min^{-1}; Gradient elution t = 0 min, MeCN:H_2O with 0.1 % TFA, 5:95; t = 10.0 min, 95:5; t = 10.5 min, 5:95; t = 12.0 min, 5:95; Rt 4.6 min, *m/z* (ES) found $[M+H]^+$ 1167.5609; $C_{52}H_{83}N_{10}O_{18}S$ requires 1167.5602.



6.18.21 Biotin-(PEG)$_2$-YQSYY-CO$_2$H



Ascentis Express (C18, 50 x 2.1 mm, 2.7 µm particle size); Flow rate 0.5 mL min^{-1}; Gradient elution t = 0 min, MeCN:H_2O with 0.1 % TFA, 5:95; t = 5.0 min, 95:5; Rt 1.6 min, *m/z* (ES) found $[M+H]^+$ 1265.5779; $C_{60}H_{85}N_{10}O_{18}S$ requires 1265.5759.



References

Agne, B., Meindl, N. M., Niederhoff, K., Einwachter, H., Rehling, P., Sickmann, A., Meyer, H. E., Girzalsky, W. and Kunau, W. H. 2003. Pex8p: an intraperoxisomal organizer of the peroxisomal import machinery. *Mol Cell,* 11, 635-46.

Agrawal, G. and Subramani, S. 2016. De novo peroxisome biogenesis: Evolving concepts and conundrums. *Biochim Biophys Acta,* 1863, 892-901.

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

Allan, R. K. and Ratajczak, T. 2011. Versatile TPR domains accommodate different modes of target protein recognition and function. *Cell Stress Chaperon,* 16, 353-67.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215, 403-10.

Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N. and Yeh, L. S. 2004. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res,* 32, D115-9.

Bain, J. D., Diala, E. S., Glabe, C. G., Dix, T. A. and Chamberlin, A. R. 1989. Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide. *J Am Chem Soc,* 111, 8013-4.

Baker, A., Charlton, W., Johnson, B., Lopez-Huertas, E., Oh, J., Sparkes, I. and Thomas, J. 2000. Biochemical and molecular approaches to understanding protein import into peroxisomes. *Biochem Soc Trans,* 28, 499-504.

Baker, A. and Paudyal, R. 2014. The life of the peroxisome: from birth to death. *Curr Opin Plant Biol,* 22, 39-47.

Baker, A., Lanyon-Hogg, T. and Warriner, S. L. 2016. Peroxisome protein import: a complex journey. *Biochem Soc Trans,* 44, 783-9.

Bankston, J. R., Camp, S. S., DiMaio, F., Lewis, A. S., Chetkovich, D. M. and Zagotta, W. N. 2012. Structure and stoichiometry of an accessory subunit TRIP8b interaction with hyperpolarization-activated cyclic nucleotide-gated channels. *PNAS,* 109, 7899-904.

Baudhuin, P., Beaufay, H. and De Duve, C. 1965. Combined biochemical and morphological study of particulate fractions from rat liver. Analysis of preparations enriched in lysosomes or in particles containing urate oxidase, D-amino acid oxidase, and catalase. *J Cell Biol*, 26, 219-43.

Beevers, H. 1979. Microbodies in higher plants. *Annu Rev Plant Physiol,* 30, 159-93.

Bernhard, W. and Rouiller, C. 1956. Microbodies and the problem of mitochondrial regeneration in liver cells. *J Biophys Biochem Cytol,* 2, 355-60.

Bhogal, M. S., Lanyon-Hogg, T., Johnston, K. A., Warriner, S. L. and Baker, A. 2016. Covalent label transfer between peroxisomal importomer components reveals export-driven import interactions. *J Biol Chem,* 291, 2460-8.

Braverman, N., Dodt, G., Gould, S. J. and Valle, D. 1998. An isoform of Pex5p, the human PTS1 receptor, is required for the import of PTS2 proteins into peroxisomes. *Hum Mol Gen,* 7, 1195-205.

Brickner, D. G., Brickner, J. H. and Olsen, L. J. 1998. Sequence analysis of a cDNA encoding Pex5p, a peroxisomal targeting signal type 1 receptor from *Arabidopsis thaliana*. *Plant Physiol,* 118, 330.

Brocard, C. and Hartig, A. 2006. Peroxisome targeting signal 1: Is it really a simple tripeptide? *BBA - Mol Cell Res,* 1763, 1565-73.

Brown, A. I., Kim, P. K. and Rutenberg, A. D. 2014. PEX5 and ubiquitin dynamics on mammalian peroxisome membranes. *PLoS Comput Biol,* 10, e1003426.
Burkhart, S. E., Lingard, M. J. and Bartel, B. 2013. Genetic dissection of peroxisome-associated matrix protein degradation in *Arabidopsis thaliana*. *Genetics,* 193, 125-41.
Cantor, C. R. and Schimmel, P. R. 1980. *Biophysical chemistry. Pt. 1,* New York, W.H. Freeman and Company.
Carvalho, A. F., Costa-Rodrigues, J., Correia, I., Costa Pessoa, J., Faria, T. Q., Martins, C. L., Fransen, M., Sa-Miranda, C. and Azevedo, J. E. 2006. The N-terminal half of the peroxisomal cycling receptor Pex5p is a natively unfolded domain. *J Mol Biol,* 356, 864-75.
Carvalho, A. F., Pinto, M. P., Grou, C. P., Alencastre, I. S., Fransen, M., Sa-Miranda, C. and Azevedo, J. E. 2007a. Ubiquitination of mammalian Pex5p, the peroxisomal import receptor. *J Biol Chem,* 282, 31267-72.
Carvalho, A. F., Grou, C. P., Pinto, M. P., Alencastre, I. S., Costa-Rodrigues, J., Fransen, M., Sá-Miranda, C. and Azevedo, J. E. 2007b. Functional characterization of two missense mutations in Pex5p—C11S and N526K. *BBA - Mol Cell Res,* 1773, 1141-8.
Chowdhary, G., Kataya, A., Lingner, T. and Reumann, S. 2012. Non-canonical peroxisome targeting signals: identification of novel PTS1 tripeptides and characterization of enhancer elements by computational permutation analysis. *BMC Plant Biol,* 12, 142.
Ciniawsky, S., Grimm, I., Saffian, D., Girzalsky, W., Erdmann, R. and Wendler, P. 2015. Molecular snapshots of the Pex1/6 AAA+ complex in action. *Nat Commun,* 6, 7331.
Cross, L. L., Ebeed, H. T. and Baker, A. 2016. Peroxisome biogenesis, protein targeting mechanisms and PEX gene functions in plants. *Biochim Biophys Acta,* 1863, 850-62.
Cui, S., Fukao, Y., Mano, S., Yamada, K., Hayashi, M. and Nishimura, M. 2013. Proteomic analysis reveals that the Rab GTPase RabE1c is involved in the degradation of the peroxisomal protein receptor PEX7 (peroxin 7). *J Biol Chem,* 288, 6014-23.
D'Andrea, L. D. and Regan, L. 2003. TPR proteins: the versatile helix. *Trends Biochem Sci,* 28, 655-62.
de Duve, C. and Baudhuin, P. 1966. Peroxisomes (microbodies and related particles). *Physiol Rev,* 46, 323-57.
Decker, E. L. and Reski, R. 2004. The moss bioreactor. *Curr Opin Plant Biol,* 7, 166-70.
DeLoache, W. C., Russ, Z. N. and Dueber, J. E. 2016. Towards repurposing the yeast peroxisome for compartmentalizing heterologous metabolic pathways. *Nat Commun,* 7, 11152.
Deshaies, R. J. and Joazeiro, C. A. 2009. RING domain E3 ubiquitin ligases. *Annu Rev Biochem,* 78, 399-434.
Distel, B., Gould, S. J., Voorn-Brouwer, T., van der Berg, M., Tabak, H. F. and Subramani, S. 1992. The carboxyl-terminal tripeptide serine-lysine-leucine of firefly luciferase is necessary but not sufficient for peroxisomal import in yeast. *New Biol,* 4, 157-65.
Dodt, G., Braverman, N., Wong, C., Moser, A., Moser, H. W., Watkins, P., Valle, D. and Gould, S. J. 1995. Mutations in the PTS1 receptor gene, PXR1, define complementation group 2 of the peroxisome biogenesis disorders. *Nat Genet,* 9, 115-25.
Ebberink, M. S., Mooyer, P. A., Koster, J., Dekker, C. J., Eyskens, F. J., Dionisi-Vici, C., Clayton, P. T., Barth, P. G., Wanders, R. J. and Waterham, H. R. 2009. Genotype-phenotype correlation in PEX5-deficient peroxisome biogenesis defective cell lines. *Hum Mutat,* 30, 93-8.

El Magraoui, F., Schrotter, A., Brinkmeier, R., Kunst, L., Mastalski, T., Muller, T., Marcus, K., Meyer, H. E., Girzalsky, W., Erdmann, R. and Platta, H. W. 2014. The cytosolic domain of Pex22p stimulates the Pex4p-dependent ubiquitination of the PTS1-receptor. *PLoS ONE,* 9, e105894.

Fan, J., Quan, S., Orth, T., Awai, C., Chory, J. and Hu, J. 2005. The Arabidopsis PEX12 gene is required for peroxisome biogenesis and is essential for development. *Plant Physiol,* 139, 231-9.

Farmer, L. M., Rinaldi, M. A., Young, P. G., Danan, C. H., Burkhart, S. E. and Bartel, B. 2013. Disrupting autophagy restores peroxisome function to an Arabidopsis lon2 mutant and reveals a role for the LON2 protease in peroxisomal matrix protein degradation. *Plant Cell,* 25, 4085-100.

Firth, A. E. and Patrick, W. M. 2008. GLUE-IT and PEDEL-AA: new programmes for analyzing protein diversity in randomized libraries. *Nucleic Acids Res,* 36, W281-5.

Fodor, K., Wolf, J., Erdmann, R., Schliebs, W. and 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.

Fodor, K., Wolf, J., Reglinski, K., Passon, D. M., Lou, Y., Schliebs, W., Erdmann, R. and Wilmanns, M. 2015. Ligand-induced compaction of the PEX5 receptor-binding cavity impacts protein import efficiency into peroxisomes. *Traffic,* 16, 85-98.

Francisco, T., Rodrigues, T. A., Pinto, M. P., Carvalho, A. F., Azevedo, J. E. and Grou, C. P. 2014. Ubiquitin in the peroxisomal protein import pathway. *Biochimie,* 98, 29-35.

Fransen, M., Terlecky, S. R. and Subramani, S. 1998. Identification of a human PTS1 receptor docking protein directly required for peroxisomal protein import. *PNAS,* 95, 8087-92.

Freitas, M. O., Francisco, T., Rodrigues, T. A., Alencastre, I. S., Pinto, M. P., Grou, C. P., Carvalho, A. F., Fransen, M., Sa-Miranda, C. and Azevedo, J. E. 2011. PEX5 protein binds monomeric catalase blocking its tetramerization and releases it upon binding the N-terminal domain of PEX14. *J Biol Chem,* 286, 40509-19.

Gatto, G. J., Geisbrecht, B. V., Gould, S. J. and Berg, J. M. 2000. Peroxisomal targeting signal-1 recognition by the TPR domains of human PEX5. *Nat Struct Biol,* 7, 1091-5.

Gatto, G. J., Maynard, E. L., Guerrerio, A. L., Geisbrecht, B. V., Gould, S. J. and Berg, J. M. 2003. Correlating structure and affinity for PEX5:PTS1 complexes. *Biochemistry,* 42, 1660-6.

Ghosh, D. and Berg, J. M. 2010. A proteome-wide perspective on Peroxisome Targeting Signal 1(PTS1)-Pex5p affinities. *J Am Chem Soc,* 132, 3973-9.

Glover, J. R., Andrews, D. W. and Rachubinski, R. A. 1994. *Saccharomyces cerevisiae* peroxisomal thiolase is imported as a dimer. *PNAS,* 91, 10541-5.

Gonzalez, N. H., Felsner, G., Schramm, F. D., Klingl, A., Maier, U.-G. and Bolte, K. 2011. A single peroxisomal targeting signal mediates matrix protein import in diatoms. *PLoS ONE,* 6, e25316.

Goto-Yamada, S., Mano, S., Oikawa, K., Shibata, M. and Nishimura, M. 2014. Interaction between chaperone and protease functions of LON2, and autophagy during the functional transition of peroxisomes. *Plant Signal Behav,* 9, e28838.

Goto-Yamada, S., Mano, S., Yamada, K., Oikawa, K., Hosokawa, Y., Hara-Nishimura, I. and Nishimura, M. 2015. Dynamics of the light-dependent transition of plant peroxisomes. *Plant Cell Physiol,* 56, 1264-71.

Goto, S., Mano, S., Nakamori, C. and Nishimura, M. 2011. Arabidopsis ABERRANT PEROXISOME MORPHOLOGY9 is a peroxin that recruits the PEX1-PEX6 complex to peroxisomes. *Plant Cell,* 23, 1573-87.

Gould, S. G., Keller, G. A. and Subramani, S. 1987. Identification of a peroxisomal targeting signal at the carboxy terminus of firefly luciferase. *J Cell Biol,* 105, 2923-31.
Gould, S. J., Keller, G. A. and Subramani, S. 1988. Identification of peroxisomal targeting signals located at the carboxy terminus of four peroxisomal proteins. *J Cell Biol,* 107, 897-905.
Gould, S. J., Keller, G. A., Hosken, N., Wilkinson, J. and Subramani, S. 1989. A conserved tripeptide sorts proteins to peroxisomes. *J Cell Biol,* 108, 1657-64.
Gould, S. J., Keller, G. A., Schneider, M., Howell, S. H., Garrard, L. J., Goodman, J. M., Distel, B., Tabak, H. and Subramani, S. 1990. Peroxisomal protein import is conserved between yeast, plants, insects and mammals. *EMBO J,* 9, 85-90.
Gouveia, A. M., Reguenga, C., Oliveira, M. E., Sa-Miranda, C. and Azevedo, J. E. 2000. Characterization of peroxisomal Pex5p from rat liver. Pex5p in the Pex5p-Pex14p membrane complex is a transmembrane protein. *J Biol Chem,* 275, 32444-51.
Gouveia, A. M., Guimaraes, C. P., Oliveira, M. E., Sa-Miranda, C. and Azevedo, J. E. 2003. Insertion of Pex5p into the peroxisomal membrane is cargo protein-dependent. *J Biol Chem,* 278, 4389-92.
Grimm, I., Erdmann, R. and Girzalsky, W. 2016. Role of AAA(+)-proteins in peroxisome biogenesis and function. *Biochim Biophys Acta,* 1863, 828-37.
Gunn, S. J. 2008. *Novel ligands of the protein PEX5: towards the study of peroxisomal protein import.* University of Leeds.
Hadden, D. A., Phillipson, B. A., Johnston, K. A., Brown, L. A., Manfield, I. W., El-Shami, M., Sparkes, I. A. and Baker, A. 2006. Arabidopsis PEX19 is a dimeric protein that binds the peroxin PEX10. *Mol Membr Biol,* 23, 325-36.
Hagstrom, D., Ma, C., Guha-Polley, S. and Subramani, S. 2014. The unique degradation pathway of the PTS2 receptor, Pex7, is dependent on the PTS receptor/coreceptor, Pex5 and Pex20. *Mol Biol Cell,* 25, 2634-43.
Harano, T., Nose, S., Uezu, R., Shimizu, N. and Fujiki, Y. 2001. Hsp70 regulates the interaction between the peroxisome targeting signal type 1 (PTS1)-receptor Pex5p and PTS1. *Biochem J,* 357, 157-65.
Harper, C. C., Berg, J. M. and Gould, S. J. 2003. PEX5 binds the PTS1 independently of Hsp70 and the peroxin PEX12. *J Biol Chem,* 278, 7897-901.
Hayashi, M., Aoki, M., Kondo, M. and Nishimura, M. 1997. Changes in targeting efficiencies of proteins to plant microbodies caused by amino acid substitutions in the carboxy-terminal tripeptide. *Plant Cell Physiol,* 38, 759-68.
Hayashi, M., Toriyama, K., Kondo, M., Kato, A., Mano, S., De Bellis, L., Hayashi-Ishimaru, Y., Yamaguchi, K., Hayashi, H. and Nishimura, M. 2000a. Functional transformation of plant peroxisomes. *Cell Biochem Biophys,* 32, 295-304.
Hayashi, M., Nito, K., Toriyama-Kato, K., Kondo, M., Yamaya, T. and Nishimura, M. 2000b. AtPex14p maintains peroxisomal functions by determining protein targeting to three kinds of plant peroxisomes. *EMBO J,* 19, 5701-10.
Hayashi, M., Yagi, M., Nito, K., Kamada, T. and Nishimura, M. 2005. Differential contribution of two peroxisomal protein receptors to the maintenance of peroxisomal functions in Arabidopsis. *J Biol Chem,* 280, 14829-35.
Heupel, R., Markgraf, T., Robinson, D. G. and Heldt, H. W. 1991. Compartmentation studies on spinach leaf peroxisomes : evidence for channeling of photorespiratory metabolites in peroxisomes devoid of intact boundary membrane. *Plant Physiol,* 96, 971-9.
Hu, J., Aguirre, M., Peto, C., Alonso, J., Ecker, J. and Chory, J. 2002. A role for peroxisomes in photomorphogenesis and development of Arabidopsis. *Science,* 297, 405-9.
Hu, J., Baker, A., Bartel, B., Linka, N., Mullen, R. T., Reumann, S. and Zolman, B. K. 2012. Plant peroxisomes: biogenesis and function. *Plant Cell,* 24, 2279-303.
Hunt, J. E. and Trelease, R. N. 2004. Sorting pathway and molecular targeting signals for the Arabidopsis peroxin 3. *Biochem Biophys Res Commun,* 314, 586-96.

Jinek, M., Rehwinkel, J., Lazarus, B. D., Izaurralde, E., Hanover, J. A. and Conti, E. 2004. The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha. *Nat Struct Mol Biol,* 11, 1001-7.
Kamada, T., Nito, K., Hayashi, H., Mano, S., Hayashi, M. and Nishimura, M. 2003. Functional differentiation of peroxisomes revealed by expression profiles of peroxisomal genes in Arabidopsis thaliana. *Plant Cell Physiol,* 44, 1275-89.
Kamisugi, Y., Mitsuya, S., El-Shami, M., Knight, C. D., Cuming, A. C. and Baker, A. 2016. Giant peroxisomes in a moss (*Physcomitrella patens*) peroxisomal biogenesis factor 11 mutant. *New Phytol,* 209, 576-89.
Kaplan, C. P., Thomas, J. E., Charlton, W. L. and Baker, A. 2001. Identification and characterisation of PEX6 orthologues from plants. *Biochim Biophys Acta,* 1539, 173-80.
Kataya, A. R. A., Heidari, B., Hagen, L., Kommedal, R., Slupphaug, G. and Lillo, C. 2015. Protein Phosphatase 2A holoenzyme is targeted to peroxisomes by piggybacking and positively affects peroxisomal β-oxidation. *Plant Physiol,* 167, 493-506.
Kato, A., Hayashi, M. and Nishimura, M. 1999. Oligomeric proteins containing N-terminal targeting signals are imported into peroxisomes in transgenic Arabidopsis. *Plant Cell Physiol,* 40, 586-91.
Kaur, N., Zhao, Q., Xie, Q. and Hu, J. 2013. Arabidopsis RING peroxins are E3 ubiquitin ligases that interact with two homologous ubiquitin receptor proteins. *J Integr Plant Biol,* 55, 108-20.
Keller, G. A., Gould, S., Deluca, M. and Subramani, S. 1987. Firefly luciferase is targeted to peroxisomes in mammalian cells. *PNAS,* 84, 3264-8.
Kerssen, D., Hambruch, E., Klaas, W., Platta, H. W., de Kruijff, B., Erdmann, R., Kunau, W. H. and Schliebs, W. 2006. Membrane association of the cycling peroxisome import receptor Pex5p. *J Biol Chem,* 281, 27003-15.
Kessel-Vigelius, S. K., Wiese, J., Schroers, M. G., Wrobel, T. J., Hahn, F. and Linka, N. 2013. An engineered plant peroxisome and its application in biotechnology. *Plant Science,* 210, 232-40.
Kim, J., Lee, H., Lee, H. N., Kim, S. H., Shin, K. D. and Chung, T. 2013. Autophagy-related proteins are required for degradation of peroxisomes in Arabidopsis hypocotyls during seedling growth. *Plant Cell,* 25, 4956-66.
Kim, P. K. and Hettema, E. H. 2015. Multiple pathways for protein transport to peroxisomes. *J Mol Biol,* 427, 1176-90.
Klein, A. T., Barnett, P., Bottger, G., Konings, D., Tabak, H. F. and Distel, B. 2001. Recognition of peroxisomal targeting signal type 1 by the import receptor Pex5p. *J Biol Chem,* 276, 15034-41.
Koller, A., Snyder, W. B., Faber, K. N., Wenzel, T. J., Rangell, L., Keller, G. A. and Subramani, S. 1999. Pex22p of *Pichia pastoris*, essential for peroxisomal matrix protein import, anchors the ubiquitin-conjugating enzyme, Pex4p, on the peroxisomal membrane. *J Cell Biol,* 146, 99-112.
Kragler, F., Lametschwandtner, G., Christmann, J., Hartig, A. and Harada, J. 1998. Identification and analysis of the plant peroxisomal targeting signal 1 receptor NtPEX5. *PNAS,* 95, 13336-41.
Kumar, A., Roach, C., Hirsh, I. S., Turley, S., deWalque, S., Michels, P. A. M. and Hol, W. G. J. 2001. An unexpected extended conformation for the third TPR motif of the peroxin PEX5 from Trypanosoma brucei. *J Mol Biol,* 307, 271-82.
Lametschwandtner, G., Brocard, C., Fransen, M., Van Veldhoven, P., Berger, J. and Hartig, A. 1998. The difference in recognition of terminal tripeptides as peroxisomal targeting signal 1 between yeast and human is due to different affinities of their receptor Pex5p to the cognate signal and to residues adjacent to it. *J Biol Chem,* 273, 33635-43.
Lanyon-Hogg, T., Warriner, S. L. and Baker, A. 2010. Getting a camel through the eye of a needle: the import of folded proteins by peroxisomes. *Biol Cell,* 102, 245-63.

Lanyon-Hogg, T. 2012. *Protein interactions in the peroxisomal docking complex of Arabidopsis thaliana.* University of Leeds.
Lanyon-Hogg, T., Hooper, J., Gunn, S., Warriner, S. L. and Baker, A. 2014. PEX14 binding to Arabidopsis PEX5 has differential effects on PTS1 and PTS2 cargo occupancy of the receptor. *FEBS Lett,* 588, 2223-9.
Lee, J. R., Jang, H. H., Park, J. H., Jung, J. H., Lee, S. S., Park, S. K., Chi, Y. H., Moon, J. C., Lee, Y. M., Kim, S. Y., Kim, J. Y., Yun, D. J., Cho, M. J., Lee, K. O. and Lee, S. Y. 2006. Cloning of two splice variants of the rice PTS1 receptor, OsPex5pL and OsPex5pS, and their functional characterization using pex5-deficient yeast and Arabidopsis. *Plant J,* 47, 457-66.
Léon, S., Goodman, J. M. and Subramani, S. 2006. Uniqueness of the mechanism of protein import into the peroxisome matrix: transport of folded, co-factor-bound and oligomeric proteins by shuttling receptors. *Biochim Biophys Acta,* 1763, 1552-64.
Li, D. M., Xiao, Y. H., Luo, M., Hou, L., Luo, X. Y., Luo, K. M. and Pei, Y. 2003. Cloning and characterization of the PTS2 receptor gene (GhPex7) from cotton (Gossypium hirsutum L.). *J Genet Genomics,* 30, 823-9.
Li, C., Wen, A., Shen, B., Lu, J., Huang, Y. and Chang, Y. 2011. FastCloning: a highly simplified, purification-free, sequence- and ligation-independent PCR cloning method. *BMC Biotechnol,* 11, 92.
Li, X. R., Li, H. J., Yuan, L., Liu, M., Shi, D. Q., Liu, J. and Yang, W. C. 2014. Arabidopsis DAYU/ABERRANT PEROXISOME MORPHOLOGY9 is a key regulator of peroxisome biogenesis and plays critical roles during pollen maturation and germination in planta. *Plant Cell,* 26, 619-35.
Lin, Y., Sun, L., Nguyen, L. V., Rachubinski, R. A. and Goodman, H. M. 1999. The Pex16p homolog SSE1 and storage organelle formation in Arabidopsis seeds. *Science,* 284, 328-30.
Lingard, M. J. and Trelease, R. N. 2006. Five Arabidopsis peroxin 11 homologs individually promote peroxisome elongation, duplication or aggregation. *J Cell Sci,* 119, 1961-72.
Lingner, T., Kataya, A., Antonicelli, G., Benichou, A., Nilssen, K., Chen, X., Siemsen, T., Morgenstern, B., Meinicke, P. and Reumann, S. 2011. Identification of novel plant peroxisomal targeting signals by a combination of machine learning methods and in vivo subcellular targeting analyses. *Plant Cell,* 23, 1556-72.
Liu, Y., Shah, K., Yang, F., Witucki, L. and Shokat, K. M. 1998. Engineering Src family protein kinases with unnatural nucleotide specificity. *Chem Biol,* 5, 91-101.
Lopez-Huertas, E., Oh, J. and Baker, A. 1999. Antibodies against Pex14p block ATP-independent binding of matrix proteins to peroxisomes in vitro. *FEBS Lett,* 459, 227-9.
Lopez-Huertas, E., Charlton, W., Johnson, B., Graham, I. and Baker, A. 2000. Stress induces peroxisome biogenesis genes. *EMBO J,* 19, 6770-7.
Madrid, K. P. and Jardim, A. 2005. Peroxin 5-peroxin 14 association in the protozoan Leishmania donovani involves a novel protein-protein interaction motif. *Biochem J,* 391, 105-14.
Mano, S., Nakamori, C., Nito, K., Kondo, M. and Nishimura, M. 2006. The Arabidopsis pex12 and pex13 mutants are defective in both PTS1- and PTS2-dependent protein transport to peroxisomes. *Plant J,* 47, 604-18.
Manton, I. 1961. Observations on phragmosomes. *J Exp Bot,* 12, 108-13.
Mast, F. D., Rachubinski, R. A. and Aitchison, J. D. 2015. Signaling dynamics and peroxisomes. *Curr Opin Cell Biol,* 35, 131-6.
Matsumura, T., Otera, H. and Fujiki, Y. 2000. Disruption of the interaction of the longer isoform of Pex5p, Pex5pL, with Pex7p abolishes peroxisome targeting signal type 2 protein import in mammals. Study with a novel Pex5-impaired Chinese hamster ovary cell mutant. *J Biol Chem,* 275, 21715-21.

Mayerhofer, P. U. 2016. Targeting and insertion of peroxisomal membrane proteins: ER trafficking versus direct delivery to peroxisomes. *Biochim Biophys Acta,* 1863, 870-80.
Maynard, E. L., Gatto, G. J., Jr. and Berg, J. M. 2004. Pex5p binding affinities for canonical and noncanonical PTS1 peptides. *Proteins,* 55, 856-61.
Maynard, E. L. and Berg, J. M. 2007. Quantitative analysis of peroxisomal targeting signal type-1 binding to wild-type and pathogenic mutants of Pex5p supports an affinity threshold for peroxisomal protein targeting. *J Mol Biol,* 368, 1259-66.
McNew, J. A. and Goodman, J. M. 1994. An oligomeric protein is imported into peroxisomes in vivo. *J Cell Biol,* 127, 1245-57.
Meinecke, M., Cizmowski, C., Schliebs, W., Kruger, V., Beck, S., Wagner, R. and Erdmann, R. 2010. The peroxisomal importomer constitutes a large and highly dynamic pore. *Nat Cell Biol,* 12, 273-7.
Meinecke, M., Bartsch, P. and Wagner, R. 2016. Peroxisomal protein import pores. *Biochim Biophys Acta,* 1863, 821-7.
Merrifield, R. B. 1963. Solid Phase Peptide Synthesis. I. The synthesis of a tetrapeptide. *J Am Chem Soc,* 85, 2149-54.
Mesa-Torres, N., Tomic, N., Albert, A., Salido, E. and Pey, A. L. 2015. Molecular recognition of PTS-1 cargo proteins by Pex5p: implications for protein mistargeting in primary hyperoxaluria. *Biomolecules,* 5, 121-41.
Miyata, N., Okumoto, K., Mukai, S., Noguchi, M. and Fujiki, Y. 2012. AWP1/ZFAND6 functions in Pex5 export by interacting with cys-monoubiquitinated Pex5 and Pex6 AAA ATPase. *Traffic,* 13, 168-83.
Mollenhauer, H. H., Morré, D. J. and Kelley, A. G. 1966. The widespread occurrence of plant cytosomes resembling animal microbodies. *Protoplasma,* 62, 44-52.
Monera, O. D., Sereda, T. J., Zhou, N. E., Kay, C. M. and Hodges, R. S. 1995. Relationship of sidechain hydrophobicity and alpha-helical propensity on the stability of the single-stranded amphipathic alpha-helix. *J Pept Sci,* 1, 319-29.
Monosov, E. Z., Wenzel, T. J., Luers, G. H., Heyman, J. A. and Subramani, S. 1996. Labeling of peroxisomes with green fluorescent protein in living P. pastoris cells. *J Histochem Cytochem,* 44, 581-9.
Monroe-Augustus, M., Ramon, N. M., Ratzel, S. E., Lingard, M. J., Christensen, S. E., Murali, C. and Bartel, B. 2011. Matrix proteins are inefficiently imported into Arabidopsis peroxisomes lacking the receptor-docking peroxin PEX14. *Plant Mol Biol,* 77, 1-15.
Mullen, R. T., Lee, M. S., Flynn, C. R. and Trelease, R. N. 1997. Diverse amino acid residues function within the type 1 peroxisomal targeting signal. Implications for the role of accessory residues upstream of the type 1 peroxisomal targeting signal. *Plant Physiol,* 115, 881-9.
Neuberger, G., Maurer-Stroh, S., Eisenhaber, B., Hartig, A. and Eisenhaber, F. 2003a. Motif refinement of the peroxisomal targeting signal 1 and evaluation of taxon-specific differences. *J Mol Biol,* 328, 567-79.
Neuberger, G., Maurer-Stroh, S., Eisenhaber, B., Hartig, A. and Eisenhaber, F. 2003b. Prediction of peroxisomal targeting signal 1 containing proteins from amino acid sequence. *J Mol Biol,* 328, 581-92.
Neufeld, C., Filipp, F. V., Simon, B., Neuhaus, A., Schüller, N., David, C., Kooshapur, H., Madl, T., Erdmann, R., Schliebs, W., Wilmanns, M. and Sattler, M. 2009. Structural basis for competitive interactions of Pex14 with the import receptors Pex5 and Pex19. *EMBO J,* 28, 745-54.
Neuhaus, A., Kooshapur, H., Wolf, J., Meyer, N. H., Madl, T., Saidowsky, J., Hambruch, E., Lazam, A., Jung, M., Sattler, M., Schliebs, W. and 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.

Nikolovska-Coleska, Z., Wang, R., Fang, X., Pan, H., Tomita, Y., Li, P., Roller, P. P., Krajewski, K., Saito, N. G., Stuckey, J. A. and Wang, S. 2004. Development and optimization of a binding assay for the XIAP BIR3 domain using fluorescence polarization. *Anal Biochem,* 332, 261-73.

Nito, K., Hayashi, M. and Nishimura, M. 2002. Direct interaction and determination of binding domains among peroxisomal import factors in Arabidopsis thaliana. *Plant Cell Physiol,* 43, 355-66.

Nito, K., Kamigaki, A., Kondo, M., Hayashi, M. and Nishimura, M. 2007. Functional classification of Arabidopsis peroxisome biogenesis factors proposed from analyses of knockdown mutants. *Plant Cell Physiol,* 48, 763-74.

Nordgren, M., Francisco, T., Lismont, C., Hennebel, L., Brees, C., Wang, B., Van Veldhoven, P. P., Azevedo, J. E. and Fransen, M. 2015. Export-deficient monoubiquitinated PEX5 triggers peroxisome removal in SV40 large T antigen-transformed mouse embryonic fibroblasts. *Autophagy,* 11, 1326-40.

Olsen, L. J. 1998. The surprising complexity of peroxisome biogenesis. *Plant Mol Biol,* 38, 163-89.

Pennington, M. W. 1994. Peptide synthesis protocols. *In:* PENNINGTON, M. W. and DUNN, B. M. (eds.) *Methods Mol Biol.* Totowa, New Jersey, USA: Humana Press Inc.

Petriv, O. I., Tang, L., Titorenko, V. I. and Rachubinski, R. A. 2004. A new definition for the consensus sequence of the peroxisome targeting signal type 2. *J Mol Biol*, 341, 119-34.

Platta, H. W., Grunau, S., Rosenkranz, K., Girzalsky, W. and Erdmann, R. 2005. Functional role of the AAA peroxins in dislocation of the cycling PTS1 receptor back to the cytosol. *Nat Cell Biol,* 7, 817-22.

Platta, H. W., El Magraoui, F., Schlee, D., Grunau, S., Girzalsky, W. and Erdmann, R. 2007. Ubiquitination of the peroxisomal import receptor Pex5p is required for its recycling. *J Cell Biol,* 177, 197-204.

Platta, H. W. and Erdmann, R. 2007. The peroxisomal protein import machinery. *FEBS Lett,* 581, 2811-9.

Platta, H. W., El Magraoui, F., Baumer, B. E., Schlee, D., Girzalsky, W. and Erdmann, R. 2009. Pex2 and pex12 function as protein-ubiquitin ligases in peroxisomal protein import. *Mol Cell Biol,* 29, 5505-16.

Platta, H. W., Brinkmeier, R., Reidick, C., Galiani, S., Clausen, M. P. and Eggeling, C. 2016. Regulation of peroxisomal matrix protein import by ubiquitination. *Biochim Biophys Acta,* 1863, 838-49.

Porter, K. R. and Caulfield, J. B. 1958. The formation of the cell plate during cytokinesis in Allium cepa L. *In:* BARGMANN, W., MÖLLENSTEDT, G., NIEHRS, H., et al. (eds.) *Vierter Internationaler Kongress für Elektronenmikroskopie / Fourth International Conference on Electron Microscopy / Quatrième Congrès International de Microscopie Électronique: Berlin 10.–17. September 1958.* Berlin, Heidelberg: Springer Berlin Heidelberg.

Porter, K. R. and Machado, R. D. 1960. Studies on the endoplasmic reticulum. IV. Its form and distribution during mitosis in cells of onion root tip. *J Biophys Biochem Cytol,* 7, 167-80.

Pracharoenwattana, I. and Smith, S. M. 2008. When is a peroxisome not a peroxisome? *Trends Plant Sci,* 13, 522-5.

Pratt, S. E., Speltz, E. B., Mochrie, S. G. and Regan, L. 2016. Designed proteins as novel imaging reagents in living *Escherichia coli*. *ChemBioChem,* 17, 1-7.

Prestele, J., Hierl, G., Scherling, C., Hetkamp, S., Schwechheimer, C., Isono, E., Weckwerth, W., Wanner, G. and Gietl, C. 2010. Different functions of the C_3HC_4 zinc RING finger peroxins PEX10, PEX2, and PEX12 in peroxisome formation and matrix protein import. *PNAS,* 107, 14915-20.

Purdue, P. E., Yang, X. and Lazarow, P. B. 1998. Pex18p and Pex21p, a novel pair of related peroxins essential for peroxisomal targeting by the PTS2 pathway. *J Cell Biol,* 143, 1859-69.

Ramón, N. M. and Bartel, B. 2010. Interdependence of the peroxisome-targeting receptors in *Arabidopsis thaliana*: PEX7 facilitates PEX5 accumulation and import of PTS1 cargo into peroxisomes. *Mol Biol Cell,* 21, 1263-71.

Ratzel, S. E., Lingard, M. J., Woodward, A. W. and Bartel, B. 2011. Reducing PEX13 expression ameliorates physiological defects of late-acting peroxin mutants. *Traffic,* 12, 121-34.

Reetz, M. T., Wang, L. W. and Bocola, M. 2006. Directed evolution of enantioselective enzymes: iterative cycles of CASTing for probing protein-sequence space. *Angew Chem Int Ed Engl,* 45, 1236-41.

Rensing, S. A., Lang, D., Zimmer, A. D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P. F., Lindquist, E. A., Kamisugi, Y., Tanahashi, T., Sakakibara, K., Fujita, T., Oishi, K., Shin, I. T., Kuroki, Y., Toyoda, A., Suzuki, Y., Hashimoto, S., Yamaguchi, K., Sugano, S., Kohara, Y., Fujiyama, A., Anterola, A., Aoki, S., Ashton, N., Barbazuk, W. B., Barker, E., Bennetzen, J. L., Blankenship, R., Cho, S. H., Dutcher, S. K., Estelle, M., Fawcett, J. A., Gundlach, H., Hanada, K., Heyl, A., Hicks, K. A., Hughes, J., Lohr, M., Mayer, K., Melkozernov, A., Murata, T., Nelson, D. R., Pils, B., Prigge, M., Reiss, B., Renner, T., Rombauts, S., Rushton, P. J., Sanderfoot, A., Schween, G., Shiu, S. H., Stueber, K., Theodoulou, F. L., Tu, H., Van de Peer, Y., Verrier, P. J., Waters, E., Wood, A., Yang, L., Cove, D., Cuming, A. C., Hasebe, M., Lucas, S., Mishler, B. D., Reski, R., Grigoriev, I. V., Quatrano, R. S. and Boore, J. L. 2008. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science,* 319, 64-9.

Reumann, S., Ma, C., Lemke, S. and Babujee, L. 2004. AraPerox. A database of putative Arabidopsis proteins from plant peroxisomes. *Plant Physiol,* 136, 2587-608.

Reumann, S. 2011. Toward a definition of the complete proteome of plant peroxisomes: Where experimental proteomics must be complemented by bioinformatics. *Proteomics,* 11, 1764-79.

Reumann, S., Buchwald, D. and Lingner, T. 2012. PredPlantPTS1: a web server for the prediction of plant peroxisomal proteins. *Front Plant Sci,* 3.

Reumann, S., Chowdhary, G. and Lingner, T. 2016. Characterization, prediction and evolution of plant peroxisomal targeting signals type 1 (PTS1s). *Biochim Biophys Acta,* 1863, 790-803.

Reumann, S. and Bartel, B. 2016. Plant peroxisomes: recent discoveries in functional complexity, organelle homeostasis, and morphological dynamics. *Curr Opin Plant Biol,* 34, 17-26.

Rhodin, J. 1954. *Correlation of ultrastructural organization and function in normal and experimentally changed proximal convoluted tubule cells of the mouse kidney.* Aktiebolaget Godvil, Stockholm.

Roy, A., Kucukural, A. and Zhang, Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc,* 5, 725-38.

Saidowsky, J., Dodt, G., Kirchberg, K., Wegner, A., Nastainczyk, W., Kunau, W. H. and 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.

Salomons, F. A., Kiel, J. A., Faber, K. N., Veenhuis, M. and van der Klei, I. J. 2000. Overproduction of Pex5p stimulates import of alcohol oxidase and dihydroxyacetone synthase in a Hansenula polymorpha Pex14 null mutant. *J Biol Chem,* 275, 12603-11.

Sampathkumar, P., Roach, C., Michels, P. A. and Hol, W. G. 2008. Structural insights into the recognition of peroxisomal targeting signal 1 by Trypanosoma brucei peroxin 5. *J Mol Biol,* 381, 867-80.

Schliebs, W., Saidowsky, J., Agianian, B., Dodt, G., Herberg, F. W. and Kunau, W. H. 1999. Recombinant human peroxisomal targeting signal receptor PEX5. Structural basis for interaction of PEX5 with PEX14. *J Biol Chem,* 274, 5666-73.

Schliebs, W., Girzalsky, W. and Erdmann, R. 2010. Peroxisomal protein import and ERAD: variations on a common theme. *Nat Rev Mol Cell Biol,* 11, 885-90.

Schumann, U., Wanner, G., Veenhuis, M., Schmid, M. and Gietl, C. 2003. AthPEX10, a nuclear gene essential for peroxisome and storage organelle formation during Arabidopsis embryogenesis. *PNAS,* 100, 9626-31.

Schwartzkopff, B., Platta, H. W., Hasan, S., Girzalsky, W. and Erdmann, R. 2015. Cysteine-specific ubiquitination protects the peroxisomal import receptor Pex5p against proteasomal degradation. *Biosci Rep,* 35.

Sereda, T. J., Mant, C. T., Sonnichsen, F. D. and Hodges, R. S. 1994. Reversed-phase chromatography of synthetic amphipathic alpha-helical peptides as a model for ligand/receptor interactions. Effect of changing hydrophobic environment on the relative hydrophilicity/hydrophobicity of amino acid side-chains. *J Chromatogr A,* 676, 139-53.

Shiozawa, K., Konarev, P. V., Neufeld, C., Wilmanns, M. and Svergun, D. I. 2009. Solution structure of human Pex5.Pex14.PTS1 protein complexes obtained by small angle X-ray scattering. *J Biol Chem,* 284, 25334-42.

Skoulding, N. 2011. *Determining the specificity of peroxisomal targeting signal 1 variants in Arabidopsis thaliana.* University of Leeds.

Skoulding, N. S., Chowdhary, G., Deus, M. J., Baker, A., Reumann, S. and Warriner, S. L. 2015. Experimental validation of plant peroxisomal targeting prediction algorithms by systematic comparison of in vivo import efficiency and in vitro PTS1 binding affinity. *J Mol Biol,* 427, 1085-101.

Smith, J. J. and Aitchison, J. D. 2013. Peroxisomes take shape. *Nat Rev Mol Cell Biol,* 14, 803-17.

Sparkes, I. A., Brandizzi, F., Slocombe, S. P., El-Shami, M., Hawes, C. and Baker, A. 2003. An Arabidopsis pex10 null mutant is embryo lethal, implicating peroxisomes in an essential role during plant embryogenesis. *Plant Physiol,* 133, 1809-19.

Speltz, E. B., Nathan, A. and Regan, L. 2015. Design of protein-peptide interaction modules for assembling supramolecular structures in vivo and in vitro. *ACS Chem Biol,* 10, 2108-15.

Spiltoir, J. I., Strickland, D., Glotzer, M. and Tucker, C. L. 2016. Optical control of peroxisomal trafficking. *ACS Synth Biol,* 5, 554-60.

Stanley, W. A., Filipp, F. V., Kursula, P., Schuller, N., Erdmann, R., Schliebs, W., Sattler, M. and Wilmanns, M. 2006. Recognition of a functional peroxisome type 1 target by the dynamic import receptor Pex5p. *Mol Cell,* 24, 653-63.

Stanley, W. A., Pursiainen, N. V., Garman, E. F., Juffer, A. H., Wilmanns, M. and Kursula, P. 2007. A previously unobserved conformation for the human Pex5p receptor suggests roles for intrinsic flexibility and rigid domain motions in ligand binding. *BMC Struct Biol,* 7, 24.

Steen, H. and Mann, M. 2004. The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol,* 5, 699-711.

Su, J. R., Takeda, K., Tamura, S., Fujiki, Y. and Miki, K. 2009. Crystal structure of the conserved N-terminal domain of the peroxisomal matrix protein import receptor, Pex14p. *PNAS,* 106, 417-21.

Tanner, P., Balasubramanian, V. and Palivan, C. G. 2013. Aiding nature's organelles: artificial peroxisomes play their role. *Nano Lett,* 13, 2875-83.

The QIAexpressionist 2001. A handbook for high-level expression and purification of 6xHis-tagged proteins. *QIAGEN.* Fifth ed. Hilden, Germany.

Titorenko, V. I., Smith, J. J., Szilard, R. K. and Rachubinski, R. A. 1998. Pex20p of the yeast Yarrowia lipolytica is required for the oligomerization of thiolase in the cytosol and for its targeting to the peroxisome. *J Cell Biol,* 142, 403-20.

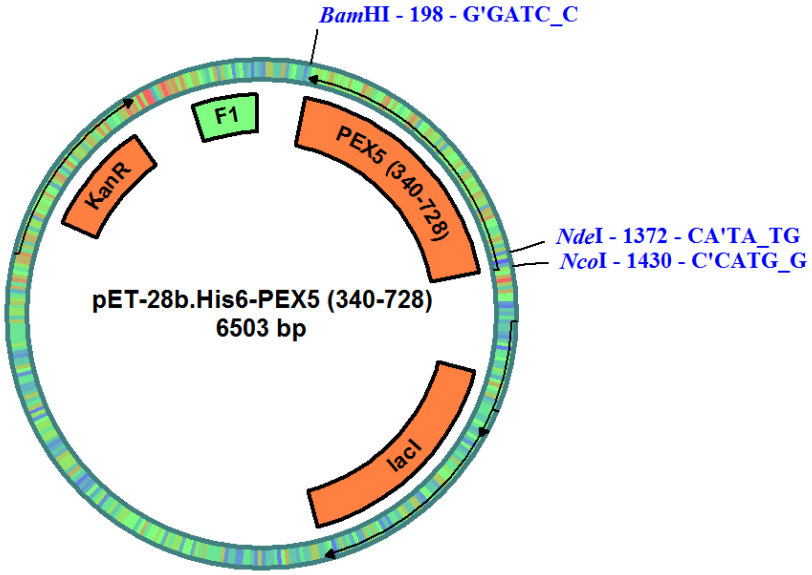
Volokita, M. 1991. The carboxy-terminal end of glycolate oxidase directs a foreign protein into tobacco leaf peroxisomes. *Plant J,* 1, 361-6.
Waller, J. C., Dhanoa, P. K., Schumann, U., Mullen, R. T. and Snedden, W. A. 2010. Subcellular and tissue localization of NAD kinases from Arabidopsis: compartmentalization of de novo NADP biosynthesis. *Planta,* 231, 305-17.
Walton, P. A., Hill, P. E. and Subramani, S. 1995. Import of stably folded proteins into peroxisomes. *Mol Biol Cell,* 6, 675-83.
Wang, K., Neumann, H., Peak-Chew, S. Y. and Chin, J. W. 2007. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat Biotechnol,* 25, 770-7.
Williams, C., van den Berg, M., Sprenger, R. R. and Distel, B. 2007. A conserved cysteine is essential for Pex4p-dependent ubiquitination of the peroxisomal import receptor Pex5p. *J Biol Chem,* 282, 22534-43.
Wimmer, C., Schmid, M., Veenhuis, M. and Gietl, C. 1998. The plant PTS1 receptor: similarities and differences to its human and yeast counterparts. *Plant J,* 16, 453-64.
Woodward, A. W. and Bartel, B. 2005. The Arabidopsis peroxisomal targeting signal type 2 receptor PEX7 is necessary for peroxisome function and dependent on PEX5. *Mol Biol Cell,* 16, 573-83.
Woodward, A. W., Fleming, W. A., Burkhart, S. E., Ratzel, S. E., Bjornson, M. and Bartel, B. 2014. A viable Arabidopsis pex13 missense allele confers severe peroxisomal defects and decreases PEX5 association with peroxisomes. *Plant Mol Biol,* 86**,** 201-14.
Xie, J. and Schultz, P. G. 2006. A chemical toolkit for proteins — an expanded genetic code. *Nat Rev Mol Cell Biol,* 7, 775-82.
Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. and Zhang, Y. 2015. The I-TASSER Suite: protein structure and function prediction. *Nat Methods,* 12, 7-8.
Yang, J. and Zhang, Y. 2015. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res,* 43, W174-81.
Yoshimoto, K., Shibata, M., Kondo, M., Oikawa, K., Sato, M., Toyooka, K., Shirasu, K., Nishimura, M. and Ohsumi, Y. 2014. Organ-specific quality control of plant peroxisomes is mediated by autophagy. *J Cell Sci,* 127, 1161-8.
Young, P. G. and Bartel, B. 2016. Pexophagy and peroxisomal protein turnover in plants. *Biochim Biophys Acta,* 1863, 999-1005.
Zhang, Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics,* 9, 40.
Zhu, X. G., Long, S. P. and Ort, D. R. 2010. Improving photosynthetic efficiency for greater yield. *Annu Rev Plant Biol,* 61, 235-61.
Zolman, B. K., Yoder, A. and Bartel, B. 2000. Genetic Analysis of Indole-3-butyric Acid Responses in Arabidopsis thaliana Reveals Four Mutant Classes. *Genetics,* 156, 1323-37.
Zolman, B. K. and Bartel, B. 2004. An Arabidopsis indole-3-butyric acid-response mutant defective in PEROXIN6, an apparent ATPase implicated in peroxisomal function. *PNAS,* 101, 1786-91.
Zolman, B. K., Monroe-Augustus, M., Silva, I. D. and Bartel, B. 2005. Identification and functional characterization of Arabidopsis PEROXIN4 and the interacting protein PEROXIN22. *Plant Cell,* 17, 3422-35.

Appendix A

Recombinant protein constructs

A.1 His$_6$-*At*PEX5(340-728)

A.1.1 His$_6$-*At*PEX5(340-728) plasmid map



A.1.2 His$_6$-*At*PEX5(340-728) DNA sequence

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCAAGCTT
CAGCCCCCGGGGAATGGGCTACTGAATATGAACAGCAGTATCTGGGGCCACCAAGTTGGGCTGATCAATT
TGCAAATGAGAAACTTTCACATGGACCAGAACAGTGGGCTGATGAGTTTGCTTCCGGGAGAGGACAGCAA
GAAACAGCTGAGGACCAATGGGTTAATGAGTTTTCAAAGTTGAATGTTGATGACTGGATAGATGAATTTG
CTGAAGGTCCCGTGGGTGATAGTTCAGCTGATGCATGGGCAAATGCTTACGATGAGTTTCTGAATGAGAA
AAATGCTGGAAAACAAACCAGTGGTGTCTACGTCTTCTCTGACATGAATCCTTATGTGGGTCACCCTGAA
CCTATGAAAGAAGGGCAAGAATTGTTTCGAAAAGGACTTCTGAGTGAAGCAGCGCTTGCTCTAGAAGCTG
AGGTTATGAAAAACCCTGAGAATGCTGAAGGTTGGAGATTACTTGGGGTCACACACGCAGAGAACGATGA
TGATCAACAGGCAATAGCTGCAATGATGCGTGCACAGGAGGCTGATCCCACAAATCTAGAGGTGCTTCTT
GCGCTTGGTGTGAGTCATACCAACGAGTTAGAGCAAGCAACTGCTTTGAAATATCTATATGGATGGCTGC
GAAATCACCCAAAGTATGGAGCAATTGCGCCTCCGGAGCTAGCGGATTCTTTGTACCATGCTGATATTGC
TAGATTATTCAATGAAGCTTCTCAGTTGAATCCTGAGGACGCCGATGTGCATATAGTGTTGGGCGTGCTC
TACAATCTGTCGAGAGAGTTCGATAGAGCAATCACATCCTTCCAAACAGCATTACAACTAAAACCAAACG
ATTATTCTCTGTGGAATAAGCTAGGTGCAACGCAAGCCAACAGTGTCCAGAGTGCTGATGCCATATCTGC
TTATCAACAGGCTCTAGATTTAAAACCAAATTATGTTCGTGCTTGGGCAAACATGGGAATCAGTTACGCA
AACCAGGGGATGTACAAAGAATCAATCCCGTATTATGTCCGTGCCCTTGCGATGAATCCTAAAGCTGATA
ACGCATGGCAATACTTGAGACTCTCGTTAAGTTGTGCATCAAGGCAAGACATGATAGAAGCTTGTGAGTC
AAGGAATCTCGATCTCTTGCAGAAAGAATTCCCGCTGTGA

Highlighted in green: start codons
Highlighted in cyan: His$_6$ tag
Highlighted in red: stop codon

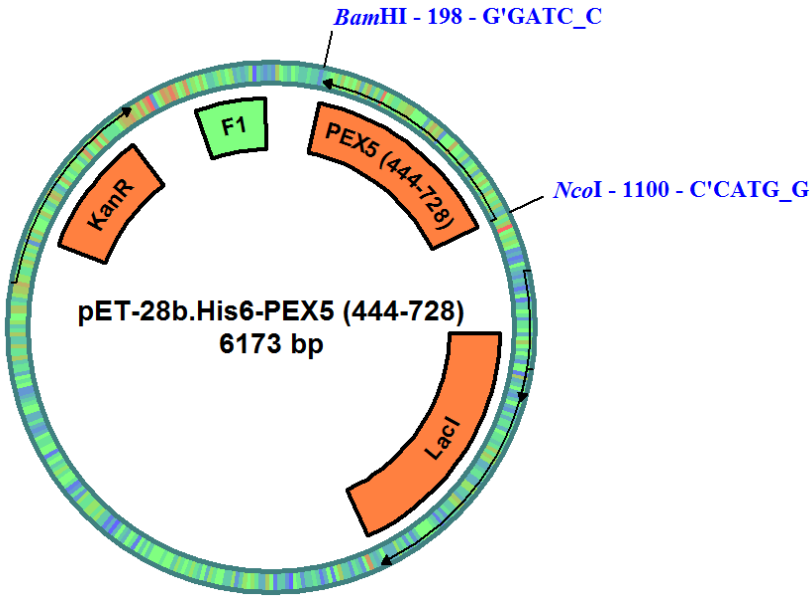
A.1.3 His$_6$-*At*PEX5(340-728) protein sequence

GSSHHHHHHSSGLVPRGSHMQASAPGEWATEYEQQYLGPPSWADQFANEKLSHGPEQWADEFASGRGQQE
TAEDQWVNEFSKLNVDDWIDEFAEGPVGDSSADAWANAYDEFLNEKNAGKQTSGVYVFSDMNPYVGHPEP
MKEGQELFRKGLLSEAALALEAEVMKNPENAEGWRLLGVTHAENDDDQQAIAAMMRAQEADPTNLEVLLA
LGVSHTNELEQATALKYLYGWLRNHPKYGAIAPPELADSLYHADIARLFNEASQLNPEDADVHIVLGVLY
NLSREFDRAITSFQTALQLKPNDYSLWNKLGATQANSVQSADAISAYQQALDLKPNYVRAWANMGISYAN
QGMYKESIPYYVRALAMNPKADNAWQYLRLSLSCASRQDMIEACESRNLDLLQKEFPL

Highlighted in cyan: His$_6$ tag
Underlined: WxxxF/Y repeats

A.2 His$_6$-*At*PEX5(444-728)

A.2.1 His$_6$-*At*PEX5(444-728) plasmid map



A.2.2 His$_6$-*At*PEX5(444-728) DNA sequence

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGGTCTACGTCTTCTCTGACATGAATC
CTTATGTGGGTCACCCTGAACCTATGAAAGAAGGGCAAGAATTGTTTCGAAAAGGACTTCTGAGTGAAGC
AGCGCTTGCTCTAGAAGCTGAGGTTATGAAAAACCCTGAGAATGCTGAAGGTTGGAGATTACTTGGGGTC
ACACACGCAGAGAACGATGATGATCAACAGGCAATAGCTGCAATGATGCGTGCACAGGAGGCTGATCCCA
CAAATCTAGAGGTGCTTCTTGCGCTTGGTGTGAGTCATACCAACGAGTTAGAGCAAGCAACTGCTTTGAA
ATATCTATATGGATGGCTGCGAAATCACCCAAAGTATGGAGCAATTGCGCCTCCGGAGCTAGCGGATTCT
TTGTACCATGCTGATATTGCTAGATTATTCAATGAAGCTTCTCAGTTGAATCCTGAGGACGCCGATGTGC
ATATAGTGTTGGGCGTGCTCTACAATCTGTCGAGAGAGTTCGATAGAGCAATCACATCCTTCCAAACAGC
ATTACAACTAAAACCAAACGATTATTCTCTGTGGAATAAGCTAGGTGCAACGCAAGCCAACAGTGTCCAG
AGTGCTGATGCCATATCTGCTTATCAACAGGCTCTAGATTTAAAACCAAATTATGTTCGTGCTTGGGCAA
ACATGGGAATCAGTTACGCAAACCAGGGGATGTACAAAGAATCAATCCCGTATTATGTCCGTGCCCTTGC
GATGAATCCTAAAGCTGATAACGCATGGCAATACTTGAGACTCTCGTTAAGTTGTGCATCAAGGCAAGAC
ATGATAGAAGCTTGTGAGTCAAGGAATCTCGATCTCTTGCAGAAAGAATTCCCGCTGTGA

Highlighted in green: start codon
Highlighted in cyan: His$_6$ tag
Highlighted in red: stop codon

A.2.3 His_6-*At*PEX5(444-728) protein sequence

GSSHHHHHHSSGLVVYVFSDMNPYVGHPEPMKEGQELFRKGLLSEAALALEAEVMKNPENAEG
WRLLGVTHAENDDDQQAIAAMMRAQEADPTNLEVLLALGVSHTNELEQATALKYLYGWLRNHP
KYGAIAPPELADSLYHADIARLFNEASQLNPEDADVHIVLGVLYNLSREFDRAITSFQTALQL
KPNDYSLWNKLGATQANSVQSADAISAYQQALDLKPNYVRAWANMGISYANQGMYKESIPYYV
RALAMNPKADNAWQYLRLSLSCASRQDMIEACESRNLDLLQKEFPL

Highlighted in cyan: His_6 tag

Appendix B

Primers for *At*PEX5(340-728) mutants and *At*PEX5(444-728)

B.1 *At*PEX5(340-728) D505A primers

Forward (5'-3'): CACACGCAGAGAACGCTGATGATCAACAGGC

Reverse (5'-3'): GCCTGTTGATCATCAGCGTTCTCTGCGTGTG

B.2 *At*PEX5(340-728) D505H and *At*PEX5(340-728) D505H(-N601A) primers

Forward (5'-3'): CACACACGCAGAGAACCATGATGATCAACAGGC

Reverse (5'-3'): GCCTGTTGATCATCATGGTTCTCTGCGTGTGTG

B.3 *At*PEX5(340-728) D505K primers

Forward (5'-3'): GGTCACACACGCAGAGAACAAGGATGATCAACAGGCAATAG

Reverse (5'-3'): CTATTGCCTGTTGATCATCCTTGTTCTCTGCGTGTGTGACC

B.4 *At*PEX5(340-728) D507A primers

Forward (5'-3'): CGCAGAGAACGATGATGCTCAACAGGCAATAGCTG

Reverse (5'-3'): CAGCTATTGCCTGTTGAGCATCATCGTTCTCTGCG

B.5 *At*PEX5(340-728) D507K primers

Forward (5'-3'): CACGCAGAGAACGATGATAAACAACAGGCAATAGCTGC

Reverse (5'-3'): GCAGCTATTGCCTGTTGTTTATCATCGTTCTCTGCGTG

B.6 *At*PEX5(340-728) V533A primers

Forward (5'-3'): CTTCTTGCGCTTGGTGCGAGTCATACCAACGAG

Reverse (5'-3'): CTCGTTGGTATGACTCGCACCAAGCGCAAGAAG

B.7 *At*PEX5(340-728) V533W primers

Forward (5'-3'): GGTGCTTCTTGCGCTTGGTTGGAGTCATACCAACGAGTTAG

Reverse (5'-3'): CTAACTCGTTGGTATGACTCCAACCAAGCGCAAGAAGCACC

B.8 *At*PEX5(340-728) T536A primers

Forward (5'-3'): CTTGGTGTGAGTCATGCCAACGAGTTAGAGC

Reverse (5'-3'): GCTCTAACTCGTTGGCATGACTCACACCAAG

B.9 *At*PEX5(340-728) T536N primers

Forward (5'-3'): CGCTTGGTGTGAGTCATAACAACGAGTTAGAGCAAGC

Reverse (5'-3'): GCTTGCTCTAACTCGTTGTTATGACTCACACCAAGCG

B.10 *At*PEX5(340-728) T536W primers

Forward (5'-3'): CTTGCGCTTGGTGTGAGTCATTGGAACGAGTTAGAGCAAGCAAC

Reverse (5'-3'): GTTGCTTGCTCTAACTCGTTCCAATGACTCACACCAAGCGCAAG

B.11 *At*PEX5(340-728) N537A primers

Forward (5'-3'): CTTGGTGTGAGTCATACCGCCGAGTTAGAGCAAGCAAC

Reverse (5'-3'): GTTGCTTGCTCTAACTCGGCGGTATGACTCACACCAAG

B.12 *At*PEX5(340-728) N537Q primers

Forward (5'-3'): CGCTTGGTGTGAGTCATACCCAGGAGTTAGAGCAAGCAACTGC

Reverse (5'-3'): GCAGTTGCTTGCTCTAACTCCTGGGTATGACTCACACCAAGCG

B.13 *At*PEX5(340-728) N537T primers

Forward (5'-3'): CTTGGTGTGAGTCATACCACCGAGTTAGAGCAAGCAAC

Reverse (5'-3'): GTTGCTTGCTCTAACTCGGTGGTATGACTCACACCAAG

B.14 *At*PEX5(340-728) E538A primers

Forward (5'-3'): GTGAGTCATACCAACGCGTTAGAGCAAGCAACTG

Reverse (5'-3'): CAGTTGCTTGCTCTAACGCGTTGGTATGACTCAC

B.15 *At*PEX5(340-728) N601A primers

Forward (5'-3'): GTTGGGCGTGCTCTACGCTCTGTCGAGAGAGTTC

Reverse (5'-3'): GAACTCTCTCGACAGAGCGTAGAGCACGCCCAAC

B.16 *At*PEX5(340-728) N601Q primers

Forward (5'-3'): GTTGGGCGTGCTCTACCAGCTGTCGAGAGAGTTC

Reverse (5'-3'): GAACTCTCTCGACAGCTGGTAGAGCACGCCCAAC

B.17 *At*PEX5(340-728) F613A primers

Forward (5'-3'): GATAGAGCAATCACATCCGCCCAAACAGCATTACAAC

Reverse (5'-3'): GTTGTAATGCTGTTTGGGCGGATGTGATTGCTCTATC

B.18 *At*PEX5(340-728) N628A primers

Forward (5'-3'): CGATTATTCTCTGTGGGCTAAGCTAGGTGCAACGC

Reverse (5'-3'): GCGTTGCACCTAGCTTAGCCCACAGAGAATAATCG

B.19 *At*PEX5(340-728) A632G primers

Forward (5'-3'): GGAATAAGCTAGGTGGAACGCAAGCCAACAG

Reverse (5'-3'): CTGTTGGCTTGCGTTCCACCTAGCTTATTCC

B.20 *At*PEX5(340-728) N636A primers

Forward (5'-3'): CTAGGTGCAACGCAAGCCGCCAGTGTCCAGAGTGCTGATG

Reverse (5'-3'): CATCAGCACTCTGGACACTGGCGGCTTGCGTTGCACCTAG

B.21 *At*PEX5(340-728) Y647F primers

Forward (5'-3'): GCTGATGCCATATCTGCTTTTCAACAGGCTCTAG

Reverse (5'-3'): CTAGAGCCTGTTGAAAAGCAGATATGGCATCAGC

B.22 *At*PEX5(340-728) R659A primers

Forward (5'-3'):

GATTTAAAACCAAATTATGTTGCTGCTTGGGCAAACATGGGAATCAG

Reverse (5'-3'):

CTGATTCCCATGTTTGCCCAAGCAGCAACATAATTTGGTTTTAAATC

B.23 *At*PEX5(340-728) N663A primers

Forward (5'-3'): GTTCGTGCTTGGGCAGCCATGGGAATCAGTTAC

Reverse (5'-3'): GTAACTGATTCCCATGGCTGCCCAAGCACGAAC

B.24 *At*PEX5(340-728) S667A primers

Forward (5'-3'): GGCAAACATGGGAATCGCTTACGCAAACCAGGGG

Reverse (5'-3'): CCCCTGGTTTGCGTAAGCGATTCCCATGTTTGCC

B.25 *At*PEX5(340-728) D505K-D507K primers

Forward (5'-3'): GGTCACACACGCAGAGAACAAAGATAAACAACAGGCAATAGCTGC

Reverse (5'-3'): GCAGCTATTGCCTGTTGTTTATCTTTGTTCTCTGCGTGTGTGACC

B.26 *At*PEX5(340-728) D505F(-N601A) primers

Forward (5'-3'): GGTCACACACGCAGAGAACTTTGATGATCAACAGGCAATAGCTGC

Reverse (5'-3'): GCAGCTATTGCCTGTTGATCATCAAAGTTCTCTGCGTGTGTGACC

B.27 *At*PEX5(340-728) D505F-D507F(-N601A) primers

Forward (5'-3'): GGTCACACACGCAGAGAACTTTGATTTTCAACAGGCAATAGCTGC

Reverse (5'-3'): GCAGCTATTGCCTGTTGAAAATCAAAGTTCTCTGCGTGTGTGACC

B.28 *At*PEX5(340-728) D505H/D507H(-N601A) primers

Forward (5'-3'): GGTCACACACGCAGAGAACCATGATCATCAACAGGCAATAGCTGC

Reverse (5'-3'): GCAGCTATTGCCTGTTGATGATCATGGTTCTCTGCGTGTGTGACC

B.29 *At*PEX5(340-728) D505H-D507T(-N601A) primers

Forward (5'-3'): CACGCAGAGAACCATGATACTCAACAGGCAATAGCTGC

Reverse (5'-3'): GCAGCTATTGCCTGTTGAGTATCATGGTTCTCTGCGTG

B.30 *At*PEX5(340-728) D505H-D507V(-N601A) primers

Forward (5'-3'): CGCAGAGAACCATGATGTTCAACAGGCAATAGCTG

Reverse (5'-3'): CAGCTATTGCCTGTTGAACATCATGGTTCTCTGCG

B.31 *At*PEX5(340-728) D505x/D507x primers

Forward (5'-3'): GGTCACACACGCAGAGAACVRNGATVRNCAACAGGCAATAGCTGC

Reverse (5'-3'): GCAGCTATTGCCTGTTGNYBATCNYBGTTCTCTGCGTGTGTGACC

B.32 *At*PEX5(340-728) V533x/T536x primers

Forward (5'-3'):

GGTGCTTCTTGCGCTTGGTVYNAGTCATNVYAACGAGTTAGAGCAAGC

Reverse (5'-3'):

GCTTGCTCTAACTCGTTRBNATGACTNRBACCAAGCGCAAGAAGCACC

B.33 *At*PEX5(340-728) N601x primers

Forward (5'-3'): GTGTTGGGCGTGCTCTACNNKCTGTCGAGAGAGTTCGATAG

Reverse (5'-3'): CTATCGAACTCTCTCGACAGMNNGTAGAGCACGCCCAACAC

B.34 *At*PEX5(444-728) primers

Forward (5'-3'): CCTGGTGGTCTACGTCTTCTCTGACATGAATCC

Reverse (5'-3'): GACGTAGACCACCAGGCCGCTGCTG

Appendix C

Mass Spectrometry of *At*PEX5-C Protein Variants

Liquid chromatography-mass spectrometry of *At*PEX5-C and variants was carried out using a Bruker maXis impact Q-TOF mass spectrometer and proteins eluted between 3 and 3.5 minutes. In some of the spectra, the mass + approximately 76 g/mol, and the mass + approximately (2*76 g/mol) is observed. These peaks correspond to the mass of the protein + one or two molecules of 2-mercaptoethanol. *At*PEX5 C has two surface-exposed cysteine residues (not near the PTS1-binding site), available to react and form disulfide bonds with 2-mercaptoethanol, and TCEP was used to reduce these disulfide bonds with varying degrees of success.

C.1 Wild-type *At*PEX5(340-728): MS trace

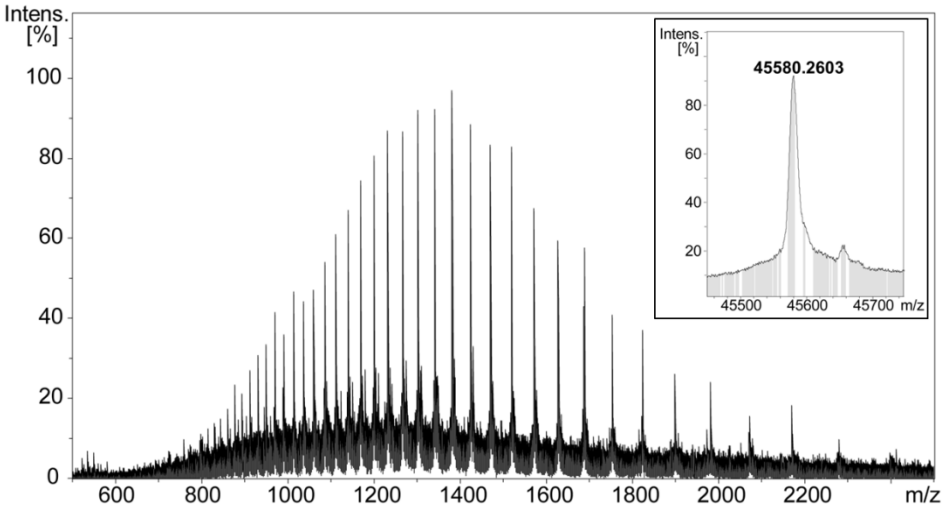


Figure C-1| Mass spectrum and deconvoluted mass data (inset) for wild-type *At*PEX5-C. Expected mass: 45,580.3 Da. Observed mass: 45,580.3 Da.

C.2 *At*PEX5(340-728) D505A: MS trace

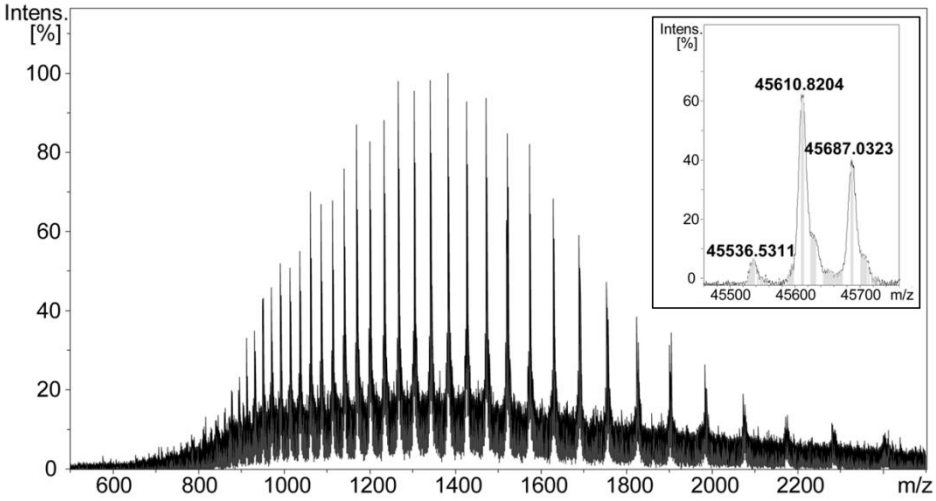


Figure C-2| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C D505A. Expected mass: 45,536.3 Da. Observed mass: 45,536.5 Da.

C.3 *At*PEX5(340-728) D505H: MS trace

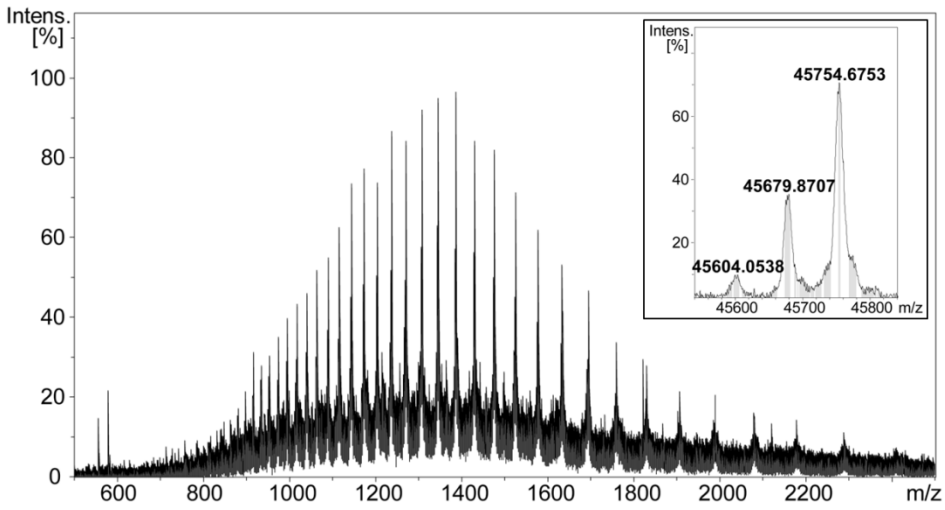


Figure C-3| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C D505H. Expected mass: 45,602.4 Da. Observed mass: 45,604.1 Da.

C.4 *At*PEX5(340-728) D505K: MS trace

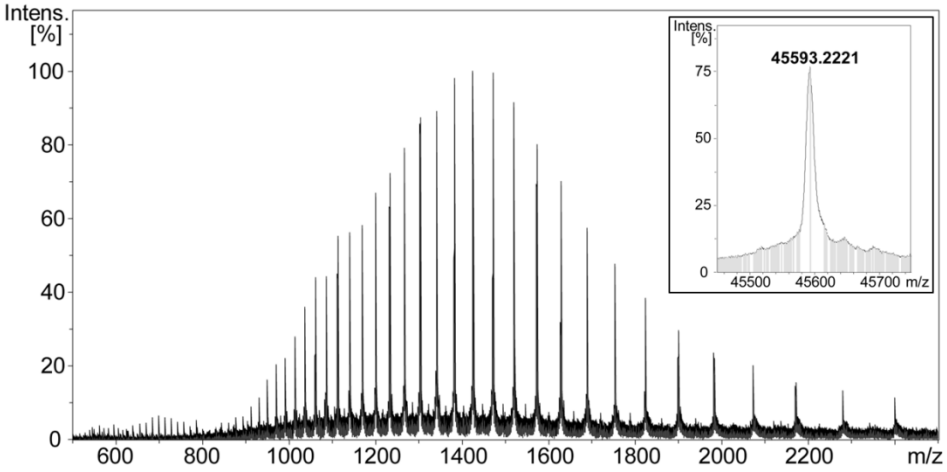


Figure C-4| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C D505K. Expected mass: 45,593.4 Da. Observed mass: 45,593.2 Da.

C.5 *At*PEX5(340-728) D507A: MS trace

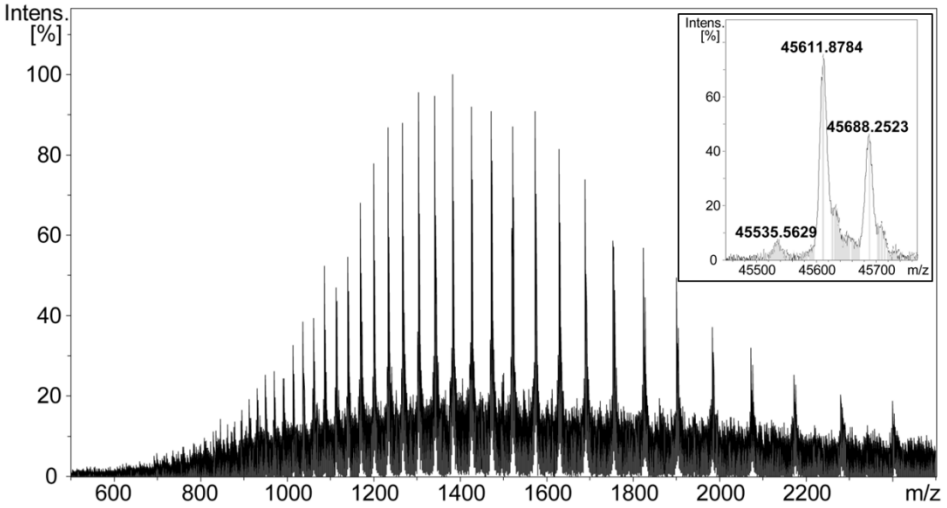


Figure C-5| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C D507A. Expected mass: 45,536.3 Da. Observed mass: 45,535.6 Da.

C.6 *At*PEX5(340-728) D507K: MS trace

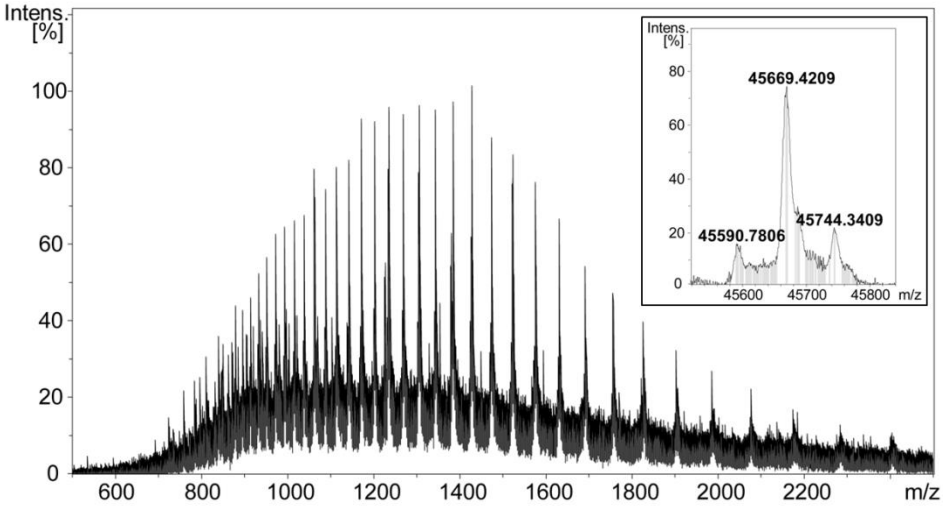


Figure C-6| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C D507K. Expected mass: 45,593.4 Da. Observed mass: 45,590.8 Da.

C.7 *At*PEX5(340-728) V533A: MS trace

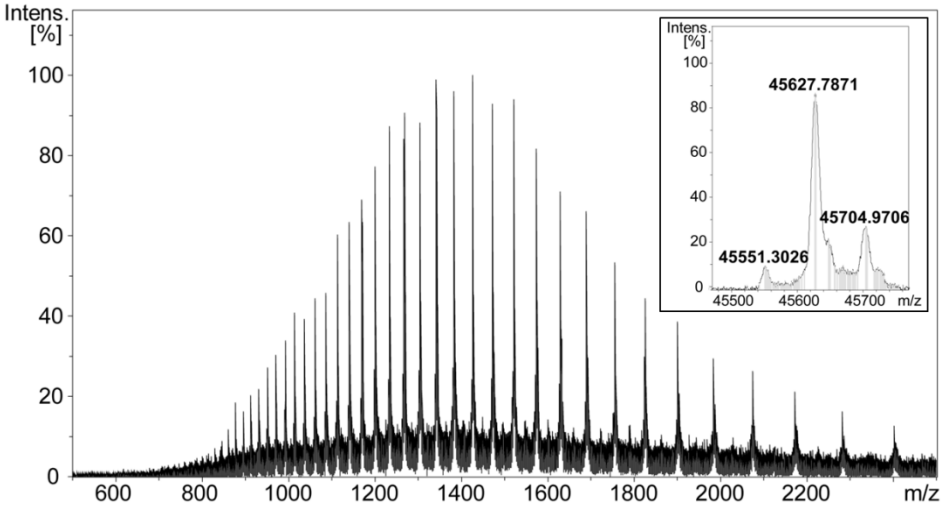


Figure C-7| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C V533A. Expected mass: 45,552.3 Da. Observed mass: 45,551.3 Da.

C.8 *At*PEX5(340-728) V533W: MS trace

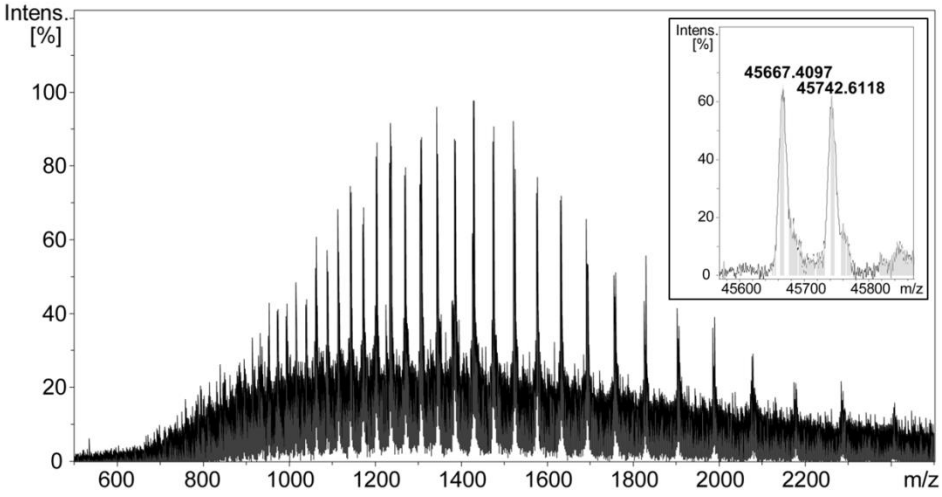


Figure C-8| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C V533W. Expected mass: 45,667.4 Da. Observed mass: 45,667.4 Da.

C.9 *At*PEX5(340-728) T536A: MS trace

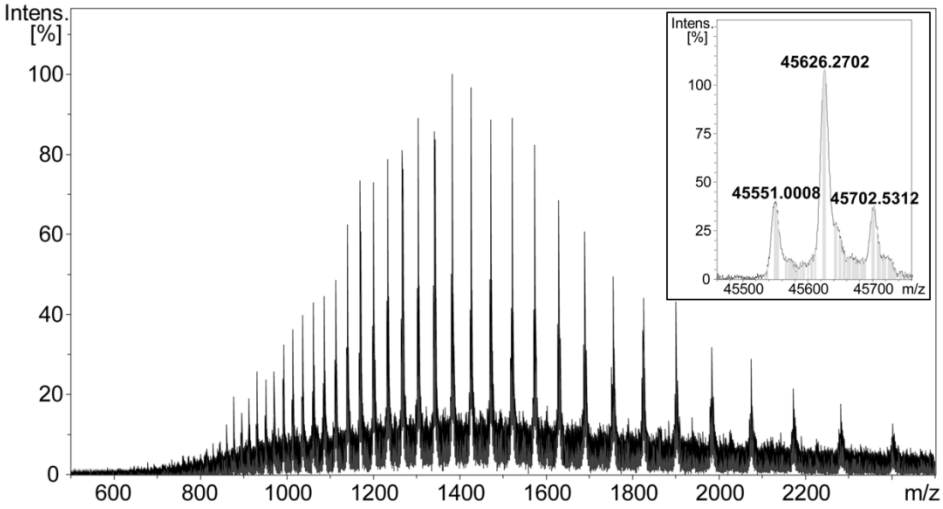


Figure C-9| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C T536A. Expected mass: 45,550.3 Da. Observed mass: 45,551.0 Da.

C.10 *At*PEX5(340-728) T536N: MS trace

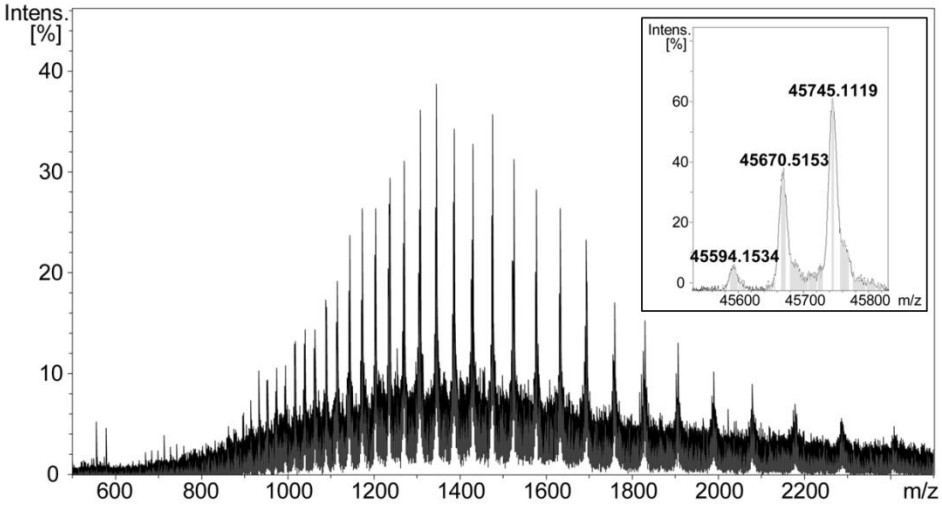


Figure C-10| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C T536N. Expected mass: 45,593.3 Da. Observed mass: 45,594.2 Da.

C.11 *At*PEX5(340-728) T536W: MS trace

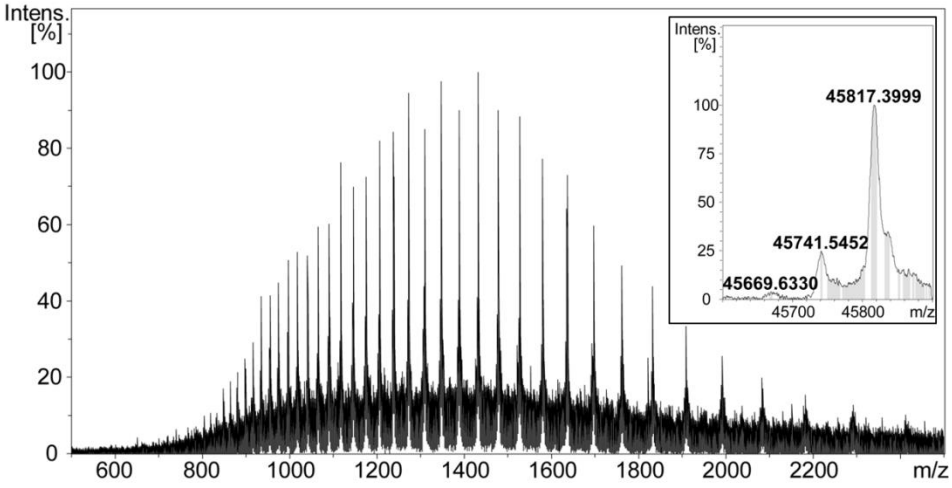


Figure C-11| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C T536W. Expected mass: 45,665.4 Da. Observed mass: 45,669.6 Da.

C.12 *At*PEX5(340-728) N537A: MS trace

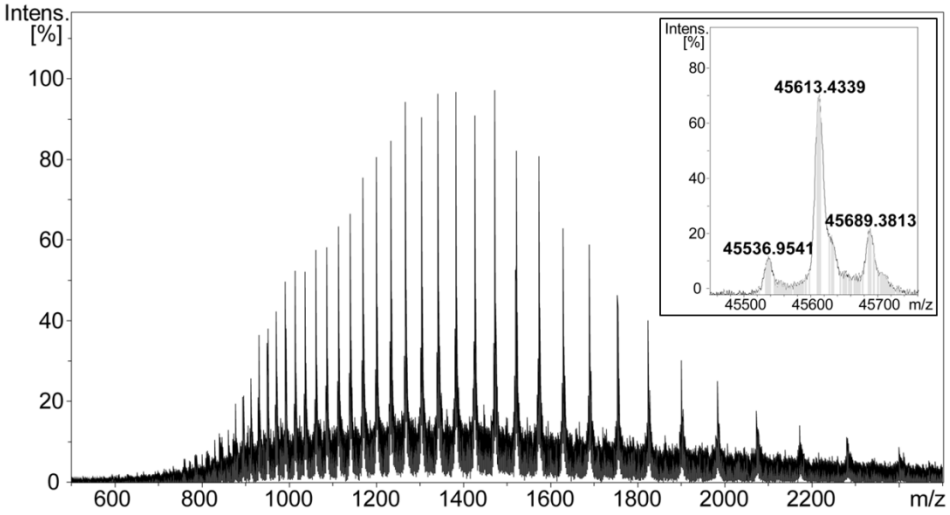


Figure C-12| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C N537A. Expected mass: 45,537.3 Da. Observed mass: 45,537.0 Da.

C.13 *At*PEX5(340-728) N537Q: MS trace

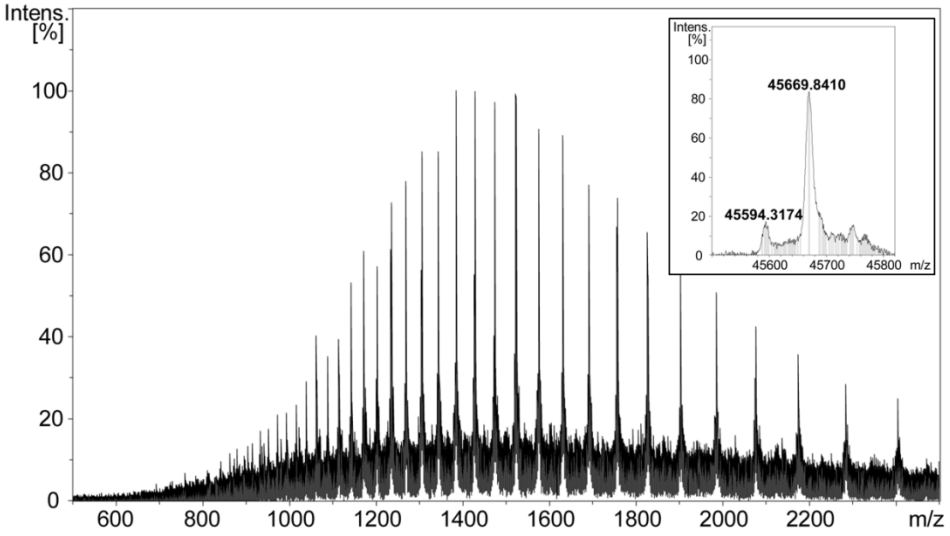


Figure C-13| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C N537Q. Expected mass: 45,594.3 Da. Observed mass: 45,594.3 Da.

C.14 *At*PEX5(340-728) N537T: MS trace

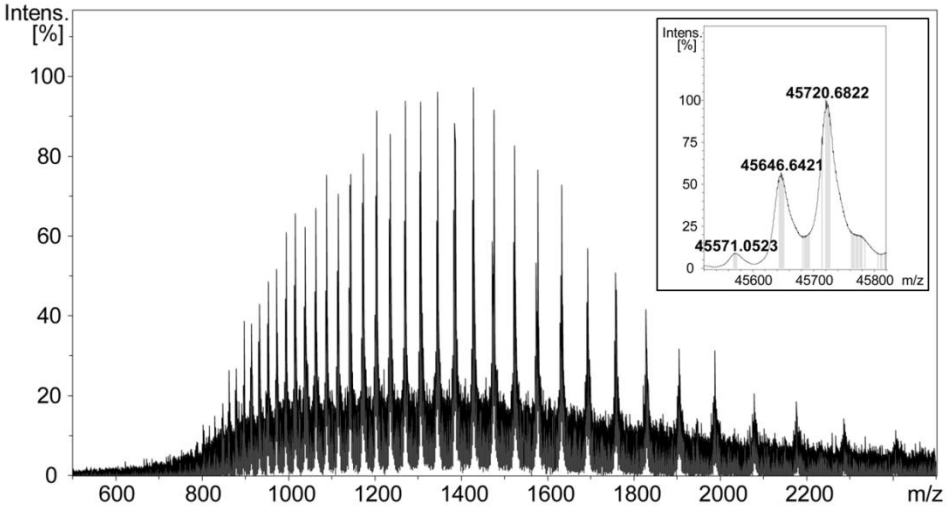


Figure C-14| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C N537T. Expected mass: 45,567.3 Da. Observed mass: 45,571.1 Da.

C.15 *At*PEX5(340-728) E538A: MS trace

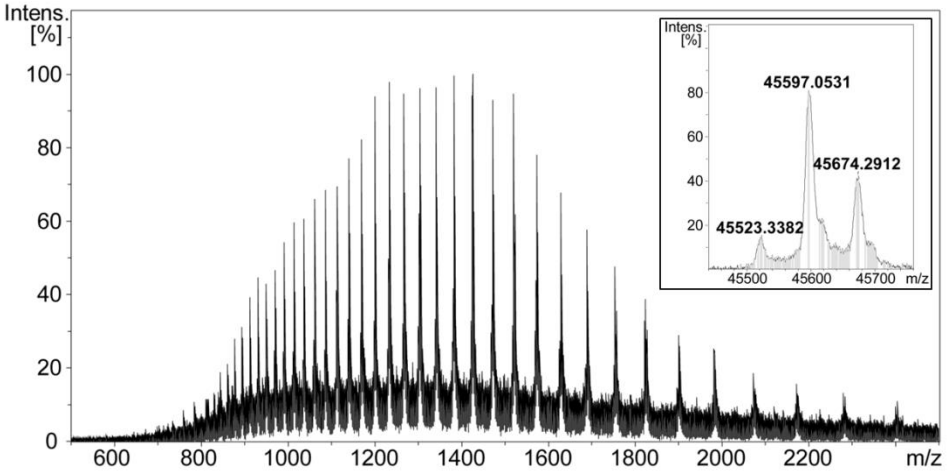


Figure C-15| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C E538A. Expected mass: 45,522.3 Da. Observed mass: 45,523.3 Da.

C.16 *At*PEX5(340-728) N601A: MS trace

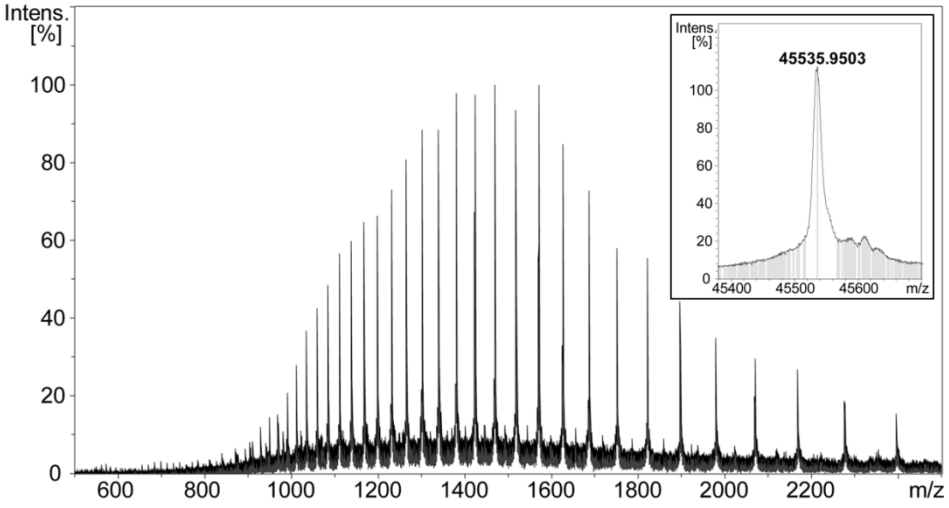


Figure C-16| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C N601A. Expected mass: 45,537.3 Da. Observed mass: 45,536.0 Da.

C.17 *At*PEX5(340-728) N601Q: MS trace

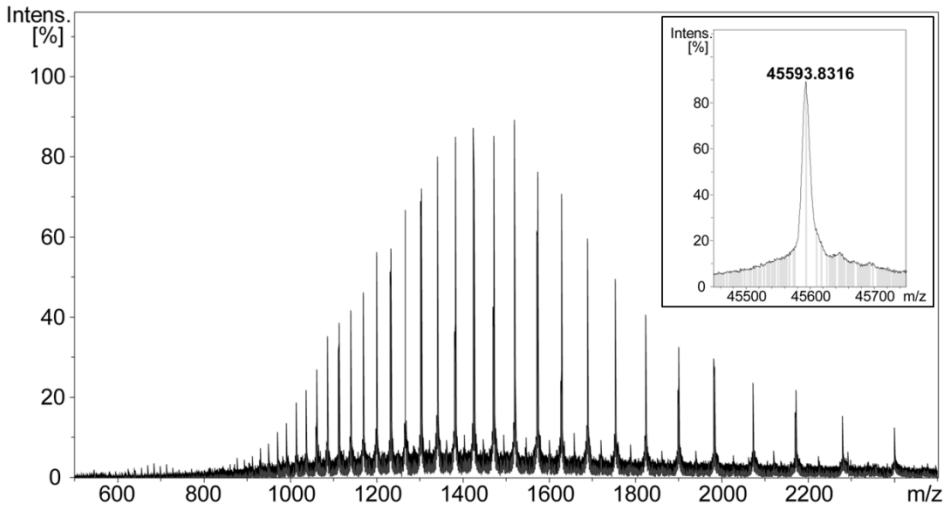


Figure C-17| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C N601Q. Expected mass: 45,594.3 Da. Observed mass: 45,593.8 Da.

C.18 *At*PEX5(340-728) F613A: MS trace

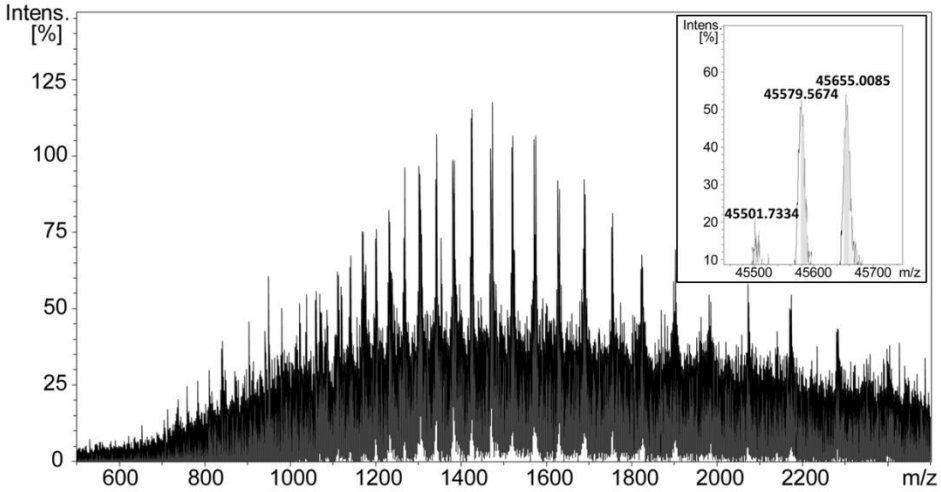


Figure C-18| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C F613A. Expected mass: 45,504.2 Da. Observed mass: 45,501.7 Da.

C.19 *At*PEX5(340-728) N628A: MS trace

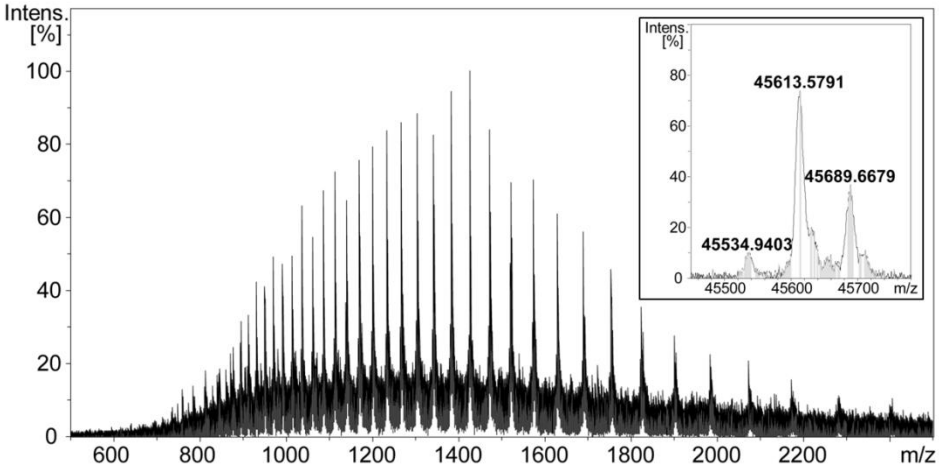


Figure C-19| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C N628A. Expected mass: 45,537.3 Da. Observed mass: 45,534.9 Da.

C.20 *At*PEX5(340-728) A632G: MS trace

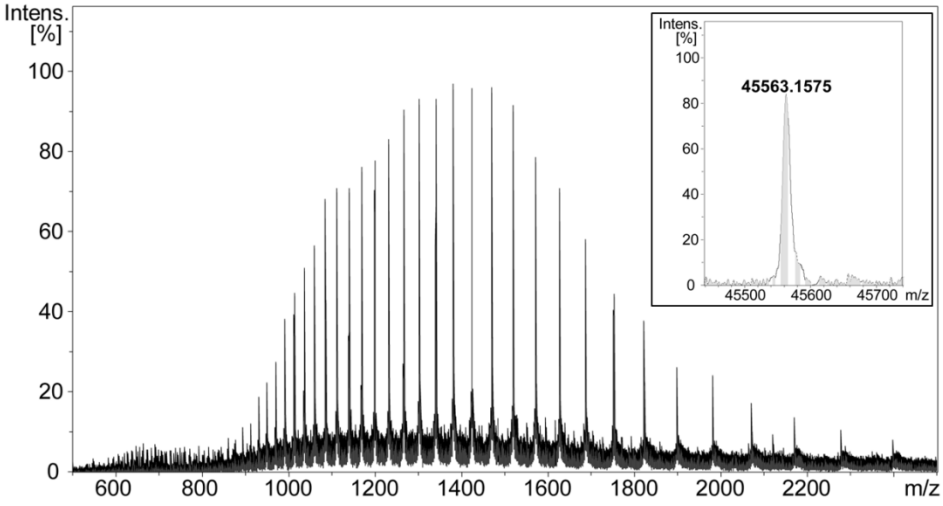


Figure C-20| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C A632G. Expected mass: 45,566.3 Da. Observed mass: 45,563.2 Da.

C.21 *At*PEX5(340-728) N636A: MS trace

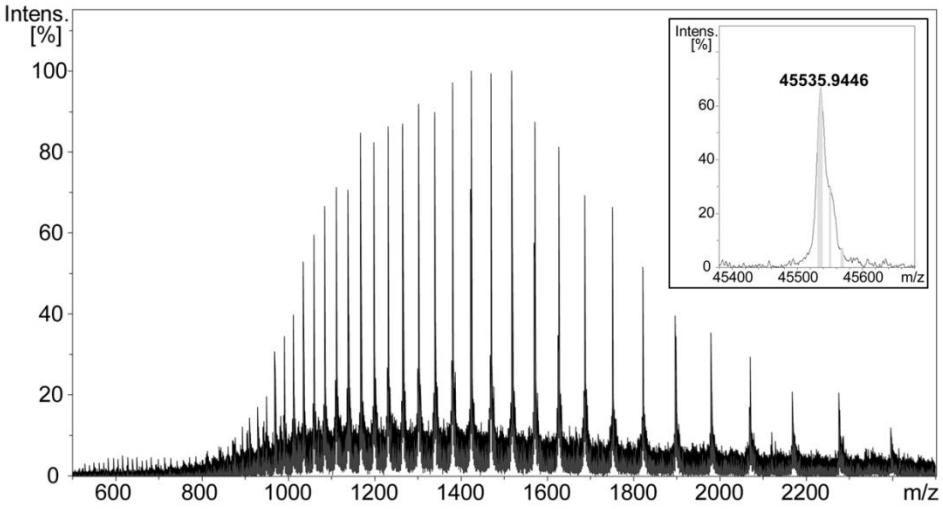


Figure C-21| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C N636A. Expected mass: 45,537.3 Da. Observed mass: 45,535.9 Da.

C.22 *At*PEX5(340-728) Y647F: MS trace

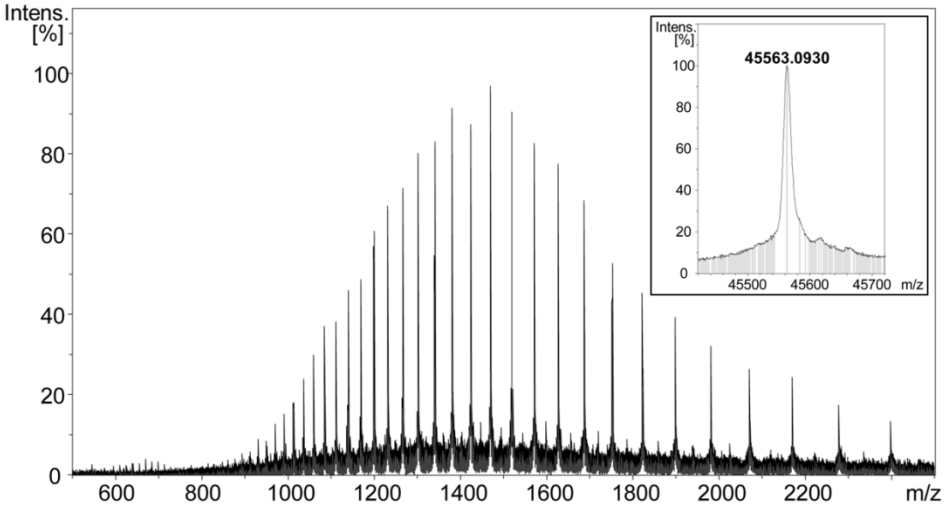


Figure C-22| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C Y647F. Expected mass: 45,564.3 Da. Observed mass: 45,563.1 Da.

C.23 *At*PEX5(340-728) R659A: MS trace

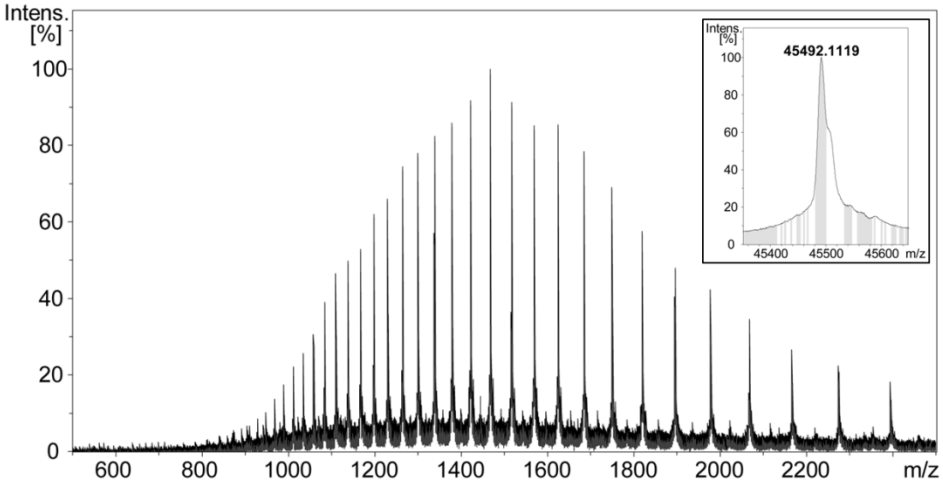


Figure C-23| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C R659A. Expected mass: 45,495.2 Da. Observed mass: 45,492.1 Da.

C.24 *At*PEX5(340-728) N663A: MS trace

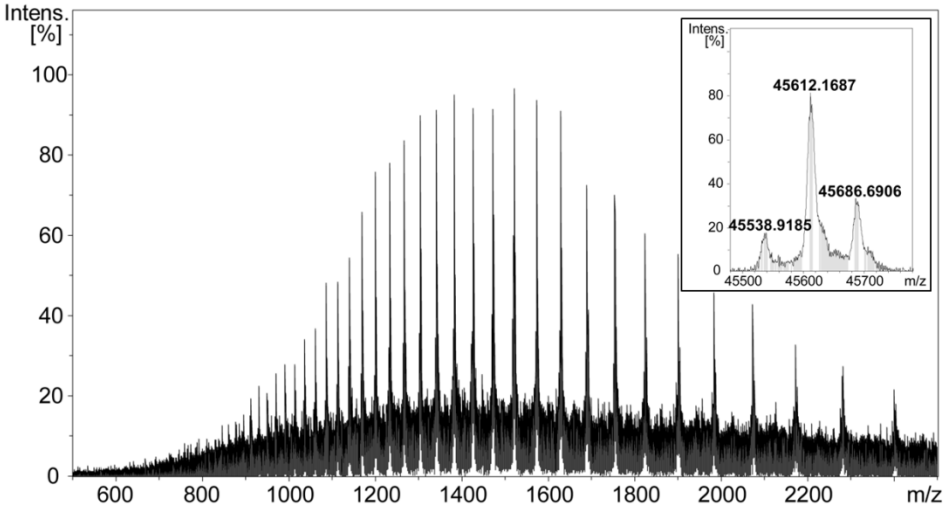


Figure C-24| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C N663A. Expected mass: 45,537.3 Da. Observed mass: 45,538.9 Da.

C.25 *At*PEX5(340-728) S667A: MS trace

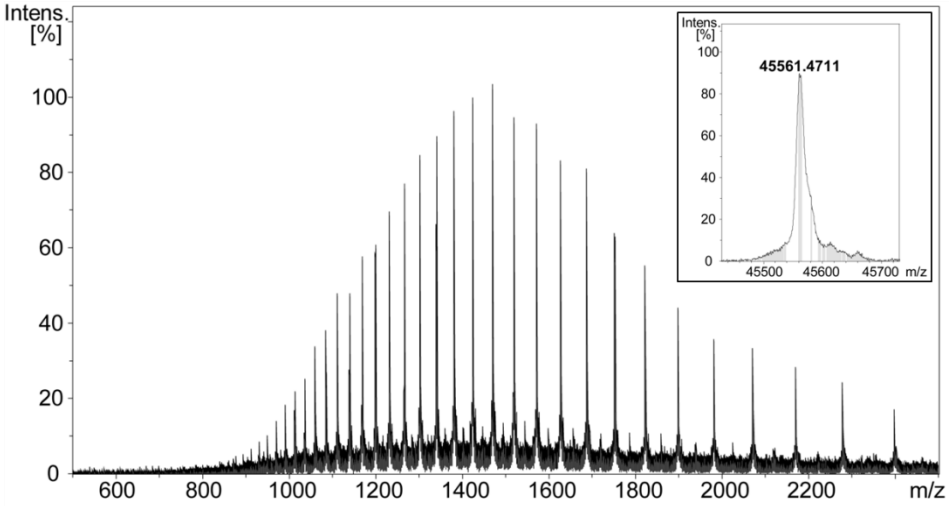


Figure C-25| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C S667A. Expected mass: 45,564.3 Da. Observed mass: 45,561.5 Da.

C.26 *At*PEX5(340-728) D505K-D507K: MS trace

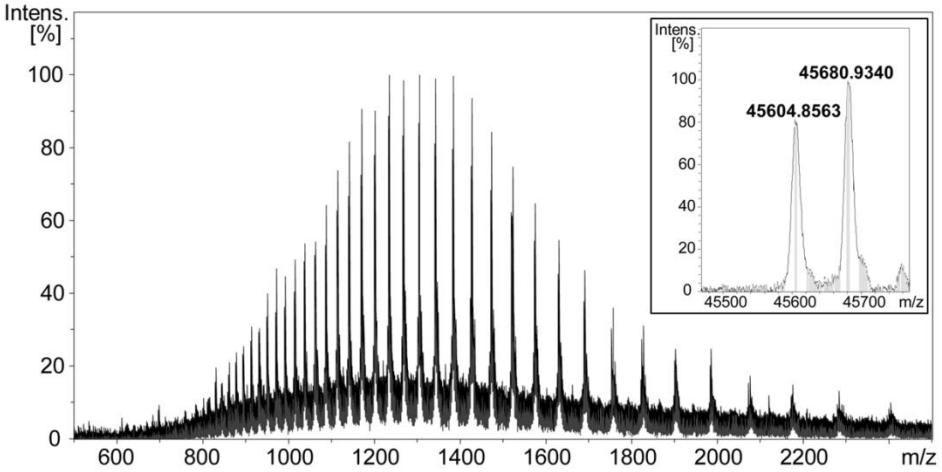


Figure C-26| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C D505K-D507K. Expected mass: 45,606.5 Da. Observed mass: 45,604.9 Da.

C.27 *At*PEX5(340-728) D505H-T536W: MS trace

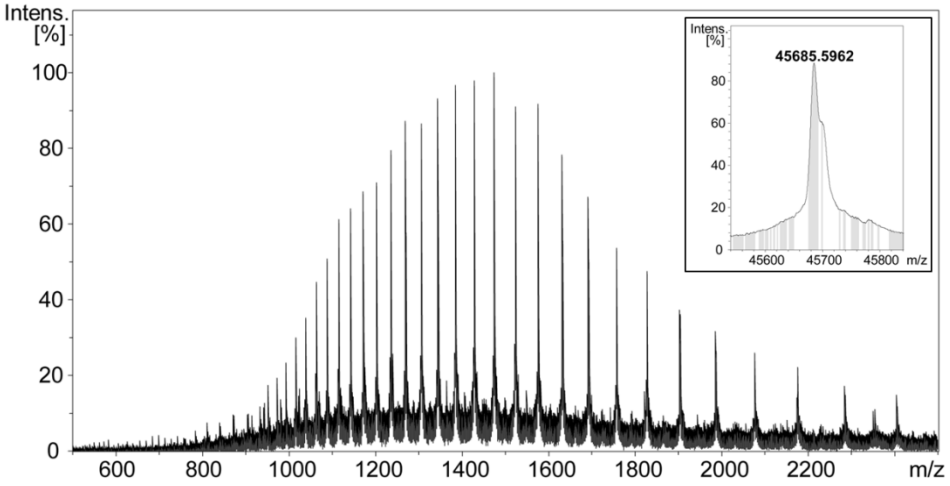


Figure C-27| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C D505H-T536W. Expected mass: 45,687.5 Da. Observed mass: 45,685.6 Da.

C.28 *At*PEX5(340-728) D505F-N601A: MS trace

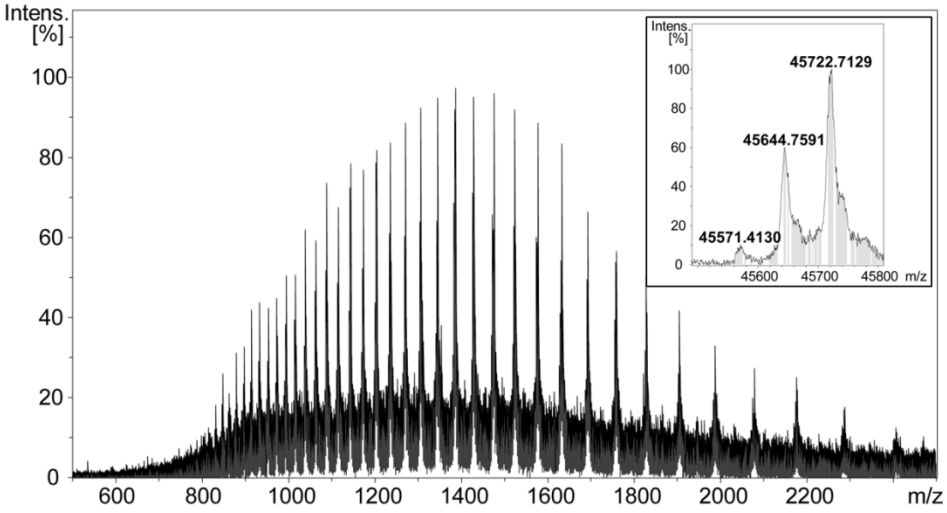


Figure C-28| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C D505F-N601A. Expected mass: 45,569.4 Da. Observed mass: 45,571.4 Da.

C.29 *At*PEX5(340-728) D505H-N601A: MS trace

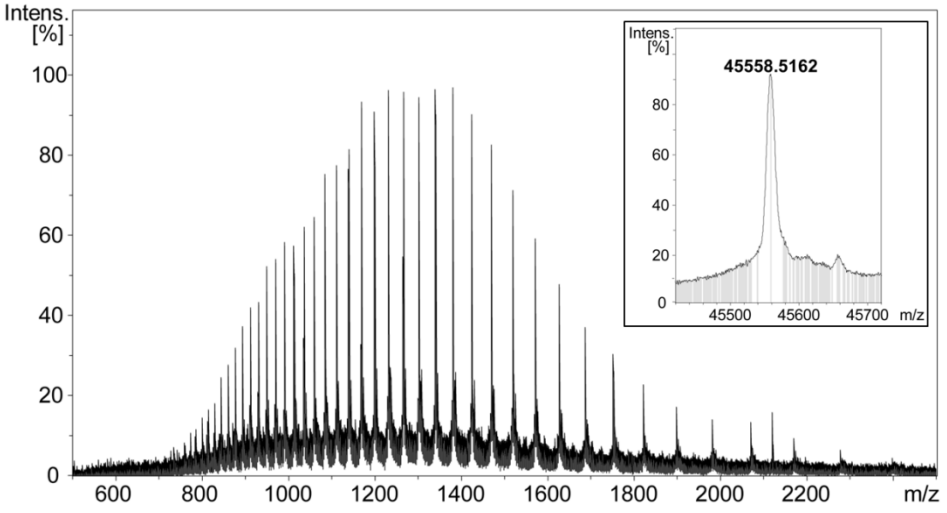


Figure C-29| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C D505H-N601A. Expected mass: 45,559.3 Da. Observed mass: 45,558.5 Da.

C.30 *At*PEX5(340-728) D505F-D507F-N601A: MS trace

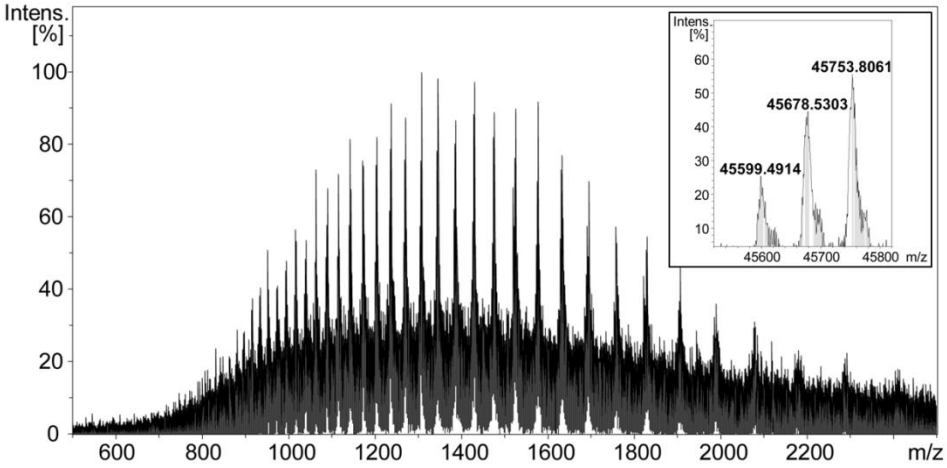


Figure C-30| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C D505F-D507F-N601A. Expected mass: 45,601.5 Da. Observed mass: 45,599.5 Da.

C.31 *At*PEX5(340-728) D505H-D507H-N601A: MS trace

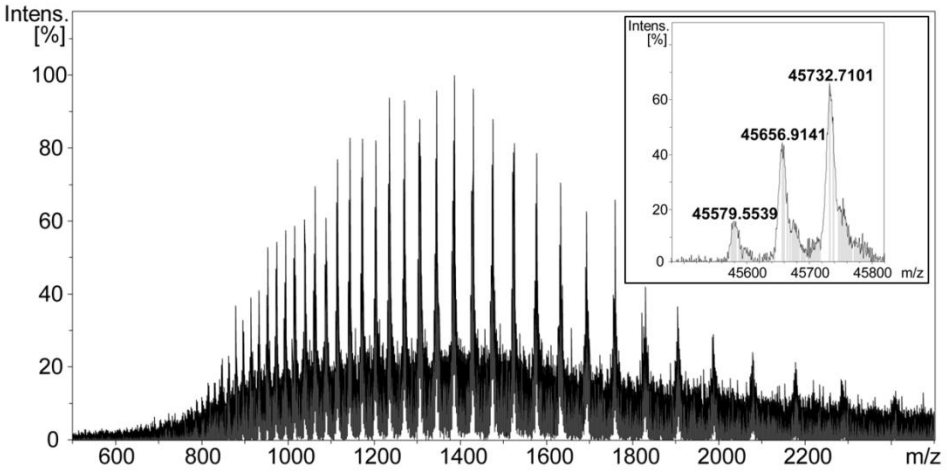


Figure C-31| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C D505H-D507H-N601A. Expected mass: 45,581.4 Da. Observed mass: 45,579.6 Da.

C.32 *At*PEX5(340-728) D505H-D507T-N601A: MS trace

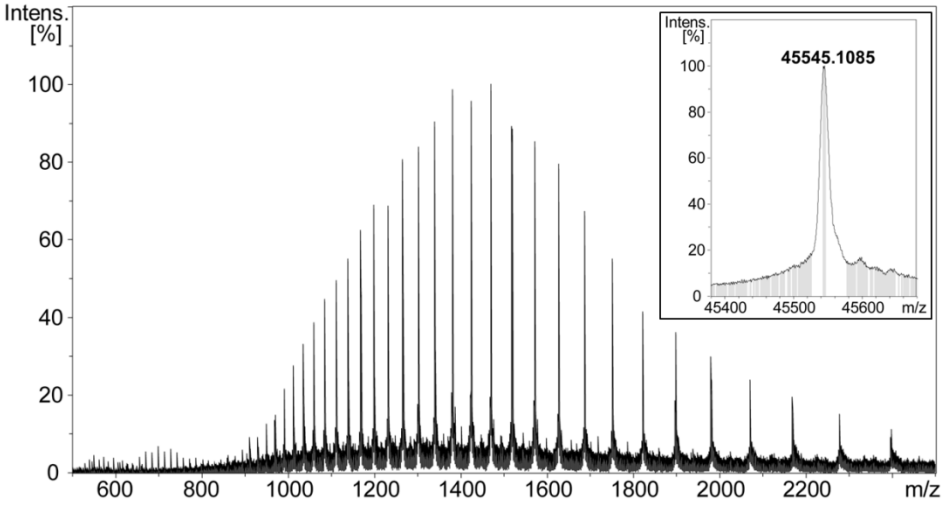


Figure C-32| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C D505H-D507T-N601A. Expected mass: 45,545.3 Da. Observed mass: 45,545.1 Da.

C.33 *At*PEX5(340-728) D505H-D507V-N601A: MS trace

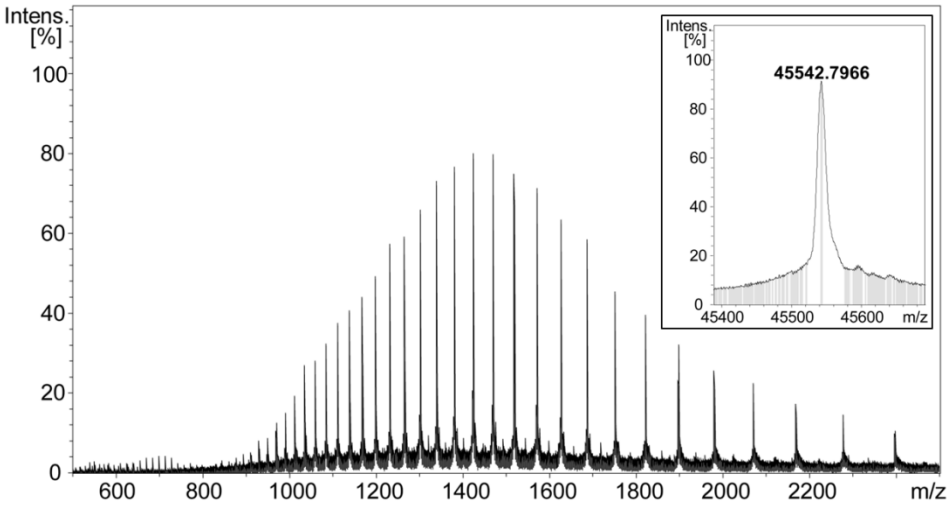


Figure C-33| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C D505H-D507V-N601A. Expected mass: 45,543.4 Da. Observed mass: 45,542.8 Da.

C.34 *At*PEX5(340-728) D505H-N601A-N636A: MS trace

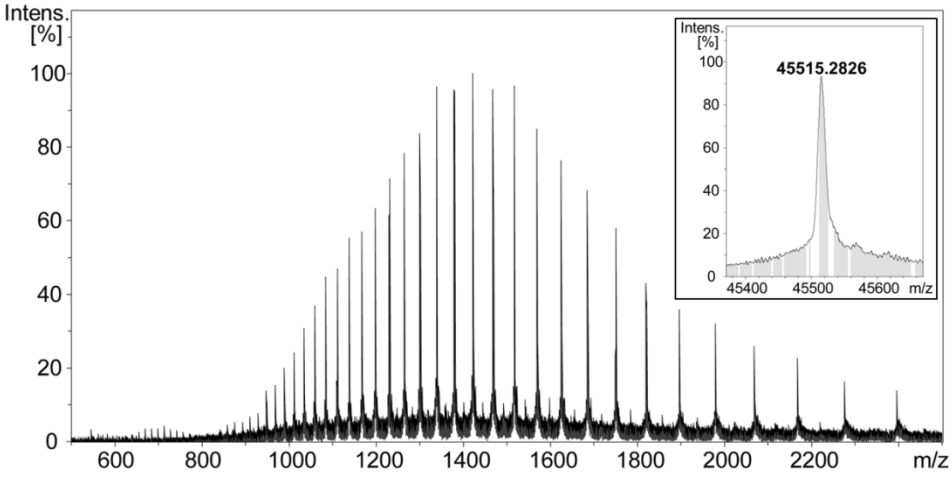


Figure C-34| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C D505H-N601A-N636A. Expected mass: 45,516.3 Da. Observed mass: 45,515.3 Da.

C.35 *At*PEX5(340-728) D505H-D507T-E538A-N601A: MS trace

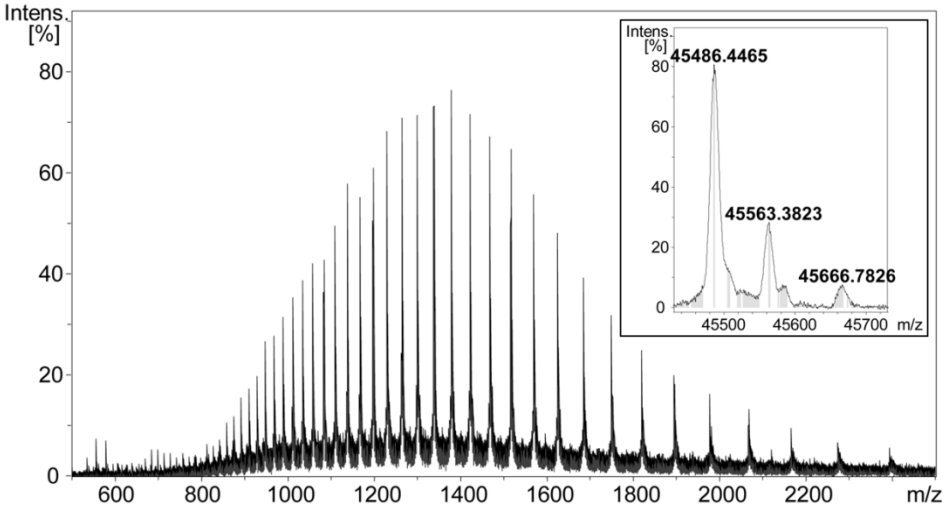


Figure C-35| Mass spectrum and deconvoluted mass data (inset) for *At*PEX5-C D505H-D507T-E538A-N601A. Expected mass: 45,487.3 Da. Observed mass: 45,486.4 Da.

C.36 Wild-type *At*PEX5(444-728): MS trace

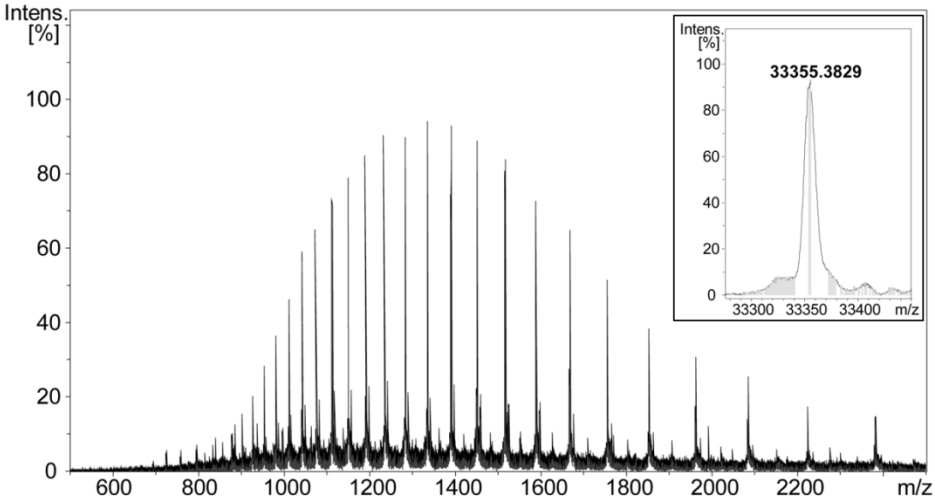


Figure C-36| Mass spectrum and deconvoluted mass data (inset) for truncated wild-type *At*PEX5-C. Expected mass: 33,356.4 Da. Observed mass: 33,355.4 Da.

Appendix D

Pull-down-LC-MS data analysis

D.1 Peptide characterisation tables – example data and processing

D.1.1 Retention time-sequence list

Peptide sequence	Retention time (s)
Dansyl-YQSAA	225.78
Dansyl-YQSAD	218.814
Dansyl-YQSAE	222.788
Dansyl-YQSAF	340.237
Dansyl-YQSAG	197.371
Dansyl-YQSAH	155.482
Dansyl-YQSAI	309.569
Dansyl-YQSAK	148.252
Dansyl-YQSAL	320.259
Dansyl-YQSAN	200.114
Dansyl-YQSAP	239.008
Dansyl-YQSAQ	206.582
Dansyl-YQSAR	150.74
Dansyl-YQSAS	206.348
Dansyl-YQSAT	220.295
Dansyl-YQSAV	272.168
Dansyl-YQSAW	348.428
Dansyl-YQSAY	267.664
Dansyl-YQSDA	216.347
Dansyl-YQSDD	215.371
Dansyl-YQSDE	212.358
Dansyl-YQSDF	325.287
Dansyl-YQSDG	210.883
Dansyl-YQSDH	155.782
Dansyl-YQSDI	295.383
Dansyl-YQSDK	148.802
Dansyl-YQSDL	304.108
Dansyl-YQSDN	202.905
Dansyl-YQSDP	233.074

Dansyl-YQSDQ	200.143
Dansyl-YQSDR	151.028
Dansyl-YQSDS	209.388
Dansyl-YQSDT	206.125
Dansyl-YQSDV	260
Dansyl-YQSDW	344.496
Dansyl-YQSDY	260.479
Dansyl-YQSEA	219.796
Dansyl-YQSED	212.58
Dansyl-YQSEE	215.558
Dansyl-YQSEF	324.529
Dansyl-YQSEH	153.986
Dansyl-YQSEI	298.598
Dansyl-YQSEK	146.755
Dansyl-YQSEL	307.298
Dansyl-YQSEN	195.875
Dansyl-YQSEP	228.039
Dansyl-YQSEQ	203.591
Dansyl-YQSER	150.74
Dansyl-YQSES	201.61
Dansyl-YQSET	218.799
Dansyl-YQSEV	262.445
Dansyl-YQSEW	340.203
Dansyl-YQSEY	260.933
Dansyl-YQSFA	337.516
Dansyl-YQSFE	329.786
Dansyl-YQSFF	391.339
Dansyl-YQSFG	323.293
Dansyl-YQSFH	220.108
Dansyl-YQSFK	205.897
Dansyl-YQSFL	388.625
Dansyl-YQSFN	300.119
Dansyl-YQSFP	356.436
Dansyl-YQSFQ	307.348
Dansyl-YQSFR	215.599
Dansyl-YQSFS	313.8
Dansyl-YQSFT	330.036
Dansyl-YQSFV	373.139
Dansyl-YQSFW	390.869
Dansyl-YQSFY	364.191
Dansyl-YQSGA	211.112
Dansyl-YQSGD	204.65
Dansyl-YQSGE	207.621
Dansyl-YQSGG	198.666
Dansyl-YQSGH	149.301

Dansyl-YQSGI	294.387
Dansyl-YQSGK	145.56
Dansyl-YQSGL	303.36
Dansyl-YQSGN	194.427
Dansyl-YQSGP	229.832
Dansyl-YQSGQ	195.156
Dansyl-YQSGR	149.282
Dansyl-YQSGS	198.915
Dansyl-YQSGT	208.12
Dansyl-YQSGV	257.258
Dansyl-YQSGW	339.759
Dansyl-YQSGY	255.992
Dansyl-YQSHA	153.521
Dansyl-YQSHE	153.272
Dansyl-YQSHG	153.289
Dansyl-YQSHH	130.35
Dansyl-YQSHI	191.186
Dansyl-YQSHK	125.364
Dansyl-YQSHL	199.8795
Dansyl-YQSHN	149.301
Dansyl-YQSHP	162.514
Dansyl-YQSHQ	150.279
Dansyl-YQSHR	130.581
Dansyl-YQSHS	150.547
Dansyl-YQSHT	153.272
Dansyl-YQSHV	169.02
Dansyl-YQSHW	234.55
Dansyl-YQSHY	171.472
Dansyl-YQSIA	307.597
Dansyl-YQSID	292.892
Dansyl-YQSIE	298.373
Dansyl-YQSIF	387.848
Dansyl-YQSIG	291.148
Dansyl-YQSII	380.12
Dansyl-YQSIK	183.207
Dansyl-YQSIL	382.392
Dansyl-YQSIN	267.723
Dansyl-YQSIP	323.293
Dansyl-YQSIQ	276.186
Dansyl-YQSIR	188.924
Dansyl-YQSIS	281.926
Dansyl-YQSIT	297.376
Dansyl-YQSIV	357.682
Dansyl-YQSIW	387.877
Dansyl-YQSIY	343.748

Dansyl-YQSKA	148.784
Dansyl-YQSKE	148.535
Dansyl-YQSKF	203.404
Dansyl-YQSKH	127.607
Dansyl-YQSKK	121.375
Dansyl-YQSKL	190.669
Dansyl-YQSKN	143.815
Dansyl-YQSKP	153.788
Dansyl-YQSKQ	145.543
Dansyl-YQSKR	127.339
Dansyl-YQSKS	145.061
Dansyl-YQSKT	148.036
Dansyl-YQSKV	163.511
Dansyl-YQSKW	223.079
Dansyl-YQSKY	164.741
Dansyl-YQSLD	306.078
Dansyl-YQSLE	313.778
Dansyl-YQSLF	390.605
Dansyl-YQSLG	148.75
Dansyl-YQSLI	385.118
Dansyl-YQSLK	190.889
Dansyl-YQSLL	385.795
Dansyl-YQSLN	280.147
Dansyl-YQSLP	342.483
Dansyl-YQSLQ	290.6
Dansyl-YQSLR	201.846
Dansyl-YQSLS	294.857
Dansyl-YQSLT	312.531
Dansyl-YQSLV	367.167
Dansyl-YQSLW	389.033
Dansyl-YQSLY	353.66
Dansyl-YQSNA	202.635
Dansyl-YQSND	197.918
Dansyl-YQSNF	305.602
Dansyl-YQSNH	151.046
Dansyl-YQSNI	276.195
Dansyl-YQSNL	285.16
Dansyl-YQSNN	188.444
Dansyl-YQSNP	212.878
Dansyl-YQSNQ	189.671
Dansyl-YQSNR	147.537
Dansyl-YQSNS	193.43
Dansyl-YQSNT	198.646
Dansyl-YQSNV	242.548
Dansyl-YQSNW	329.038

Dansyl-YQSNY	245.021
Dansyl-YQSPA	234.052
Dansyl-YQSPE	230.559
Dansyl-YQSPG	221.604
Dansyl-YQSPL	331.781
Dansyl-YQSPP	248.282
Dansyl-YQSPQ	214.353
Dansyl-YQSPR	157.012
Dansyl-YQSPS	219.609
Dansyl-YQSPT	228.066
Dansyl-YQSPV	286.91
Dansyl-YQSPW	358.456
Dansyl-YQSPY	282.418
Dansyl-YQSQD	202.109
Dansyl-YQSQF	304.583
Dansyl-YQSQG	159.222
Dansyl-YQSQH	150.994
Dansyl-YQSQI	278.9
Dansyl-YQSQK	144.012
Dansyl-YQSQL	288.605
Dansyl-YQSQN	187.646
Dansyl-YQSQP	388.36
Dansyl-YQSQQ	193.119
Dansyl-YQSQR	147.499
Dansyl-YQSQS	191.886
Dansyl-YQSQT	205.834
Dansyl-YQSQV	244.244
Dansyl-YQSQW	325.246
Dansyl-YQSQY	247.221
Dansyl-YQSRD	151.492
Dansyl-YQSRF	208.092
Dansyl-YQSRG	146.755
Dansyl-YQSRH	129.552
Dansyl-YQSRI	187.397
Dansyl-YQSRK	124.067
Dansyl-YQSRL	196.36
Dansyl-YQSRN	146.256
Dansyl-YQSRP	154.484
Dansyl-YQSRR	128.054
Dansyl-YQSRS	147.503
Dansyl-YQSRT	150.989
Dansyl-YQSRV	166.951
Dansyl-YQSRW	228.522
Dansyl-YQSRY	168.691
Dansyl-YQSSD	204.401

Dansyl-YQSSE	211.112
Dansyl-YQSSF	316.315
Dansyl-YQSSH	153.04
Dansyl-YQSSI	287.907
Dansyl-YQSSL	296.877
Dansyl-YQSSQ	194.408
Dansyl-YQSSR	149.033
Dansyl-YQSSS	198.167
Dansyl-YQSST	203.882
Dansyl-YQSSV	252.272
Dansyl-YQSSW	334.274
Dansyl-YQSSY	252.252
Dansyl-YQSTD	216.32
Dansyl-YQSTE	213.813
Dansyl-YQSTF	327.022
Dansyl-YQSTH	155.482
Dansyl-YQSTI	298.349
Dansyl-YQSTK	147.753
Dansyl-YQSTL	306.55
Dansyl-YQSTN	199.117
Dansyl-YQSTP	232.277
Dansyl-YQSTQ	202.344
Dansyl-YQSTS	204.353
Dansyl-YQSTT	215.558
Dansyl-YQSTV	261.946
Dansyl-YQSTW	341.2
Dansyl-YQSTY	260.684
Dansyl-YQSVA	268.956
Dansyl-YQSVE	261.975
Dansyl-YQSVF	375.134
Dansyl-YQSVG	253.518
Dansyl-YQSVH	170.492
Dansyl-YQSVL	361.448
Dansyl-YQSVN	235.318
Dansyl-YQSVQ	241.281
Dansyl-YQSVR	167.484
Dansyl-YQSVS	246.288
Dansyl-YQSVT	260.728
Dansyl-YQSVV	326.034
Dansyl-YQSVW	376.907
Dansyl-YQSVY	313.082
Dansyl-YQSWD	336.248
Dansyl-YQSWE	336.464
Dansyl-YQSWF	392.35
Dansyl-YQSWG	187.397

Dansyl-YQSWH	231.03
Dansyl-YQSWI	388.61
Dansyl-YQSWK	215.323
Dansyl-YQSWN	313.059
Dansyl-YQSWP	390.605
Dansyl-YQSWQ	317.268
Dansyl-YQSWS	325.775
Dansyl-YQSWT	339.206
Dansyl-YQSWV	378.886
Dansyl-YQSWW	391.275
Dansyl-YQSWY	369.604
Dansyl-YQSYA	270.656
Dansyl-YQSYD	256.71
Dansyl-YQSYF	366.918
Dansyl-YQSYG	250.228
Dansyl-YQSYH	171.938
Dansyl-YQSYI	348.718
Dansyl-YQSYK	162.214
Dansyl-YQSYN	234.769
Dansyl-YQSYP	284.136
Dansyl-YQSYQ	242.483
Dansyl-YQSYS	244.743
Dansyl-YQSYT	263.675
Dansyl-YQSYV	316.549
Dansyl-YQSYY	306.799

D.1.2 Peptides with identical exact mass and retention times to 45 peptides in the retention time-sequence list

In sequence-retention time list	Same retention time and mass
Dansyl-YQSAD	Dansyl-YQSEG
Dansyl-YQSAL	Dansyl-YQSLA
Dansyl-YQSAQ	Dansyl-YQSQA
Dansyl-YQSAR	Dansyl-YQSRA
Dansyl-YQSAS	Dansyl-YQSTG
Dansyl-YQSAT	Dansyl-YQSTA
Dansyl-YQSAW	Dansyl-YQSWA
Dansyl-YQSDF	Dansyl-YQSFD
Dansyl-YQSDH	Dansyl-YQSHD
Dansyl-YQSDK	Dansyl-YQSKD
Dansyl-YQSDP	Dansyl-YQSPD
Dansyl-YQSDV	Dansyl-YQSVD
Dansyl-YQSEQ	Dansyl-YQSQE
Dansyl-YQSEY	Dansyl-YQSYE
Dansyl-YQSFP	Dansyl-YQSPF

Dansyl-YQSFG	Dansyl-YQSGF
Dansyl-YQSGK	Dansyl-YQSKG
Dansyl-YQSGS	Dansyl-YQSSG
Dansyl-YQSGT	Dansyl-YQSSA
Dansyl-YQSFH	Dansyl-YQSHF
Dansyl-YQSHI	Dansyl-YQSIH
Dansyl-YQSHL	Dansyl-YQSLH
Dansyl-YQSHP	Dansyl-YQSPH
Dansyl-YQSIF	Dansyl-YQSFI
Dansyl-YQSIK	Dansyl-YQSKI
Dansyl-YQSIP	Dansyl-YQSPI
Dansyl-YQSIV	Dansyl-YQSVI
Dansyl-YQSKP	Dansyl-YQSPK
Dansyl-YQSKS	Dansyl-YQSSK
Dansyl-YQSKV	Dansyl-YQSVK
Dansyl-YQSLW	Dansyl-YQSWL
Dansyl-YQSLY	Dansyl-YQSYL
Dansyl-YQSDQ	Dansyl-YQSNE
Dansyl-YQSGN	Dansyl-YQSNG
Dansyl-YQSKN	Dansyl-YQSNK
Dansyl-YQSNP	Dansyl-YQSPN
Dansyl-YQSNS	Dansyl-YQSSN
Dansyl-YQSPS	Dansyl-YQSSP
Dansyl-YQSPV	Dansyl-YQSVP
Dansyl-YQSER	Dansyl-YQSRE
Dansyl-YQSQR	Dansyl-YQSRQ
Dansyl-YQSRT	Dansyl-YQSTR
Dansyl-YQSRW	Dansyl-YQSWR
Dansyl-YQSRY	Dansyl-YQSYR
Dansyl-YQSWY	Dansyl-YQSYW

D.1.3 Example data processing

The following example data shown are for D505F-D507F-N601A.

Name	Chromatogram	ReT (s)	Area
Cmpd 1	XYQSHK & XYQSKH	128.78	5622.347
Cmpd 2	XYQSRH & XYQSHR	130.275	8791.736
Cmpd 3	XYQSHH	132.021	3972.709
Cmpd 4	XYQSNW & XYQSHY & XYQSWN & XYQSYH	171.415	1427.574
Cmpd 5	XYQSDV & XYQSVD	182.386	604.4569
Cmpd 6	XYQSRE & XYQSER	220.038	554.2702
Cmpd 7	XYQSHF & XYQSFH	220.038	2091.318

Cmpd 8	XYQSHW & XYQSWH	231.757	2027.002
Cmpd 9	XYQSGG	282.869	364.8235
Cmpd 10	XYQSNG & XYQSGN	318.277	347.5375
Cmpd 11	XYQSIF & XYQSLF & XYQSFI & XYQSFL	389.335	1146.977
Cmpd 12	XYQSIW & XYQSLW & XYQSWI & XYQSWL	390.083	912.2089
Cmpd 13	XYQSVV	402.8	1350.508
Cmpd 14	XYQSRH & XYQSHR	406.29	2993.248

A drift time of -1 seconds was applied to the retention times as this produced the lowest average difference between the retention time master list and the raw data.

Rows with a light red fill show the peptides which were too far outside the retention time identifier for the peptide sequences shown.

Highlighted in yellow are the peptide sequences that were identified as having a retention time closest to that in the master list. Cells where two peptide sequences are highlighted in yellow show that these two peptide sequences have identical retention times so I have plotted both peptide sequences in the final plot. A plot with an altered scale is shown in **Figure D-1**

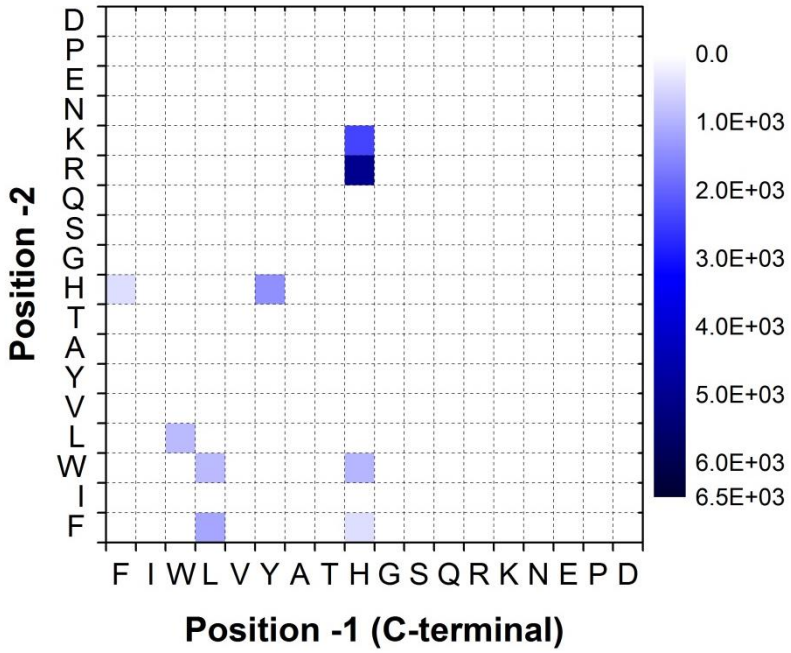


Figure D-1| Pull-down-LC-MS screen result for *At*PEX5-C D505F-D507F-N601A. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS. A smaller scale bar than screen results in Appendix E are used to show the less abundant peptides pulled down.

D.2 Code for adding EICs for each peptide sequence after mass spectrometry, and exporting data

```
'*********************************
'All code in this section (D.2) was written by Stuart Warriner
'*********************************

Dim Chrom, f , i,rstart, rstop
rstart = 100 ' retention time window start
rstop = 480 ' retention time window stop for peak picking
rstart = rstart/60
rstop =  rstop/60
Analysis.RecalibrateAutomatically
Analysis.ClearResults
Analysis.Chromatograms.Clear
Analysis.Compounds.Clear
ReadData
Analysis.Chromatograms.Smooth
For Each Chrom in Analysis.Chromatograms
Chrom.AddRangeSelection rstart, rstop, 0, 0
Chrom.FindCompounds
Next
'Analysis.Chromatograms.FindCompounds
Analysis.Save

Sub ReadData
Dim sequenceformula, listhandler, f
Set listhandler = New clsCreateFormulaList
listhandler.FileInfo "D:\targetlists\peptideLibrary_Dan_full.csv",1,2
For Each f In listhandler.ForumlaList_Dict.Keys
AddEIC f, listhandler.ForumlaList_Dict.Item(f)
Next
Set listhandler = Nothing
End Sub

'Sub AddEIC(formula,text)
'End Sub


Sub AddEIC(formula, text)
Dim EIC
 Set EIC = CreateObject("DataAnalysis.EICChromatogramDefinition")
EIC.MSFilter.Type = daMSFilterAllMS
EIC.Polarity = daPositive
EIC.WidthLeft = 0.008
EIC.WidthRight = 0.008
EIC.Formula = formula
EIC.Adducts = "M+H"
EIC.Charges = MaxCharge(text)
Analysis.Chromatograms.AddChromatogram EIC
Analysis.Chromatograms(Analysis.Chromatograms.Count).Name_ = text
End Sub


Function MaxCharge(sequence)
Dim seq, arr, count, seq_startlength
arr = Split(sequence, " & ")
seq=arr(0)
seq_startlength = Len(seq)
seq = Replace(seq,"H","")
seq = Replace(seq,"K","")
```

```
seq = Replace(seq,"R","")
count = seq_startlength - Len(seq)
If count >1 Then
MaxCharge = "1-2"
Else
MaxCharge = "1"
End If
End Function


Class clsCreateFormulaList
Private filepath, fso, sequence_data, formula_col, sequence_col

Sub Class_Initialize
    Set fso = createobject("scripting.filesystemobject")
    Set sequence_data = createobject("Scripting.Dictionary")
    sequence_data.CompareMode = vbTextCompare
End Sub

Sub Class_Terminate
    Set fso  = Nothing
    Set sequence_data = Nothing
End Sub


 Public Property Get ForumlaList_Dict
  Set  ForumlaList_Dict = sequence_data
 End Property


 Public Sub FileInfo (setfilepath, sequencecolumn, formulacolumn)
 filepath = setfilepath
 formula_col = formulacolumn - 1 ' array numbering
 sequence_col= sequencecolumn - 1
PopulateLists
 End Sub

 Private Sub PopulateLists
 Dim i, j, imnputfile, fields
 Set inputfile = fso.OpentextFile(filepath,1)
 inputfile.ReadLine
 Do While Not(inputfile.AtEndofStream)
  fields = Split(inputfile.ReadLine,",")
If sequence_data.Exists(fields(formula_col)) Then
 sequence_data.Item(fields(formula_col)) =
sequence_data.Item(fields(formula_col)) & " & " & fields(sequence_col)
 Else
 sequence_data.Add fields(formula_col), fields(sequence_col)
 End If
 'sgBox(fields(sequence_col))

 Loop


 End Sub


End Class

Form.Close
```

```
'*********************************
'This code was used to export data
'ExportCmpds
'ExportPdf
'*********************************

Sub ExportPdf
Dim filepath
filepath = "D:\Laura_Reports\Reports\" & strip(Analysis.Name) & ".pdf"
Analysis.PrintToPDF "Leeds_Laura", filepath
End Sub


Sub ExportCmpds
Dim cmpd, fso, filepath, outfile, line
Set fso = createobject("scripting.filesystemobject")
filepath = "D:\Laura_Reports\Lists\" & strip(Analysis.Name) & ".csv"
Set outfile = fso.CreatetextFile(filepath,True)
line = "Name, Chromatogram,RT,Area"
outfile.WriteLine(line)
For each cmpd In Analysis.Compounds
line = stripRT(cmpd.Name) & "," & cmpd.Chromatogram & "," &
cmpd.RetentionTime & "," & cmpd.Area
outfile.WriteLine(line)
Next
outfile.Close
Set outfile = Nothing
Set fso = Nothing
End Sub


Function stripRT(str)
dim newstr
newstr = Split(str,",")
stripRT = Trim(newstr(0))
End Function


Function strip(str)
Dim newstr
newstr = Left(str,Len(str)-2)
strip = newstr
End Function

Form.Close


'*********************************
'This code was used to add EICs for tandem MS data
'*********************************

Dim Chrom, f , i,rstart, rstop, AAS
rstart = 100 ' retention time window start
rstop = 480 ' retention time window stop for peak picking
rstart = rstart/60
rstop =  rstop/60
SetAAS
'Analysis.RecalibrateAutomatically
Analysis.ClearResults
Analysis.Chromatograms.Clear
Analysis.Compounds.Clear
ReadData
```

```
Analysis.Chromatograms.Smooth
For Each Chrom in Analysis.Chromatograms
Chrom.AddRangeSelection rstart, rstop, 0, 0
'Chrom.FindCompounds
Next
'Analysis.Chromatograms.FindCompounds
Analysis.Save

Sub ReadData
Dim sequenceformula, listhandler, seq, p,t
Set listhandler = New clsCreateFormulaList
listhandler.FileInfo "D:\targetlists\peptideLibrary_Dan_1_1.csv",1,2
For Each seq In listhandler.ForumlaList_Dict.Keys
p = GetMass(listhandler.ForumlaList_Dict.Item(seq))
t = GetTarget(seq)
AddEIC p, t,seq
Next
Set listhandler = Nothing
End Sub

'Sub AddEIC(formula,text)
'End Sub
GetTarget(seq)


Function GetMass(fm)
dim SF
set SF = CreateObject("DataAnalysis.SumFormula")
SF.Formula = fm
SF.Add("H2")
GetMass = SF.MonoIsotopicMass(+2)
End Function

Sub AddEIC(prec, target, text)
Dim EIC
Set EIC = CreateObject("DataAnalysis.EICChromatogramDefinition")
EIC.MSFilter.Type = daMSFilterMSMS
EIC.MSFilter.FragmentationPath = prec
EIC.Polarity = daPositive
EIC.WidthLeft = 0.008
EIC.WidthRight = 0.008
EIC.Range = target
'MsgBox (prec & " " & target & " " & text)
'EIC.Adducts = "M+H"
'EIC.Charges = MaxCharge(text)
Analysis.Chromatograms.AddChromatogram EIC
Analysis.Chromatograms(Analysis.Chromatograms.Count).Name_ = text
End Sub


Function MaxCharge(sequence)
Dim seq, arr, count, seq_startlength
arr = Split(sequence, " & ")
seq=arr(0)
seq_startlength = Len(seq)
seq = Replace(seq,"H","")
seq = Replace(seq,"K","")
seq = Replace(seq,"R","")
count = seq_startlength - Len(seq)
If count >1 Then
MaxCharge = "1-2"
Else
```

```
MaxCharge = "1"
End If
End Function



Sub SetAAS
dim aa, mass
Set AAS = createobject("Scripting.Dictionary")
AAS.CompareMode = vbTextCompare
aa = Split("A,C,D,E,F,G,H,I,K,L,M,N,P,Q,R,S,T,V,W,Y",",")
mass =
Array(90.054955,122.027026,134.044784,148.060434,166.086255,76.039305,
156.076753,132.101905,147.112804,132.101905,150.058326,133.060769,116.
070605,147.076419,175.118952,106.04987,120.06552,118.086255,205.097154
,182.08117)
For i = 0 to Ubound(aa)
AAS.Add aa(i), mass(i)
Next
End Sub

Function GetTarget(seq)
Dim txt
txt = Right(seq,1)
GetTarget = AAS.Item(txt)
End Function


Class clsCreateFormulaList
Private filepath, fso, sequence_data, formula_col, sequence_col

Sub Class_Initialize
    Set fso = createobject("scripting.filesystemobject")
    Set sequence_data = createobject("Scripting.Dictionary")
    sequence_data.CompareMode = vbTextCompare
End Sub

Sub Class_Terminate
    Set fso  = Nothing
    Set sequence_data = Nothing
End Sub


 Public Property Get ForumlaList_Dict
  Set  ForumlaList_Dict = sequence_data
 End Property


 Public Sub FileInfo (setfilepath, sequencecolumn, formulacolumn)
 filepath = setfilepath
 formula_col = formulacolumn - 1 ' array numbering
 sequence_col= sequencecolumn - 1
PopulateLists
 End Sub

 Private Sub PopulateLists
 Dim i, j, imnputfile, fields
 Set inputfile = fso.OpentextFile(filepath,1)
 inputfile.ReadLine
 Do While Not(inputfile.AtEndofStream)
  fields = Split(inputfile.ReadLine,",")
  If Not(sequence_data.Exists(fields(sequence_col))) Then
```

```
 sequence_data.Add fields(sequence_col),fields(formula_col)
 End If
 'sgBox(fields(sequence_col))
  Loop


 End Sub


End Class

Form.Close
```

D.3 Code for data-processing macros

D.3.1 Generation of 'drifted' retention times and matching of these to the retention times of peptides in the 'seq-RT' master list

```
'*********************************
' Drive the 'Master' Sub, iterating through a drift in retention time
from -5 to +5 seconds
' Written by Stuart Warriner
'*********************************

Sub drive()
Dim i As Integer
For i = 0 To 10
Master i
Next

End Sub

'*********************************
' Add a drift time error on to each of the retention times in the
exported '.csv' file and print these 'drifted' retention times onto
the worksheet
' Written by Stuart Warriner
'*********************************

Sub Master(x As Integer)

Dim RtList As Dictionary
Dim PeakChroms As Dictionary
Dim PeakAreas As Dictionary
Dim PeakRT As Dictionary
Dim PeakBestFit As Dictionary
Dim PeakBestOff As Dictionary
Dim DataSheet As Worksheet
Dim j, k, l
Set RtList = New Scripting.Dictionary
RtList.CompareMode = TextCompare
Set PeakChroms = New Scripting.Dictionary
PeakChroms.CompareMode = TextCompare
Set PeakAreas = New Scripting.Dictionary
PeakAreas.CompareMode = TextCompare
Set PeakRT = New Scripting.Dictionary
PeakRT.CompareMode = TextCompare
Set PeakBestFit = New Scripting.Dictionary
Set PeakBestOff = New Scripting.Dictionary
```

```
Set DataSheet = Worksheets(44) ' This is the Worksheet number that
this particular '.csv' file was copied and pasted into
drift = -5 + x ' Add -5 to 0, then 1, then ..., then 10 (this is the
line being driven by the 'drive' Sub)
      For j = 2 To 280
            If Not RtList.Exists(Worksheets(1).Cells(j, 1).Value) Then
            RtList.Add Trim(Worksheets(1).Cells(j, 1).Value),
            Worksheets(1).Cells(j, 2).Value
            End If
      Next

      For j = 2 To LastRow(DataSheet)
      PeakChroms.Add DataSheet.Cells(j, 1), DataSheet.Cells(j, 2) ' A
      'dictionary' is created, containing compound name (e.g. cmpd1)
      against peptide sequence name (e.g. XYQSKL)
      PeakAreas.Add DataSheet.Cells(j, 1), DataSheet.Cells(j, 4) ' A
      dictionary is created, containing compound name against EIC area
      for each peak
      PeakRT.Add DataSheet.Cells(j, 1), DataSheet.Cells(j, 3) ' A
      dictionary is created, containing compound name against
      retention time for each peak
      Next


      For k = 0 To PeakChroms.Count - 1
      l = PeakChroms.Keys(k)
      Chrom = Split(PeakChroms.Items(k), "&")
      best = 5
      bestSeq = ""
            For j = 0 To UBound(Chrom)
            off = PeakRT.Items(k) - RtList.Item(Trim(Chrom(j)))
            off = off + drift
                  If Abs(off) < Abs(best) Then
                  best = off
                  bestSeq = Trim(Chrom(j)) ' For each peak, look for
                  the peptide sequence name (within the possible
                  peptides defined in 'Chromatogram' column) that has
                  the closest matching retention time (in the seq-RT
                  master list) for each 'drift' time applied to the
                  retention time of the peak
                  End If
            Next
      PeakBestFit.Add l, bestSeq
      PeakBestOff.Add l, best
            If Abs(best) < 5 Then
            DataSheet.Cells(2 + k, 6 + 2 * x).Value = bestSeq
            DataSheet.Cells(2 + k, 7 + 2 * x).Value = best ' Print
            each best matching peptide sequence name, next to the
            drift that would have to be applied to the peak in order
            to have an exact retention time match
            End If
      Next


End Sub
```

```
'*********************************
' Calculate which row in the worksheet was the last row, to then allow
assessment of the entire dataset
' Written by Stuart Warriner
'*********************************

Function LastRow(s As Worksheet)

LastRow = s.Range("A2000").End(xlUp).Row

End Function
```

D.3.2 Average and count of peptide sequences found when each 'drift' time was applied to retention times

```
'*********************************
' Calculate the average of each of the errors on the 'drifted'
retention times and to count the number of peptides found when each
'drift' time was applied
' Written by Laura Cross
'*********************************

Sub AverageandCount()

Dim PeakAssign As Worksheet
Dim PeakAssignbook
Dim RTdiff1 As Dictionary
Dim RTdiff2 As Dictionary
Dim RTdiff3 As Dictionary
Dim RTdiff4 As Dictionary
Dim RTdiff5 As Dictionary
Dim RTdiff6 As Dictionary
Dim RTdiff7 As Dictionary
Dim RTdiff8 As Dictionary
Dim RTdiff9 As Dictionary
Dim RTdiff10 As Dictionary
Dim RTdiff11 As Dictionary

Dim h

Set RTdiff1 = New Scripting.Dictionary
Set RTdiff2 = New Scripting.Dictionary
Set RTdiff3 = New Scripting.Dictionary
Set RTdiff4 = New Scripting.Dictionary
Set RTdiff5 = New Scripting.Dictionary
Set RTdiff6 = New Scripting.Dictionary
Set RTdiff7 = New Scripting.Dictionary
Set RTdiff8 = New Scripting.Dictionary
Set RTdiff9 = New Scripting.Dictionary
Set RTdiff10 = New Scripting.Dictionary
Set RTdiff11 = New Scripting.Dictionary


Set PeakAssign = Worksheets(44)
Set PeakAssignbook =
Workbooks("DataProcess1.xlsm").Worksheets("Sheet44")

With PeakAssignbook

      For h = 2 To LastRow(PeakAssign)
```

```
                If Not IsEmpty(.Cells(h, 6)) Then
                RTdiff1.Add PeakAssign.Cells(h, 6), PeakAssign.Cells(h, 7)
                ' A 'dictionary' will be created, containing assigned
                peptide sequence against retention time drift (if that
                sequence were the correctly assigned peptide)
                End If

                If Not IsEmpty(.Cells(h, 8)) Then
                RTdiff2.Add PeakAssign.Cells(h, 8), PeakAssign.Cells(h, 9)
                ' Separate dictionaries are created for each 'drifted'
                column in the Worksheet, in the same way as the note above
                End If

                If Not IsEmpty(.Cells(h, 10)) Then
                RTdiff3.Add PeakAssign.Cells(h, 10),PeakAssign.Cells(h,11)
                End If

                If Not IsEmpty(.Cells(h, 12)) Then
                RTdiff4.Add PeakAssign.Cells(h, 12),PeakAssign.Cells(h,13)
                End If

                If Not IsEmpty(.Cells(h, 14)) Then
                RTdiff5.Add PeakAssign.Cells(h, 14),PeakAssign.Cells(h,15)
                End If

                If Not IsEmpty(.Cells(h, 16)) Then
                RTdiff6.Add PeakAssign.Cells(h, 16),PeakAssign.Cells(h,17)
                End If

                If Not IsEmpty(.Cells(h, 18)) Then
                RTdiff7.Add PeakAssign.Cells(h, 18),PeakAssign.Cells(h,19)
                End If

                If Not IsEmpty(.Cells(h, 20)) Then
                RTdiff8.Add PeakAssign.Cells(h, 20),PeakAssign.Cells(h,21)
                End If

                If Not IsEmpty(.Cells(h, 22)) Then
                RTdiff9.Add PeakAssign.Cells(h, 22),PeakAssign.Cells(h,23)
                End If

                If Not IsEmpty(.Cells(h, 24)) Then
                RTdiff10.Add PeakAssign.Cells(h,24),PeakAssign.Cells(h,25)
                End If

                If Not IsEmpty(.Cells(h, 26)) Then
                RTdiff11.Add PeakAssign.Cells(h,26),PeakAssign.Cells(h,27)
                End If

        Next

.Cells(265, 6).Value =
Application.WorksheetFunction.Average(RTdiff1.Items)
.Cells(265, 8).Value =
Application.WorksheetFunction.Average(RTdiff2.Items)
.Cells(265, 10).Value =
Application.WorksheetFunction.Average(RTdiff3.Items)
.Cells(265, 12).Value =
Application.WorksheetFunction.Average(RTdiff4.Items)
.Cells(265, 14).Value =
Application.WorksheetFunction.Average(RTdiff5.Items)
```

```
.Cells(265, 16).Value =
Application.WorksheetFunction.Average(RTdiff6.Items)
.Cells(265, 18).Value =
Application.WorksheetFunction.Average(RTdiff7.Items)
.Cells(265, 20).Value =
Application.WorksheetFunction.Average(RTdiff8.Items)
.Cells(265, 22).Value =
Application.WorksheetFunction.Average(RTdiff9.Items)
.Cells(265, 24).Value =
Application.WorksheetFunction.Average(RTdiff10.Items)
.Cells(265, 26).Value =
Application.WorksheetFunction.Average(RTdiff11.Items) ' The average of
retention time drift in each column is calculated and the result
placed in row 265, columns 6 (F), 8 (H), 10 (J)... to 26 (Z)

.Cells(266, 6).Value = RTdiff1.Count
.Cells(266, 8).Value = RTdiff2.Count
.Cells(266, 10).Value = RTdiff3.Count
.Cells(266, 12).Value = RTdiff4.Count
.Cells(266, 14).Value = RTdiff5.Count
.Cells(266, 16).Value = RTdiff6.Count
.Cells(266, 18).Value = RTdiff7.Count
.Cells(266, 20).Value = RTdiff8.Count
.Cells(266, 22).Value = RTdiff9.Count
.Cells(266, 24).Value = RTdiff10.Count
.Cells(266, 26).Value = RTdiff11.Count ' The count of retention time
drift in each column is calculated and the result placed in row 266,
columns 6 (F), 8 (H), 10 (J)... to 26 (Z)

End With

End Sub
```

D.3.3 Sorting of peptides by hydrophobicity

```
'*********************************
' Select the column with the highest count and lowest average, and
copy the data from this column into another Worksheet
' Written by Laura Cross except where stated otherwise
'*********************************

Sub SortByHydrophobicity()

Dim PeakAssign As Worksheet
Dim PeakAssignbook
Dim Output

Dim Averages As Dictionary
Dim BestCol As Dictionary

Dim i, j, k, l, m

Dim target As Variant
Dim seqs As Variant
Dim out, outarr

Set BestCol = New Scripting.Dictionary
Set Averages = New Scripting.Dictionary


Set PeakAssign = Worksheets(44)
```

```
Set PeakAssignbook =
Workbooks("DataProcess1.xlsm").Worksheets("Sheet44") ' The data are
being read from Worksheet 44. This number was changed for the
processing of each screen, as the '.csv' data from each screen was
copied and pasted into a separate Worksheet in the same Workbook
Set Output = Workbooks("DataProcess2.xlsm").Worksheets("Sheet1") ' The
'target file' for the data is 'DataProcess2', which contains all amino
acid sequences ordered by hydrophobicity, plus a column named 'area'
for the data to be transferred to, plus a column with areas from the
'blank' screen recorded


With PeakAssignbook

      For i = 6 To 26 ' This will start with column 6 (F) and apply
      the code up until column 26 (Z)

            If Not IsEmpty(.Cells(266, i)) Then ' If there is a value
            in row 266 and column i (this will contain the peptide-RT
            'count' if the overall column contains the peptide
            sequence names)
                  If .Cells(266, i).Value <>
                  Application.WorksheetFunction.Max(Range("F266:Z266")
                  ) Then
                  .Cells(266, i).ClearContents
                  .Cells(265, i).ClearContents ' Delete the contents
                  of rows 265 and 266 (containing average and count)
                  if the count is not the maximum across the range
                  End If
            End If

      Next

      For i = 6 To 26 ' This will look through the same columns again

            If Not IsEmpty(.Cells(266, i)) Then
                  If Not IsEmpty(.Cells(265, i)) Then ' If there are
                  values in rows 265 and 266 (which there will be if
                  the count is the maximum)...
                  Averages.Add Abs(Worksheets(44).Cells(264, i)),
                  Abs(Worksheets(44).Cells(265, i)) ' A 'dictionary'
                  is created containing identifier numbers (in row
                  264) against the averages of retention time drift
                  for all columns where count was the highest
                  End If
            End If

      Next

      For i = 6 To 26 ' This will look through the same columns again
            For j = 2 To LastRow(PeakAssign)

                  If Not IsEmpty(.Cells(266, i)) Then
                        If Abs(.Cells(265, i).Value) =
                        Application.WorksheetFunction.Min(Averages.Ite
                        ms) Then ' If the average of retention times
                        is the lowest absolute number in the
                        dictionary createed above...
                              If Not IsEmpty(.Cells(j, i)) Then
                                    If Not
                                    BestCol.Exists(Worksheets(44).Cell
                                    s(j, i).Value) Then
```

```
                                    BestCol.Add
                                    Trim(Worksheets(44).Cells(j, i)),
                                    Trim(Worksheets(44).Cells(j, 4)) '
                                    Another dictionary is created
                                    which takes the peptide names in
                                    the 'best row' (highest count,
                                    lowest average) against the EIC
                                    areas for each of those peptides
                                    End If
                              End If
                        End If
                  End If
            Next
      Next

End With

'*********************************
' From this point onwards, this part of the code was written by Stuart
Warriner and modified by Laura Cross
'*********************************

      For k = 2 To 325

      target = Output.Cells(k, 1).Value
      out = ""

            For l = 0 To BestCol.Count - 1
            seqs = BestCol.Keys(l) ' Keys from the dictionary
            containing data from the best row (the peptide names of
            this dictionary) are named 'seqs'

                  If InStr(1, seqs, target, vbTextCompare) > 0 Then '
                  If the peptide name from the 'best row' dictionary
                  is found in the target file (containing all peptide
                  sequences sorted by hydrophobicity) (which it will
                  be)...

                        If out <> "" Then
                        out = out & "_"
                        End If
                        out = out & BestCol.Items(l) ' Record the
                        peptide EIC areas for transfer into the
                        'target file'


                  End If
            Next
      outarr = Split(out, "_")
            For m = 0 To UBound(outarr)
            Output.Cells(k, m + 3).Value = outarr(m) ' Transfer the
            EIC areas next to their corresponding peptide sequence in
            the list of peptides ordered by hydrophobicity
            Next
      Next


End Sub

'*********************************
' Calculate which row in the worksheet was the last row, to then allow
assessment of the entire dataset
```

```
' Written by Stuart Warriner
'*********************************

Function LastRow(s As Worksheet)

LastRow = s.Range("A2000").End(xlUp).Row

End Function
```

D.3.4 Adding EIC areas for pair set 2 (of the 45 pairs of peptides without unique RT-mass identifiers) when a peptide of pair set 1 was identified in the screen

```
'*********************************
' Add EIC areas for the 45 peptides not in the master seq-RT list (of
the pairs without unique RT-mass identifiers) if the corresponding
peptide in the pair was identified
' Written by Laura Cross and modified by Stuart Warriner
'*********************************

Sub AddMatchingRTseqs()

Dim IdentifyCmpd
Dim SameRTs
Dim i, h, j, k

Dim target1 As Variant
Dim target2 As Variant
Dim seqs1 As Variant
Dim seqs2 As Variant
Dim t1arr()
Dim t2arr()
Dim s1arr()
Dim s2arr()

Set IdentifyCmpd = Workbooks("DataProcess2.xlsm").Worksheets("Sheet1")
Set SameRTs =
Workbooks("DataProcessSameRTs.xlsx").Worksheets("Sheet1")


t1arr = IdentifyCmpd.Range("A2:A325").Value ' Create a target array
containing all peptide 'names', e.g. XYQSKL
t2arr = IdentifyCmpd.Range("B2:B325").Value ' Create another target
array containing the same peptide 'names', e.g. XYQSKL
s1arr = SameRTs.Range("A2:A46").Value ' Create an array of peptide
names in the master 'seq-RT' list which belong to the 45 pairs of
peptides without a unique RT-mass identifier (set 1 of the pairs)
s2arr = SameRTs.Range("B2:B46").Value ' Create an array of the peptide
names which do not appear in the master 'seq-RT' list and which belong
to the 43 pairs of peptides without a unique RT-mass identifier (set 2
of the pairs)
area = IdentifyCmpd.Range("C2:C325").Value ' Name the range containing
EIC areas of peptides identified in the screen

Application.ScreenUpdating = False
Application.Calculation = xlCalculationManual

      For i = 1 To UBound(t1arr)
      target1 = t1arr(i, 1)

            For h = 1 To UBound(t2arr)
```

```
            target2 = t2arr(h, 1)

                  For j = 1 To UBound(s1arr)
                  seqs1 = s1arr(j, 1)
                  seqs2 = s2arr(j, 1)

                        If (InStr(1, seqs1, target1, vbTextCompare) >
                        0 And area(i, 1) > 0) Then ' If set 1 of the
                        43 pairs of peptides are found (which they
                        will be), do something if there is a value in
                        the 'EIC area' column
                              If InStr(1, seqs2, target2,
                              vbTextCompare) > 0 Then

                              IdentifyCmpd.Cells(h + 1, 3) =
                              IdentifyCmpd.Cells(i + 1, 3) ' If set 2
                              of the 45 pairs of peptides are found
                              (which they will be), take the EIC area
                              next to set 1 and place it next to set 2
                              of each pair
                              End If

                        End If

                  Next
            Next
      Next

Application.ScreenUpdating = True
Application.Calculation = xlCalculationAutomatic

End Sub
```

D.3.5 Blank correcting of data

```
'*********************************
' Subtract EIC areas from the 'blank' screen (run using the same
protocol as a screen but without protein in the initial incubation)
' Also replace any empty cells (where no EIC area was detected) with
the value 0 (for use of the data in OriginPro 9.1)
' Written by Laura Cross
'*********************************

Sub BlankCorrect()

Dim DataProcess2
Dim i, j

Set DataProcess2 = Workbooks("DataProcess2.xlsm").Worksheets("Sheet1")

With DataProcess2

    For i = 2 To 325
      .Cells(i, 5).Value = .Cells(i, 3).Value - .Cells(i, 4).Value '
      Minus EIC area for peptide in blank screen (already pasted into
      Worksheet) from EIC area of peptide in current screen
    Next

      For j = 2 To 325
            If .Cells(j, 5).Value < 0 Then ' Place the value "0" in
            any blank cells with no EIC area
```

```
            .Cells(j, 5).Value = "0"
            End If
      Next

End With

End Sub
```

D.3.6 Splitting -1 and -2 amino acids into separate columns

```
'*********************************
' Split the two C-terminal amino acids into separate columns
' Written by Laura Cross
'*********************************

Sub SplitAAs()

Dim DataProcess2
Dim q
Set DataProcess2 = Workbooks("DataProcess2.xlsm").Worksheets("Sheet1")

With DataProcess2

Range("E2").EntireColumn.Insert
Range("F2").EntireColumn.Insert

End With

With DataProcess2

Range("E1").Value = "Position -1"
Range("F1").Value = "Position -2"

End With

With DataProcess2

      For q = 2 To 325
            While Not IsEmpty(.Cells(q, 2).Value)
            .Cells(q, 5).Value = Mid((.Cells(q, 2).Value), 6, 1) '
            Take the C-terminal amino acid one-letter code and place
            this in column E
            q = q + 1
            Wend
      Next

      For q = 2 To 325
            While Not IsEmpty(.Cells(q, 2).Value)
            .Cells(q, 6).Value = Mid((.Cells(q, 2).Value), 5, 1) '
            Take the '-2' amino acid one-letter code and place this in
            column F
            q = q + 1
            Wend
      Next

End With

End Sub
```

D.3.7 Converting amino acids to numbers

```
'*********************************
' Replace the two C-terminal amino acid one-letter codes with numbers
for plotting a matrix of the data in OriginPro 9.1
' Written by Laura Cross
'*********************************

Sub ReplaceAAsWithNumbers()

Dim IdentifyCmpd
Dim a, b

Set IdentifyCmpd = Workbooks("DataProcess2.xlsm").Worksheets("Sheet1")

With IdentifyCmpd

      For a = 2 To 325

            If .Cells(a, 5).Text = "F" Then
            .Cells(a, 5).Value = "1"
            End If

            If .Cells(a, 5).Text = "I" Then
            .Cells(a, 5).Value = "2"
            End If

            If .Cells(a, 5).Text = "W" Then
            .Cells(a, 5).Value = "3"
            End If

            If .Cells(a, 5).Text = "L" Then
            .Cells(a, 5).Value = "4"
            End If

            If .Cells(a, 5).Text = "V" Then
            .Cells(a, 5).Value = "5"
            End If

            If .Cells(a, 5).Text = "Y" Then
            .Cells(a, 5).Value = "6"
            End If

            If .Cells(a, 5).Text = "A" Then
            .Cells(a, 5).Value = "7"
            End If

            If .Cells(a, 5).Text = "T" Then
            .Cells(a, 5).Value = "8"
            End If

            If .Cells(a, 5).Text = "H" Then
            .Cells(a, 5).Value = "9"
            End If

            If .Cells(a, 5).Text = "G" Then
            .Cells(a, 5).Value = "10"
            End If

            If .Cells(a, 5).Text = "S" Then
            .Cells(a, 5).Value = "11"
            End If
```

```
        If .Cells(a, 5).Text = "Q" Then
        .Cells(a, 5).Value = "12"
        End If

        If .Cells(a, 5).Text = "R" Then
        .Cells(a, 5).Value = "13"
        End If

        If .Cells(a, 5).Text = "K" Then
        .Cells(a, 5).Value = "14"
        End If

        If .Cells(a, 5).Text = "N" Then
        .Cells(a, 5).Value = "15"
        End If

        If .Cells(a, 5).Text = "E" Then
        .Cells(a, 5).Value = "16"
        End If

        If .Cells(a, 5).Text = "P" Then
        .Cells(a, 5).Value = "17"
        End If

        If .Cells(a, 5).Text = "D" Then
        .Cells(a, 5).Value = "18"
        End If

Next

For b = 2 To 325

        If .Cells(b, 6).Text = "D" Then
        .Cells(b, 6).Value = "1"
        End If

        If .Cells(b, 6).Text = "P" Then
        .Cells(b, 6).Value = "2"
        End If

        If .Cells(b, 6).Text = "E" Then
        .Cells(b, 6).Value = "3"
        End If

        If .Cells(b, 6).Text = "N" Then
        .Cells(b, 6).Value = "4"
        End If

        If .Cells(b, 6).Text = "K" Then
        .Cells(b, 6).Value = "5"
        End If

        If .Cells(b, 6).Text = "R" Then
        .Cells(b, 6).Value = "6"
        End If

        If .Cells(b, 6).Text = "Q" Then
        .Cells(b, 6).Value = "7"
        End If

        If .Cells(b, 6).Text = "S" Then
        .Cells(b, 6).Value = "8"
```

```
            End If

            If .Cells(b, 6).Text = "G" Then
            .Cells(b, 6).Value = "9"
            End If

            If .Cells(b, 6).Text = "H" Then
            .Cells(b, 6).Value = "10"
            End If

            If .Cells(b, 6).Text = "T" Then
            .Cells(b, 6).Value = "11"
            End If

            If .Cells(b, 6).Text = "A" Then
            .Cells(b, 6).Value = "12"
            End If

            If .Cells(b, 6).Text = "Y" Then
            .Cells(b, 6).Value = "13"
            End If

            If .Cells(b, 6).Text = "V" Then
            .Cells(b, 6).Value = "14"
            End If

            If .Cells(b, 6).Text = "L" Then
            .Cells(b, 6).Value = "15"
            End If

            If .Cells(b, 6).Text = "W" Then
            .Cells(b, 6).Value = "16"
            End If

            If .Cells(b, 6).Text = "I" Then
            .Cells(b, 6).Value = "17"
            End If

            If .Cells(b, 6).Text = "F" Then
            .Cells(b, 6).Value = "18"
            End If

      Next

End With

End Sub
```

D.3.8 Reset of worksheet for processing of the next dataset

```
'*********************************
' Reset the 'DataProcess2' Worksheet for processing of the next
dataset
' Written by Laura Cross
'*********************************

Sub ResetSheet()

Dim DataProcess2
Dim a, b

Set DataProcess2 = Workbooks("DataProcess2.xlsm").Worksheets("Sheet1")

With DataProcess2

Range("F2").EntireColumn.Delete
Range("E2").EntireColumn.Delete

End With

With DataProcess2

      For a = 2 To 325
    .Cells(a, 3).ClearContents
      Next

      For b = 2 To 325
    .Cells(b, 5).ClearContents
      Next

End With

End Sub
```

Appendix E

Pull-down-LC-MS results for all *At*PEX5-C variants

E.1 Wild-type *At*PEX5(340-728): pull-down-LC-MS heat map

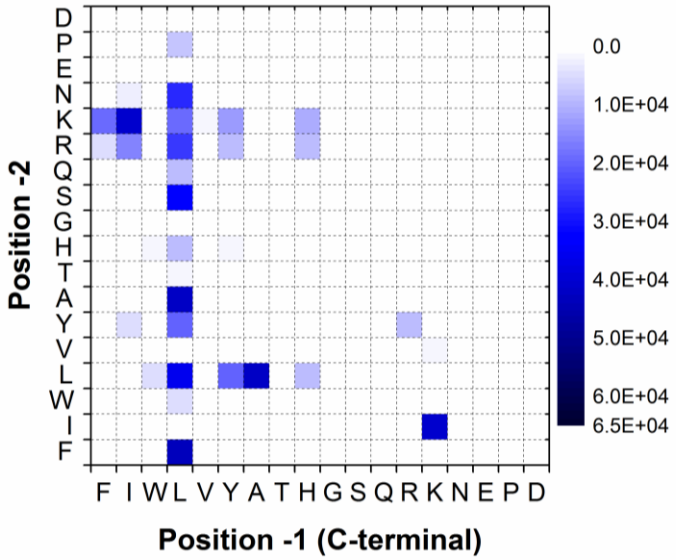


Figure E-1| Pull-down-LC-MS screen result for wild-type *At*PEX5-C. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.2 *At*PEX5(340-728) D505A: pull-down-LC-MS heat map

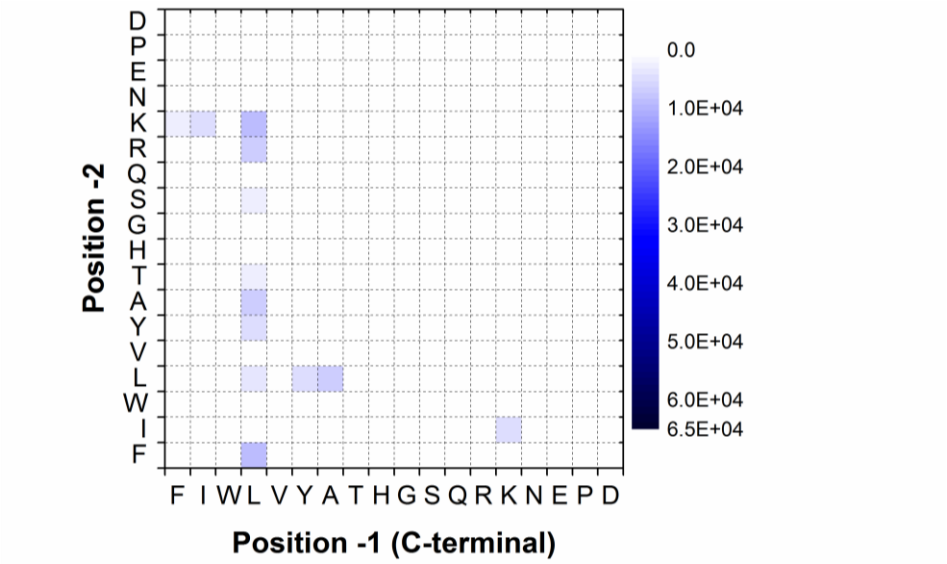


Figure E-2| Pull-down-LC-MS screen result for *At*PEX5-C D505A. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.3 *At*PEX5(340-728) D505H: pull-down-LC-MS heat map

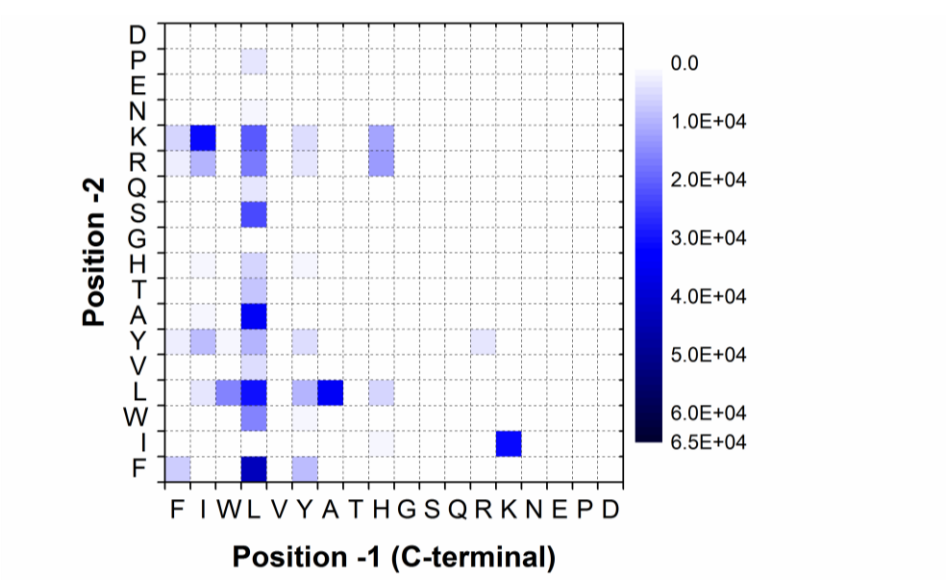


Figure E-3| Pull-down-LC-MS screen result for *At*PEX5-C D505H. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.4 *At*PEX5(340-728) D505K: pull-down-LC-MS heat map

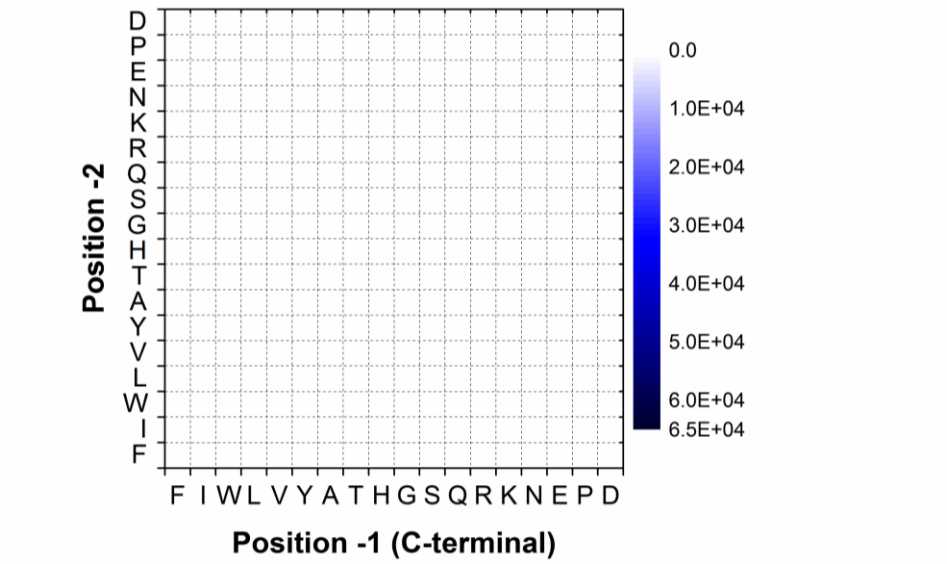


Figure E-4| Pull-down-LC-MS screen result for *At*PEX5-C D505K. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.5 *At*PEX5(340-728) D507A: pull-down-LC-MS heat map

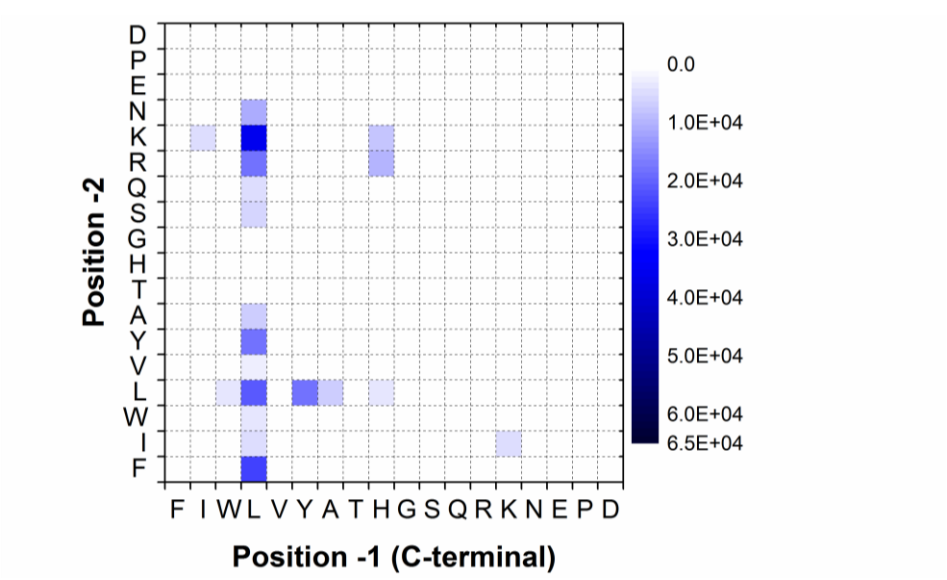


Figure E-5| Pull-down-LC-MS screen result for *At*PEX5-C D507A. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.6 *At*PEX5(340-728) D507K: pull-down-LC-MS heat map

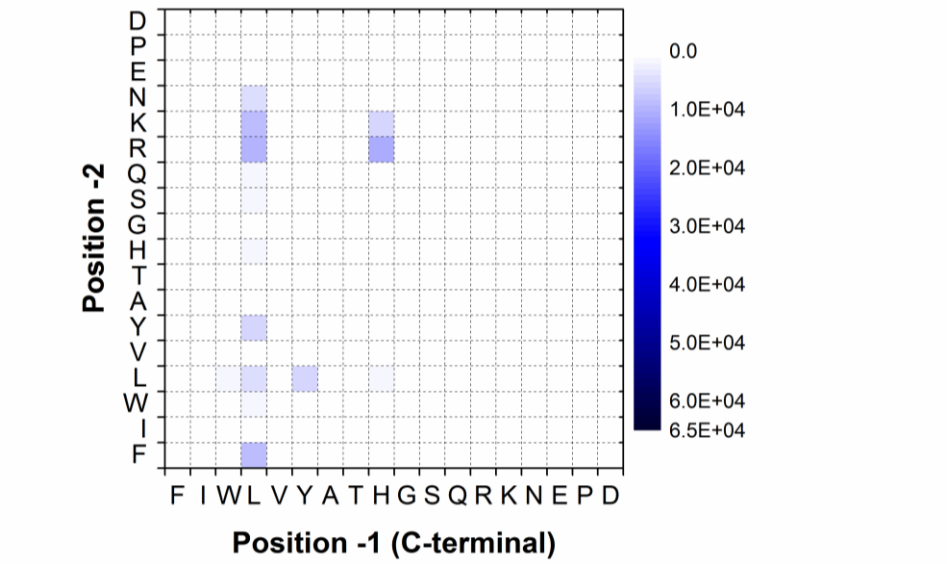


Figure E-6| Pull-down-LC-MS screen result for *At*PEX5-C D507K. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.7 *At*PEX5(340-728) V533A: pull-down-LC-MS heat map

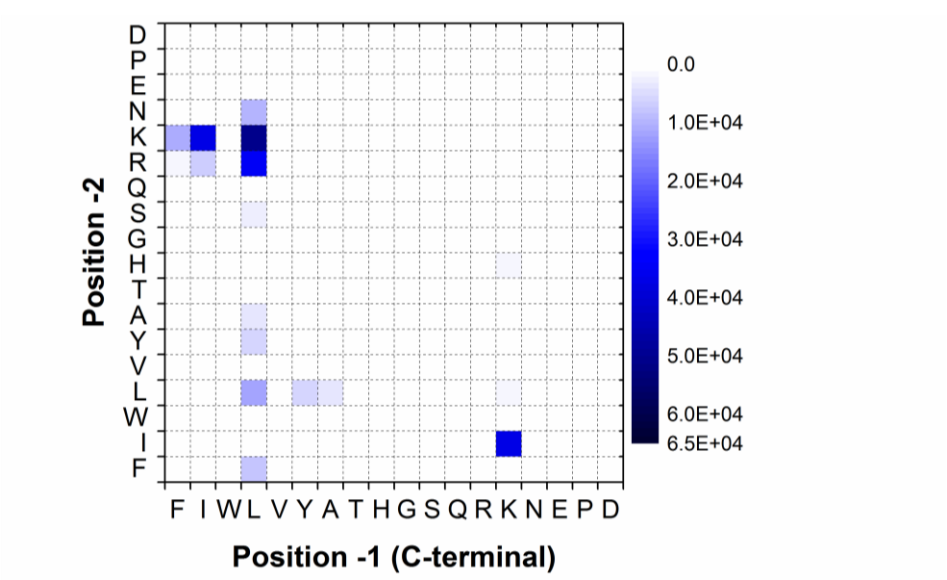


Figure E-7| Pull-down-LC-MS screen result for *At*PEX5-C V533A. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.8 *At*PEX5(340-728) V533W: pull-down-LC-MS heat map

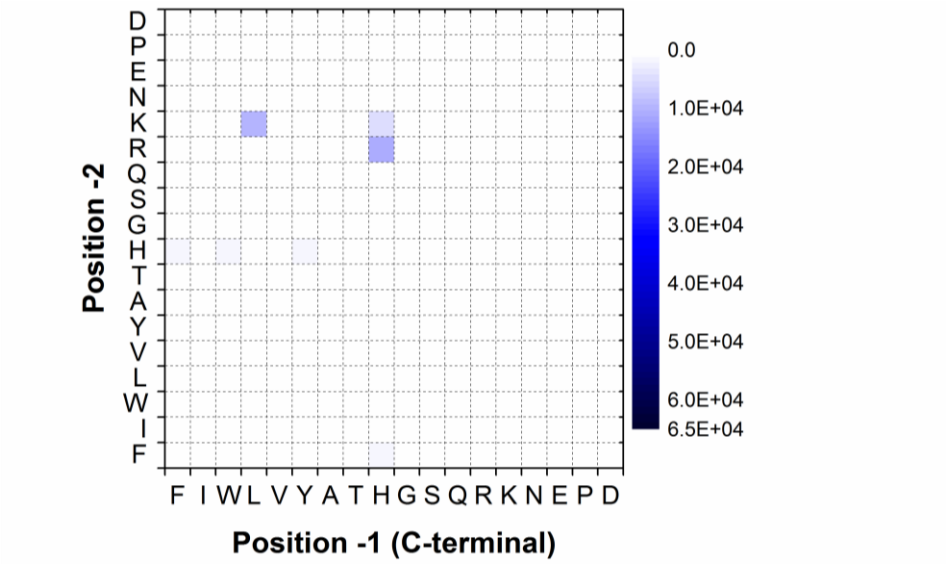


Figure E-8| Pull-down-LC-MS screen result for *At*PEX5-C V533W. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.9 *At*PEX5(340-728) T536A: pull-down-LC-MS heat map

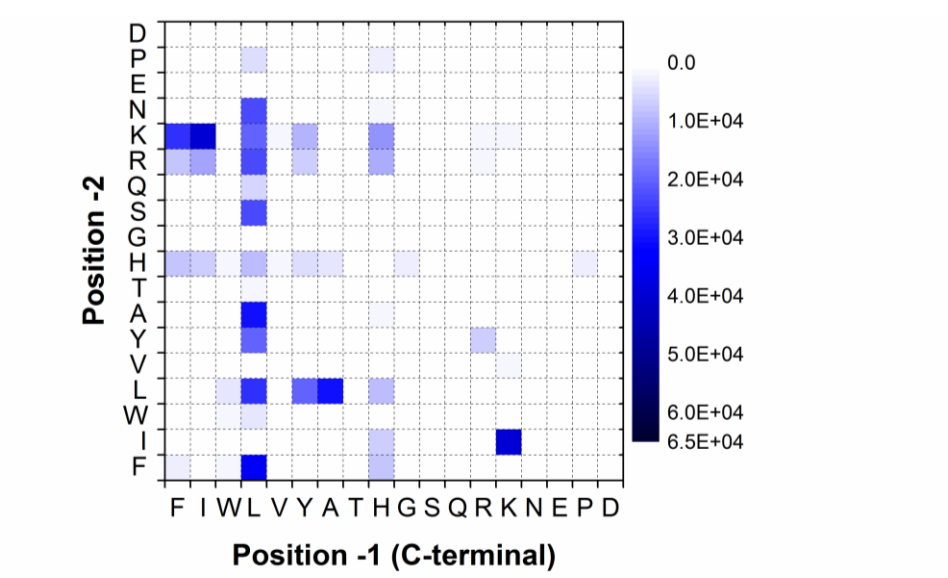


Figure E-9| Pull-down-LC-MS screen result for *At*PEX5-C T536A. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.10 *At*PEX5(340-728) T536N: pull-down-LC-MS heat map

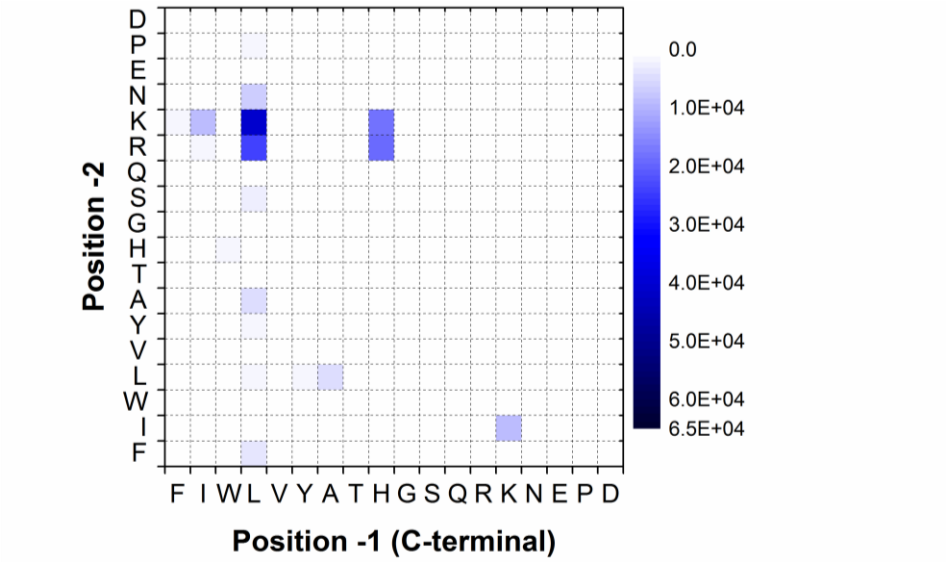


Figure E-10| Pull-down-LC-MS screen result for *At*PEX5-C T536N. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.11 *At*PEX5(340-728) T536W: pull-down-LC-MS heat map

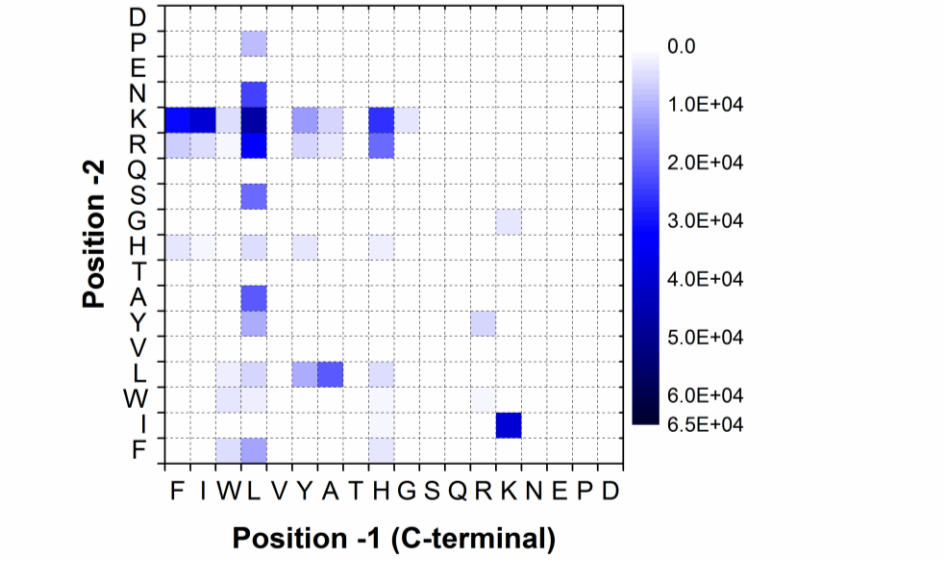


Figure E-11| Pull-down-LC-MS screen result for *At*PEX5-C T536W. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.12 *At*PEX5(340-728) N537A: pull-down-LC-MS heat map

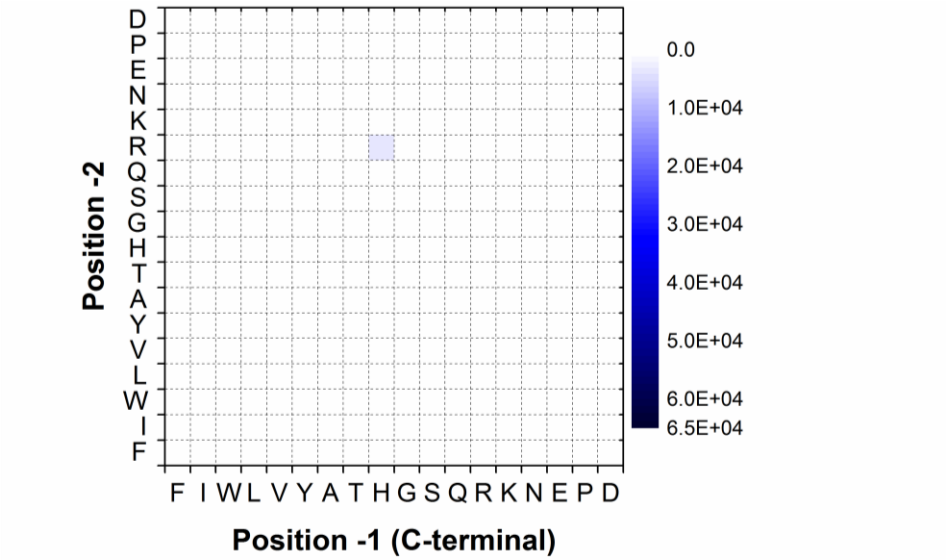


Figure E-12| Pull-down-LC-MS screen result for *At*PEX5-C N537A. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.13 *At*PEX5(340-728) N537Q: pull-down-LC-MS heat map

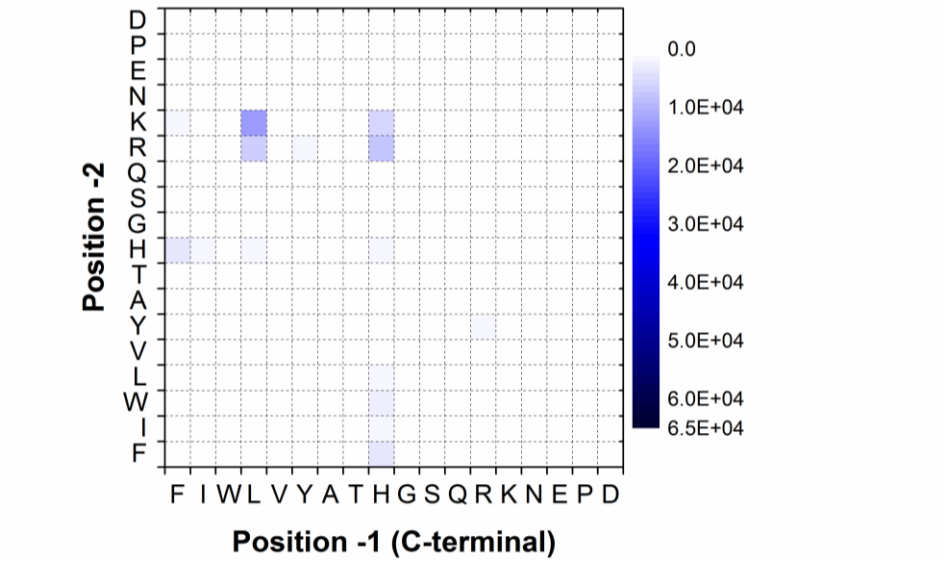


Figure E-13| Pull-down-LC-MS screen result for *At*PEX5-C N537Q. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.14 *At*PEX5(340-728) N537T: pull-down-LC-MS heat map

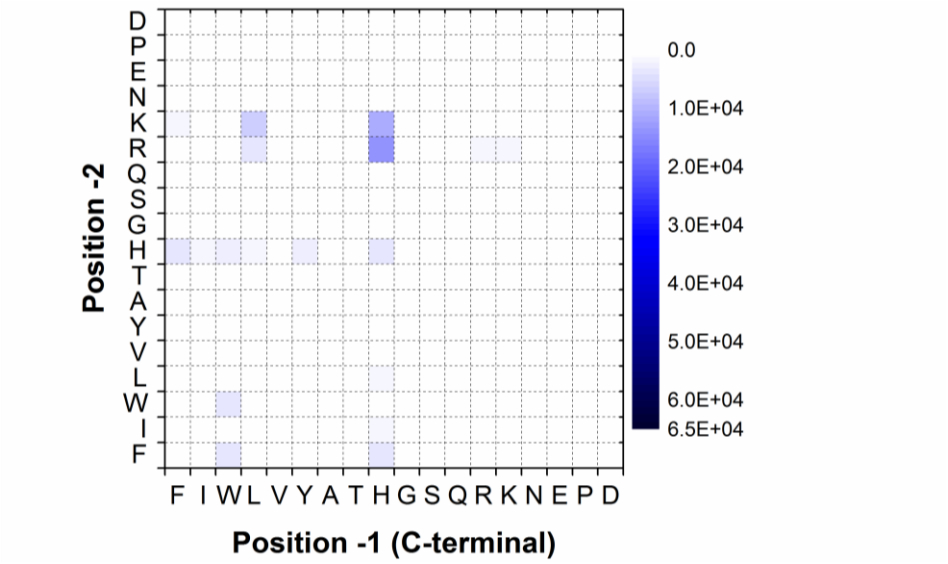


Figure E-14| Pull-down-LC-MS screen result for *At*PEX5-C N537T. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.15 *At*PEX5(340-728) E538A: pull-down-LC-MS heat map

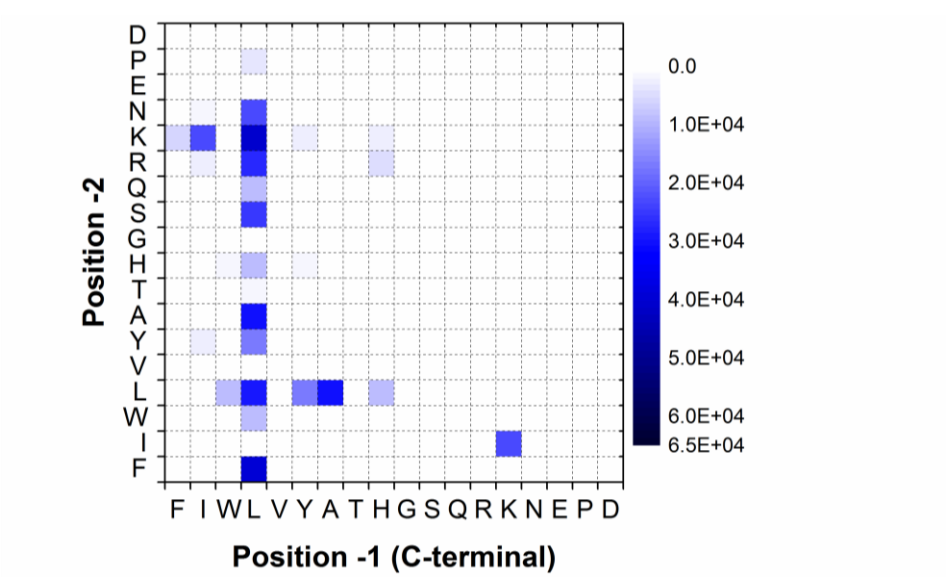


Figure E-15| Pull-down-LC-MS screen result for *At*PEX5-C E538A. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.16 *At*PEX5(340-728) N601A: pull-down-LC-MS heat map

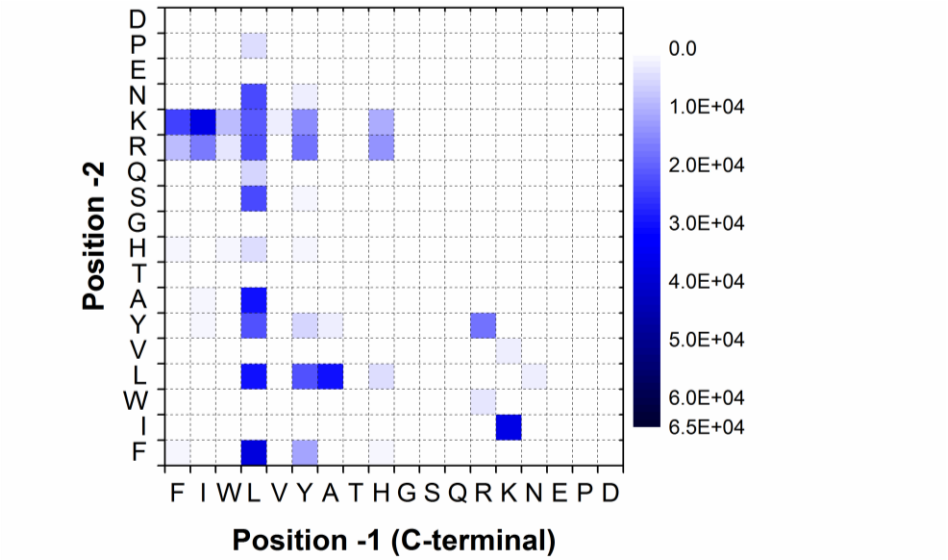


Figure E-16| Pull-down-LC-MS screen result for *At*PEX5-C N601A. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.17 *At*PEX5(340-728) N601Q: pull-down-LC-MS heat map

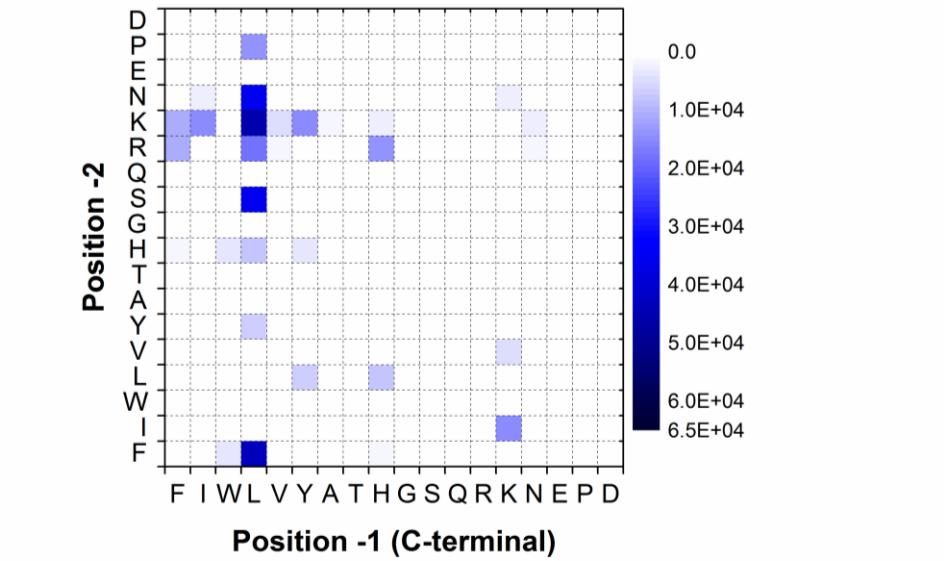


Figure E-17| Pull-down-LC-MS screen result for *At*PEX5-C N601Q. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.18 *At*PEX5(340-728) F613A: pull-down-LC-MS heat map

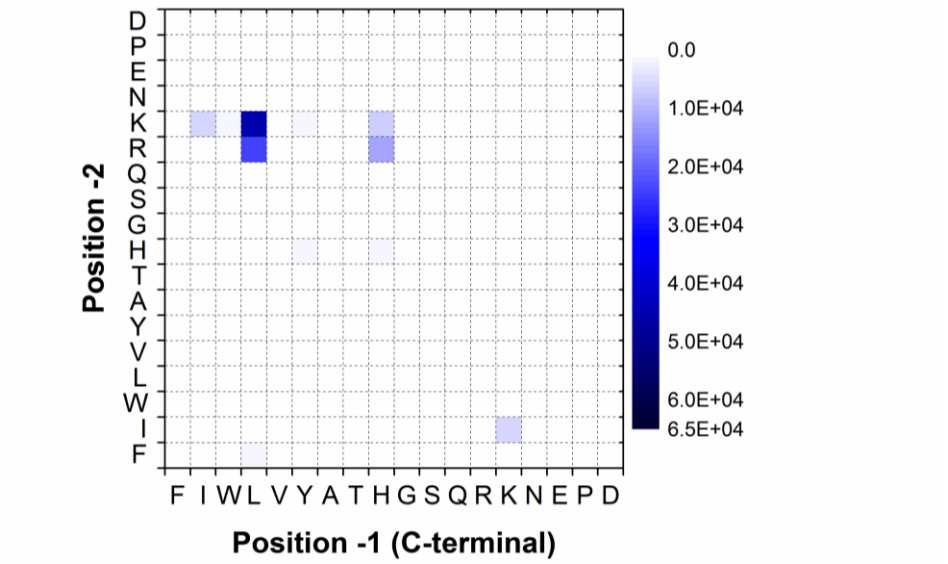


Figure E-18| Pull-down-LC-MS screen result for *At*PEX5-C F613A. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.19 *At*PEX5(340-728) N628A: pull-down-LC-MS heat map

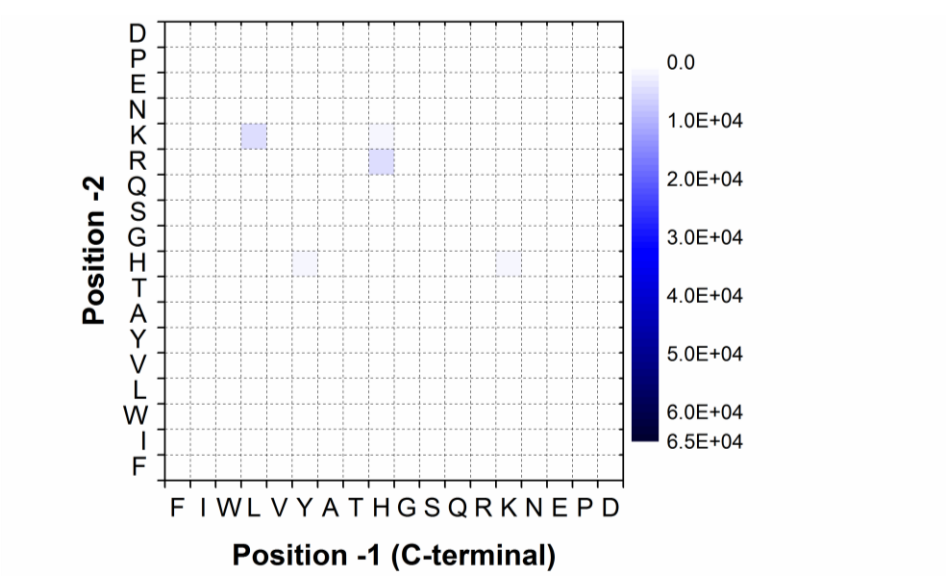


Figure E-19| Pull-down-LC-MS screen result for *At*PEX5-C N628A. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.20 *At*PEX5(340-728) A632G: pull-down-LC-MS heat map

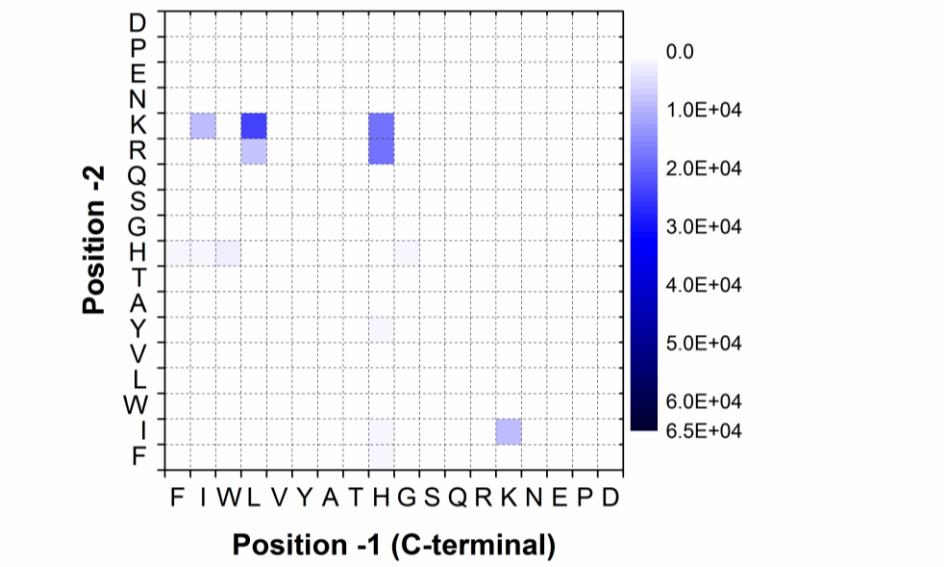


Figure E-20| Pull-down-LC-MS screen result for *At*PEX5-C A632G. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.21 *At*PEX5(340-728) N636A: pull-down-LC-MS heat map

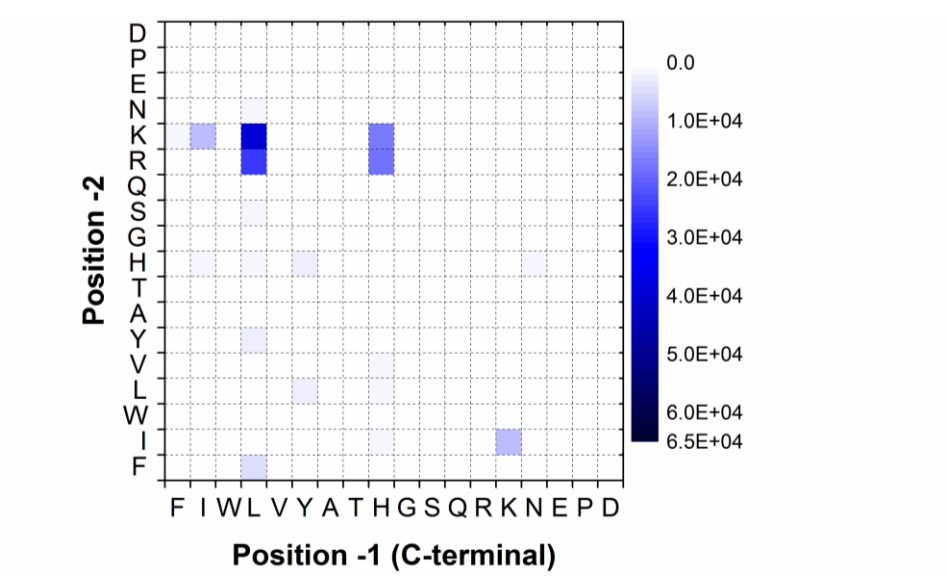


Figure E-21| Pull-down-LC-MS screen result for *At*PEX5-C N636A. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.22 *At*PEX5(340-728) Y647F: pull-down-LC-MS heat map

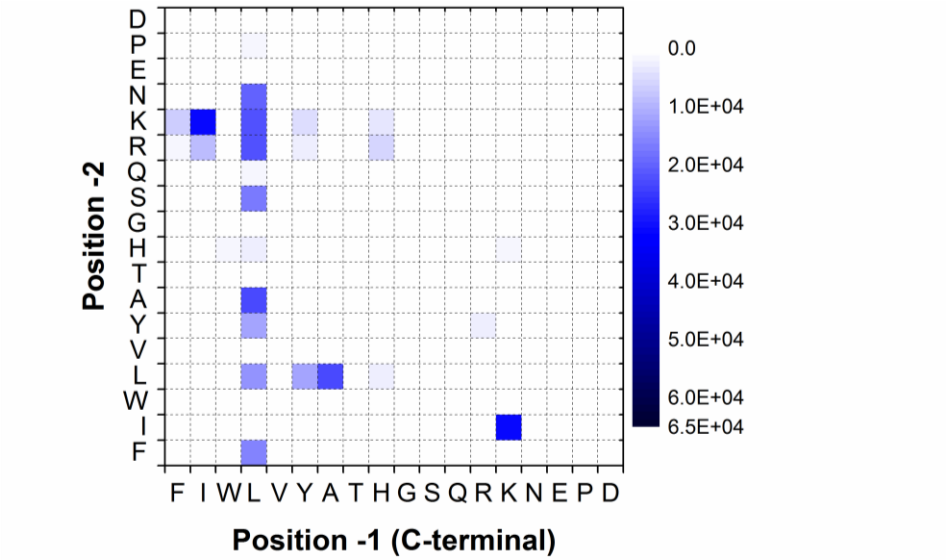


Figure E-22| Pull-down-LC-MS screen result for *At*PEX5-C Y647F. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.23 *At*PEX5(340-728) R659A: pull-down-LC-MS heat map

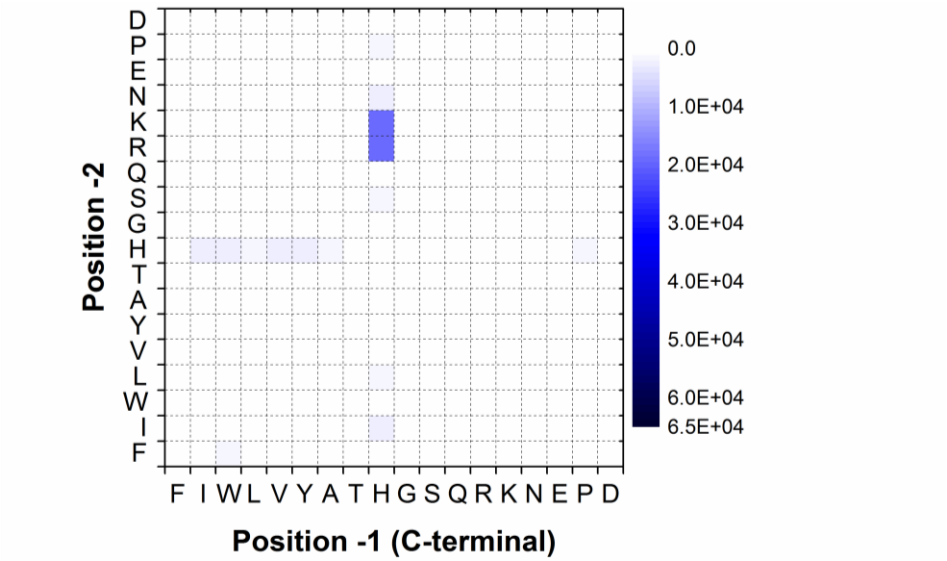


Figure E-23| Pull-down-LC-MS screen result for *At*PEX5-C R659A. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.24 *At*PEX5(340-728) N663A: pull-down-LC-MS heat map

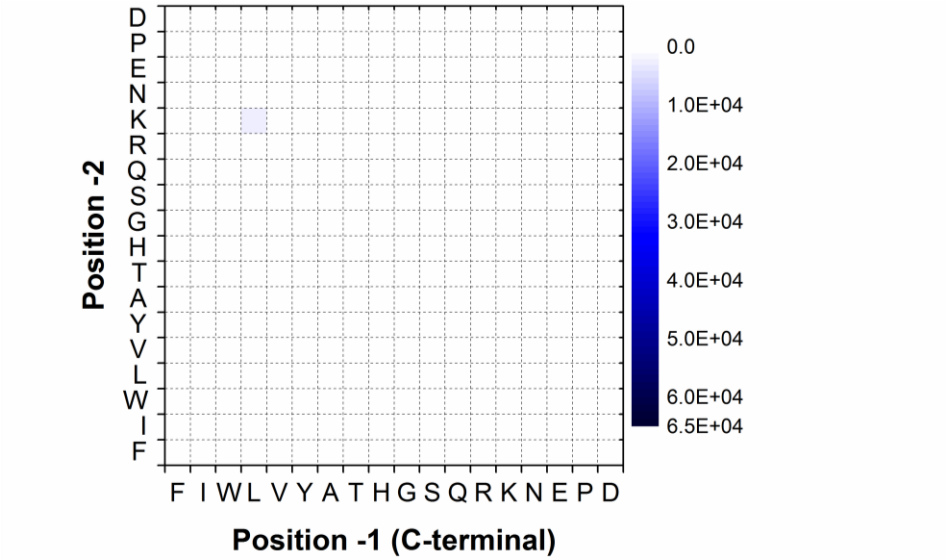


Figure E-24| Pull-down-LC-MS screen result for *At*PEX5-C N663A. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.25 *At*PEX5(340-728) S667A: pull-down-LC-MS heat map

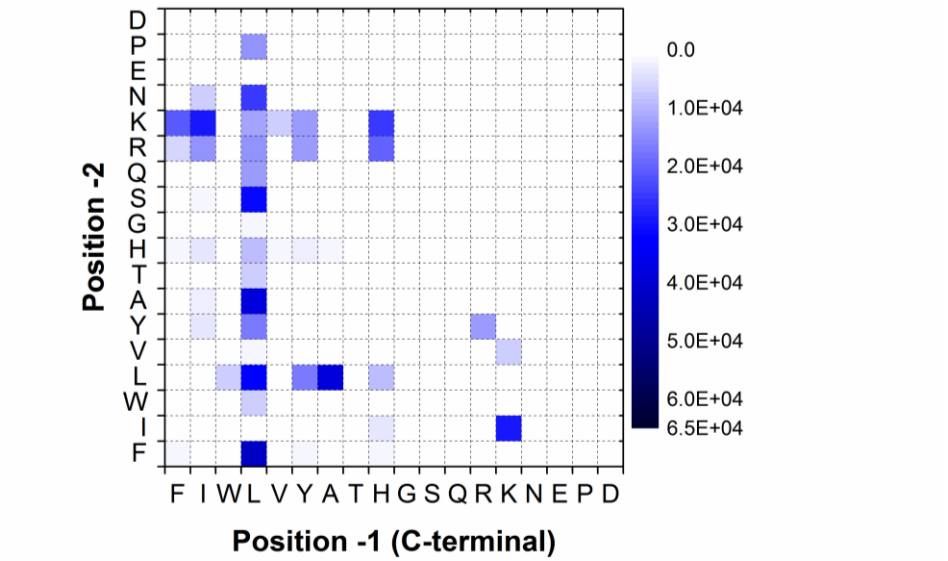


Figure E-25| Pull-down-LC-MS screen result for *At*PEX5-C S667A. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.26 *At*PEX5(340-728) D505K-D507K: pull-down-LC-MS heat map

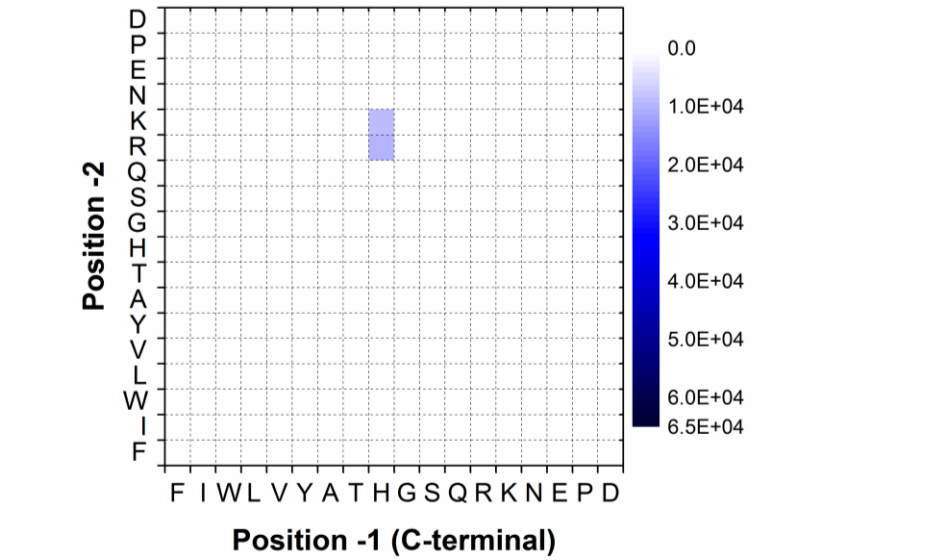


Figure E-26| Pull-down-LC-MS screen result for *At*PEX5-C D505K-D507K. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.27 *At*PEX5(340-728) D505H-T536W: pull-down-LC-MS heat map

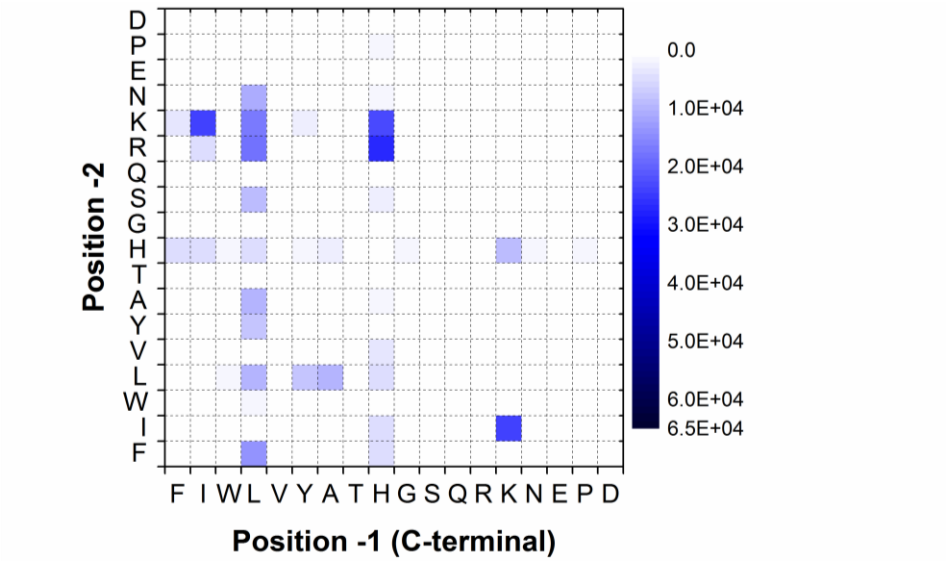


Figure E-27| Pull-down-LC-MS screen result for *At*PEX5-C D505H-T536W. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.28 *At*PEX5(340-728) D505F-N601A: pull-down-LC-MS heat map

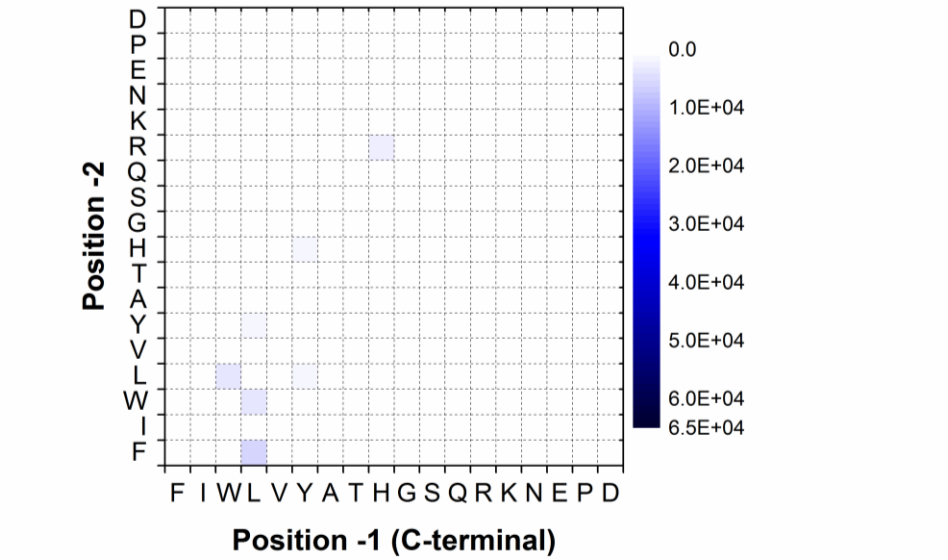


Figure E-28| Pull-down-LC-MS screen result for *At*PEX5-C D505F-N601A. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.29 *At*PEX5(340-728) D505H-N601A: pull-down-LC-MS heat map

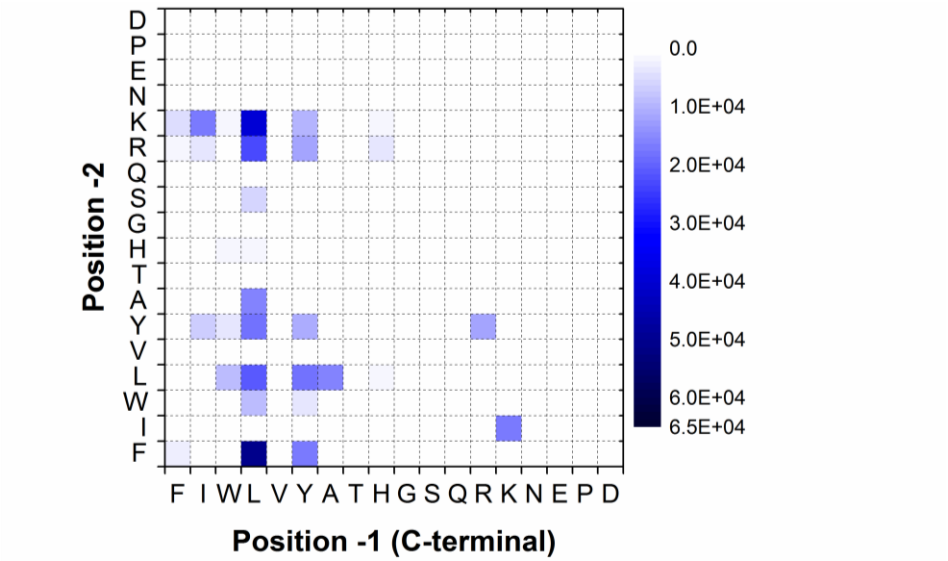


Figure E-29| Pull-down-LC-MS screen result for *At*PEX5-C D505H-N601A. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.30 *At*PEX5(340-728) D505F-D507F-N601A: pull-down-LC-MS heat map

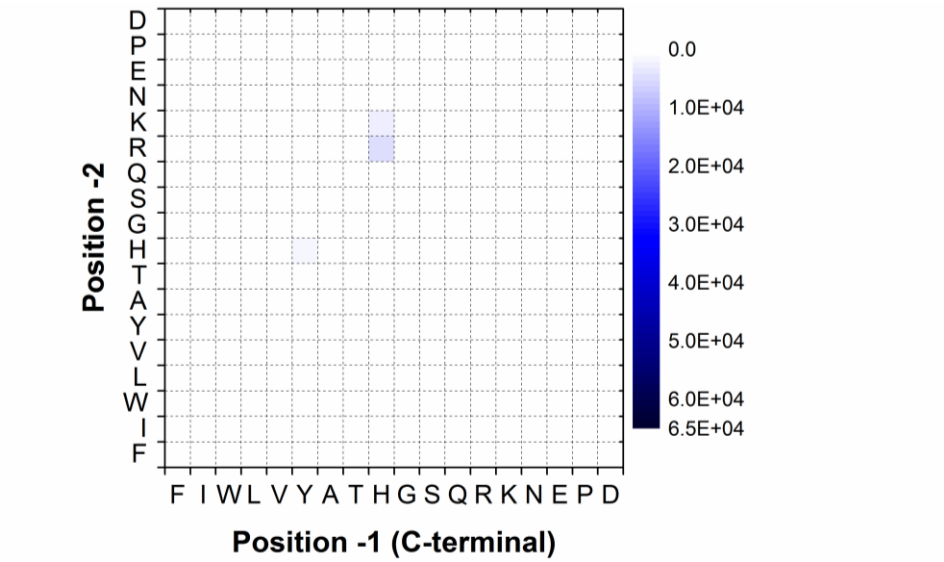


Figure E-30| Pull-down-LC-MS screen result for *At*PEX5-C D505F-D507F-N601A. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.31 *At*PEX5(340-728) D505H-D507H-N601A: pull-down-LC-MS heat map

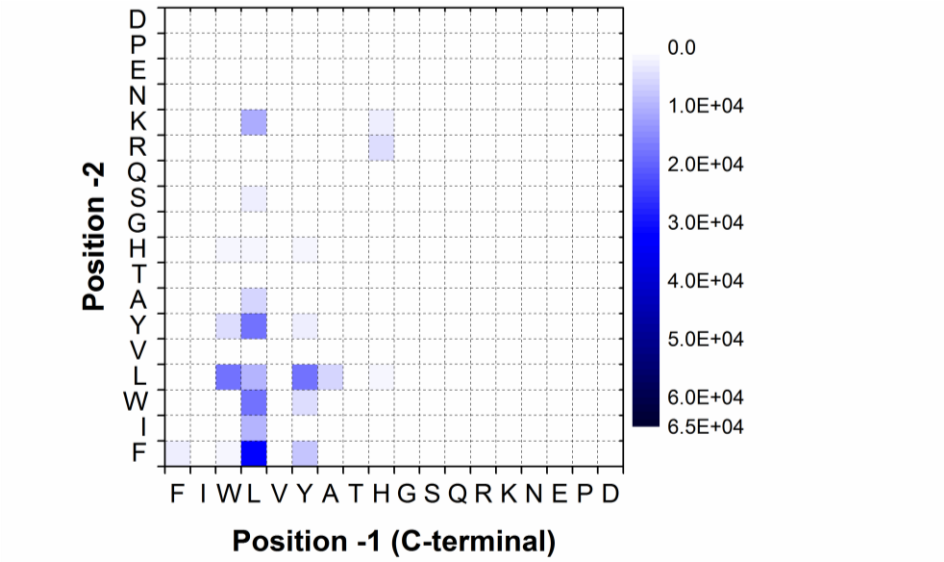


Figure E-31| Pull-down-LC-MS screen result for *At*PEX5-C D505H-D507H-N601A. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.32 *At*PEX5(340-728) D505H-D507T-N601A: pull-down-LC-MS heat map

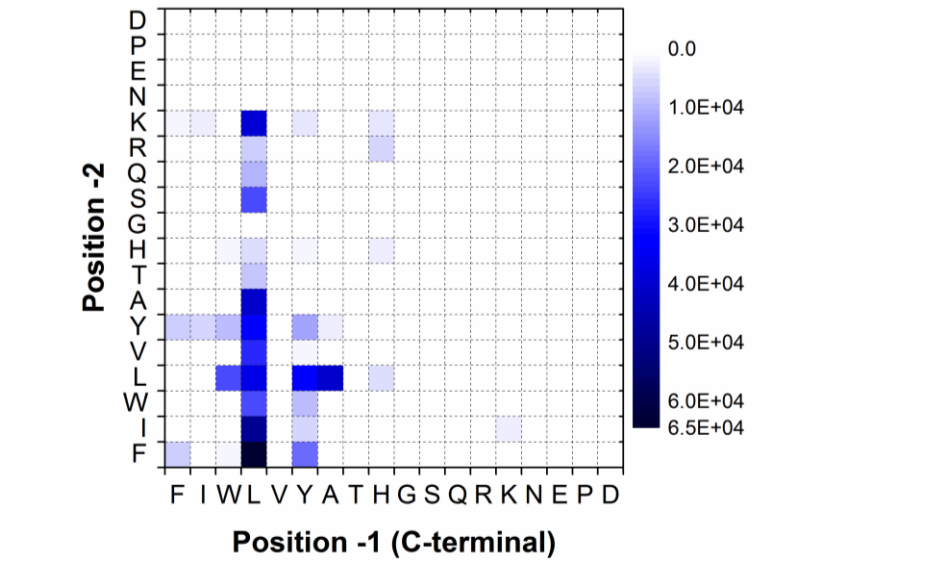


Figure E-32| Pull-down-LC-MS screen result for *At*PEX5-C D505H-D507T-N601A. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.33 *At*PEX5(340-728) D505H-D507V-N601A: pull-down-LC-MS heat map

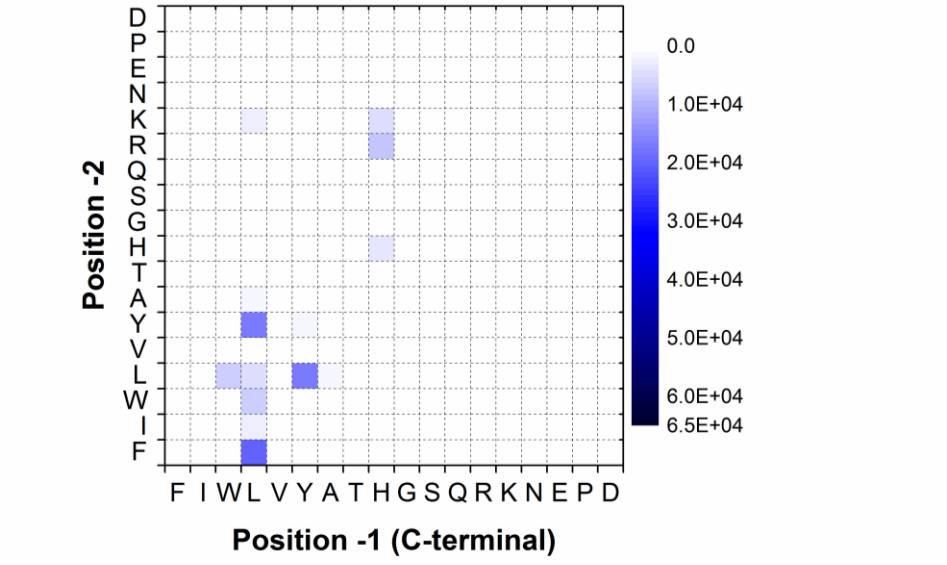


Figure E-33| Pull-down-LC-MS screen result for *At*PEX5-C D505H-D507V-N601A. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.34 *At*PEX5(340-728) D505H-N601A-N636A: pull-down-LC-MS heat map

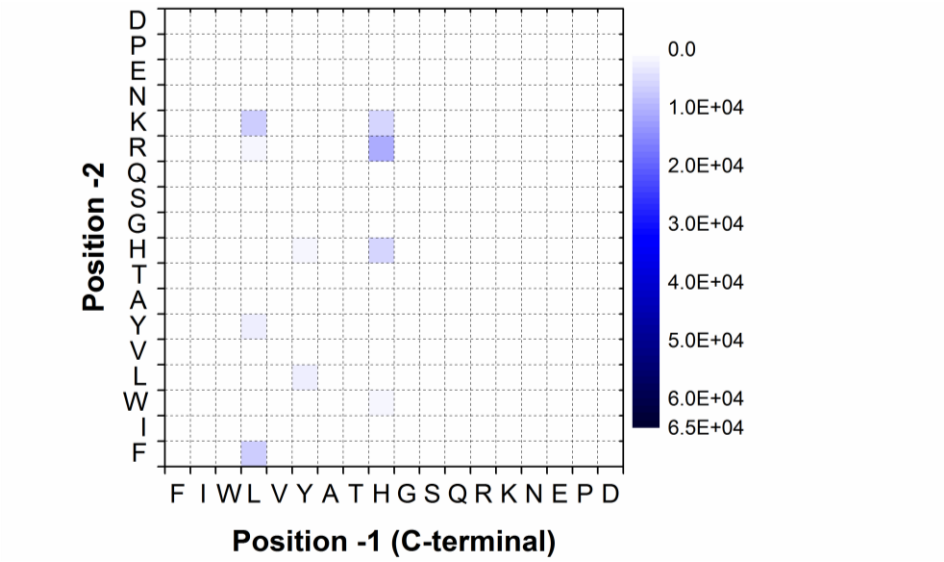


Figure E-34| Pull-down-LC-MS screen result for *At*PEX5-C D505H-N601A-N636A. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.35 *At*PEX5(340-728) D505H-D507T-E538A-N601A: pull-down-LC-MS heat map

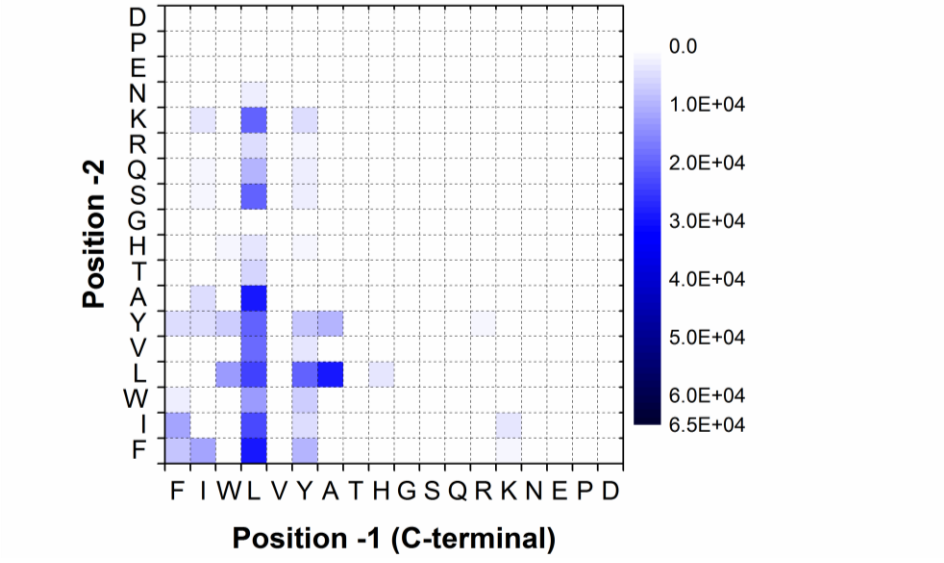


Figure E-35| Pull-down-LC-MS screen result for *At*PEX5-C D505H-D507T-E538A-N601A. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

E.36 Wild-type *At*PEX5(444-728) : pull-down-LC-MS heat map

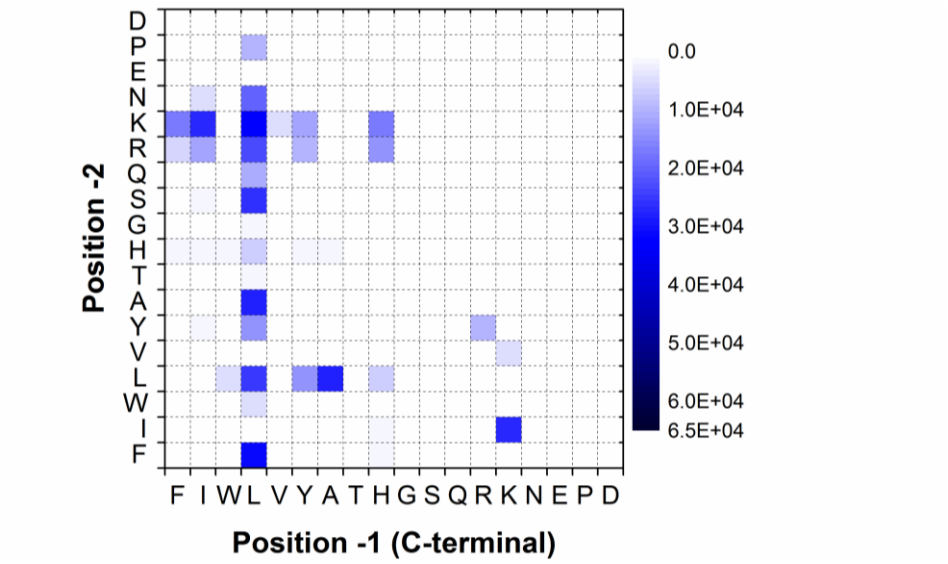


Figure E-36| Pull-down-LC-MS screen result for wild-type truncated *At*PEX5-C. Shading represents the area-under-EIC (extracted ion chromatogram) for each of the peptides after pull-down-LC-MS.

Appendix F

Additional data (generated by H. Ebeed and L. Cross)

F.1 H. Ebeed *P. patens* PTS1 homologues, based on *A. thaliana* PTS1 sequences

	Total PWM score
14 C-terminal sequences from *P. patens* homologues to *A. thaliana* proteins with major PTS1 sequences	
PRLEDWPVMPVANM	0.41
RKNLQQLAILKSNL	0.42
ISPMNVQQLSRLRL	0.42
SQYGPVTCPKKSSL	0.47
QCSVRQQQSLQSCL	0.5
EAKLSRMEIFQCKL	0.54
RKDLVQLSRAKSNL	0.56
RRELLQSATAKPRL	0.57
FVRSIPRPPLRASL	0.57
GTIAKLELQRVAKL	0.58
DFLVDAMDHVMSKL	0.67
AFEGGCPWKPASRL	0.69
STDDWGWYPKPSKL	0.71
GTIAKLELQRVSKL	0.72
AFDGGCPWKPASRL	0.72
TEQDSMKLALISKL	0.74
TEQDSMKLALISKL	0.74
AYRNIVTFAPTARM	0.74
DSYHSGEVYAKSKL	0.75
DSYHSGEVYAKSKL	0.75
AFETGCPLKPASRL	0.75
LAKMQKRTPKFSKL	0.77
QFGLTRNLLINSKL	0.78
VQDRERNSRSLSRL	0.79
ADFMAQYSDSKSKL	0.8
ADFMAQYSDSKSKL	0.8
PTSETAPMFMRAKL	0.81
KNNVELAPQVVSKL	0.83
QKFTPDIFPKPAKL	0.84
VKLSVPIRNSMSRL	0.85
SQPEMVPGGLLSKL	0.85
QDFMAQYADAKSKL	0.87
QDFMAQYADAKSKL	0.87
EALQIPNSKANSKL	0.89

KEDELITPLLISRM	0.89
ADFMAQHSDTKSKL	0.89
ADFMAQHSDTKSKL	0.89
KLIWFVDKPAASKL	0.9
KEDELTTPLLISRM	0.91
KLQKSDRSVAKSKL	0.92
KGGQSEMIPLLSKL	0.92
ECAKLLYSAPQSRL	0.93
ALGLPTGDGSKSRL	0.93
FVSPEDEKSLVSKL	0.94
PVLTQNLWAPHSKL	0.95
PVLTQNLWAPHSKL	0.95
LSQIVEKRSGASRL	0.95
VQQREGRSMSRSRL	0.96
EALHLSEDRQRSRL	0.97
LDPNASSSQHASRL	0.97
EALQIPSNKANSKL	0.98
RRKLVELSRLQSKL	0.99
RRKLVELSRLQSKL	0.99
GETIVVAGGMKSRL	1.01
ECAKIVNSKQRSRL	1.01
QRSRSKTQLPNSKL	1.04
ILPGGSSSSPASKL	1.05
LLPPPKVTTPASKL	1.07
VLPPDEITQLASKL	1.08
EMAKALPQSVASKL	1.08
EQRSRTTKLPSSKL	1.09
SYAARASFPIHSRL	1.09
KALGNLNKNPRSRM	1.09
EPTDSMQPTLLSKL	1.1
GPTDSIQPTLLSKL	1.1
AGRSSNQSTPKSKL	1.16
AHVQTEVERFHAKL	1.21
AGRSSKQSKPKSKL	1.22
SHIQTEAERFRAKL	1.24
CHIQTEAERIHSKL	1.35
CHIQTEAERIHSKL	1.35
SHIQTEAERLYSKL	1.39
14 C-terminal sequences from *P. patens* homologues to *A. thaliana* proteins with minor PTS1 sequences	
IQKWLKNEDTSSTL	0.45
YTIQNSTQGPSRKL	0.46
GGQSLPRPRMKSYL	0.6
REVQSKKVRLRPKL	0.85
VGRSMSRTKGMSKL	0.9
VATIAKLQSEKAKL	0.91
WHKIRIHGPPESKL	1.01
DVHRRQSKPPVSKL	1.07
14 C-terminal sequences from *P. patens* homologues to *A. thaliana* proteins with rare PTS1 sequences	
IIAAVDASYNSSTL	0.41
IIAAVDASYNSSTL	0.41
RQVLSHFQIMTSSL	0.49

DELVSGGLGFASKI	0.49
TFLHEQTLASSSRL	0.58
KSRTDGGKRRPSHL	0.66
LSGSIILRSTLSRL	0.78
WTKRFVFGRDSSRM	0.94

F.2 L. Cross additional data, testing binding of *At*PEX5-C with biotinylated peptides

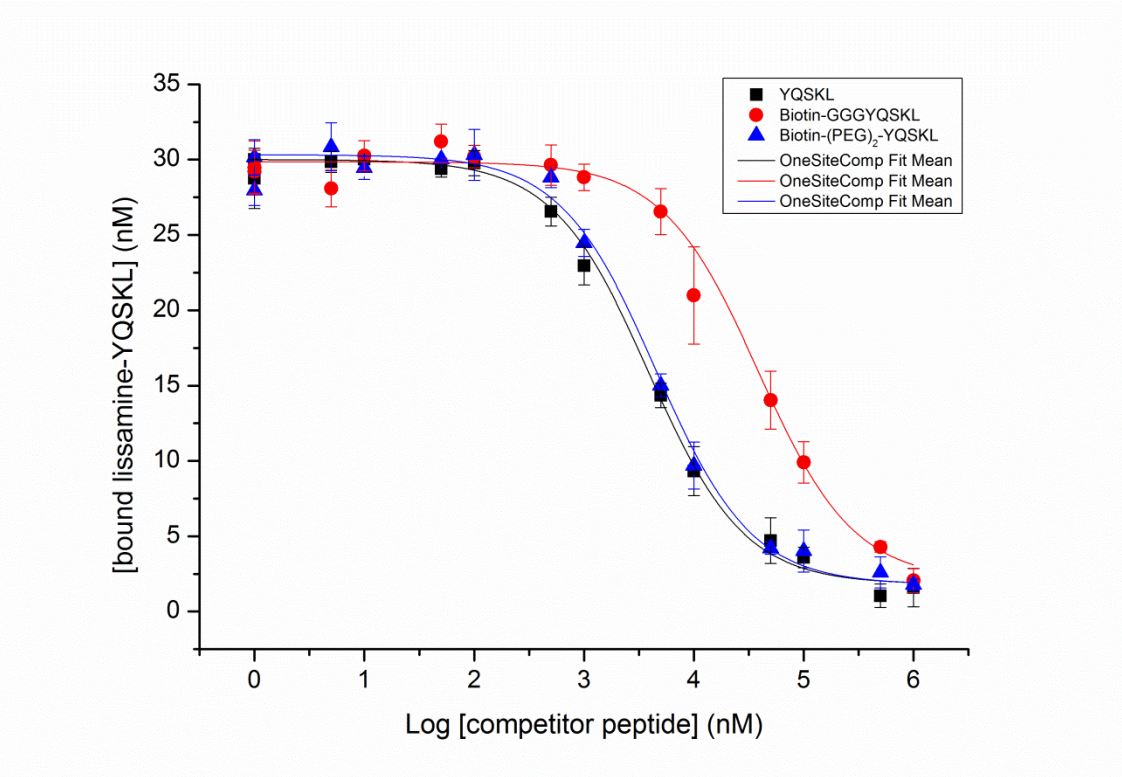


Figure F-1| Fluorescence anisotropy competition assays using a fixed concentration of lissamine-YQSKL (30 nM final concentration), and of AtPEX5 C (200 nM final concentration), and competitor peptide (ranging from 0-1 mM final concentration). Biotin-$(PEG)_2$-YQSKL appears to out-compete lissamine-YQSKL much more effectively than biotin-GGGYQSKL. K_i of unlabelled YQSKL = 22 ± 12 nM. K_i of biotin-GGGYQSKL = 230 ± 140 nM. K_i of biotin-$(PEG)_2$-YQSKL = 26 ± 15 nM. Data were fitted to a one-site competition model using OriginPro 9.1.