Experimental validation of plant peroxisomal targeting prediction algorithms by systematic comparison of *in vivo* import efficiency and *in vitro* PTS1 binding affinity

Nicola S. Skoulding[#1], Gopal Chowdhary[#2,3], Mara J. Deus[2], Alison Baker[4], Sigrun Reumann[2,5], and Stuart L. Warriner[1]

[#]These authors contributed equally to this work.

[1] - School of Chemistry and the Astbury Centre, University of Leeds, Leeds, LS2 9JT, UK.

[2] - Centre for Organelle Research (CORE), Faculty of Science and Technology, University of Stavanger, Richard Johnsons gate 4, N-4021 Stavanger, Norway.

[3] - KIIT School of Biotechnology, Campus XI, KIIT University, I-751024 Bhubaneswar, India.

[4] – Centre for Plant Sciences, School of Molecular and Cellular Biology, University of Leeds, Leeds, LS2 9JT, UK.

[5] - Department of Biology, Biocentre Klein Flottbek, University of Hamburg, D-22609 Hamburg, Germany.

Correspondence to Stuart Warriner, s.l.warriner@leeds.ac.uk, +44 113 343 6437.

Abstract

Most peroxisomal matrix proteins possess a C-terminal targeting signal type 1 (PTS1). Accurate prediction of functional PTS1 sequences and their relative strength by computational methods is essential for determination of peroxisomal proteomes *in silico*, but has proved challenging, due to high sequence variability of non-canonical targeting signals, particularly in higher plants, and low availability of experimentally validated non-canonical examples. In this study *in silico* predictions were compared with *in vivo* targeting analyses

and *in vitro* thermodynamic binding of mutated variants within the context of one model targeting sequence. There was broad agreement between the methods for entire PTS1 domains and position-specific single amino acid (aa) residues, including residues upstream of the PTS1 tripeptide. The hierarchy Leu>Met>Ile>Val at the C-terminal position was determined for all methods but both experimental approaches suggest Tyr is under weighted in the prediction algorithm due to the absence of this residue in the positive training dataset. A combination of methods better defines the score range that discriminates a functional PTS1. *In vitro* binding to the PEX5 receptor could discriminate amongst strong targeting signals whilst *in vivo* targeting assays were more sensitive, allowing detection of weak functional import signals that were below the limit of detection in the binding assay. Together the data provide a comprehensive assessment of the factors driving PTS1 efficacy and provide a framework for the more quantitative assessment of the protein import pathway in higher plants.

Keywords (not in title): PEX5, Fluorescence anisotropy, YFP fusion, peptide, specificity

List of acronyms: aa, amino acid(s); ACX4, acyl-CoA oxidase 4; At, Arabidopsis thaliana; EYFP, enhanced yellow fluorescent protein; Hs, human; PTS1/2, peroxisome targeting signal type 1/2; PWM, position weight matrices; ROS, reactive oxygen species; TPR, tetratricopeptide repeat; Ze, *Zinnia elegans*

Introduction

Peroxisomes are ubiquitous organelles within eukaryotes, responsible for a wide range of intracellular roles which are critical to cell and organism function. Compared to other cell organelles, peroxisomes are very dynamic and metabolically versatile. For example in cotyledons of *Arabidopsis thaliana* and other oil seed plants, a major role of peroxisomes is in mobilisation of storage lipids and conversion to carbohydrates to support early heterotrophic seedling growth. As the cotyledons become photoautotrophic, photorespiration becomes the predominant pathway. Additionally, it is increasingly apparent that peroxisomes are connected into many if not all aspects of plant life, including primary metabolism, hormone synthesis and signalling of reactive oxygen species (ROS) [1]. Proteomic studies from different tissues are revealing new and unexpected peroxisomal capabilities, for example in synthesis of secondary metabolites and in plant defence [2; 3; 4; 5]. Collectively, these roles are of critical importance for plant fitness and productivity, underscored by the severe, sometimes lethal phenotypes of peroxisome biogenesis mutants [6; 7] . Different peroxisome functions are determined by their precise enzyme set which in turn reflects the balance between import and turnover of individual proteins and the organelle as a whole.[1]

Proteins destined for the peroxisomal matrix are typically synthesised in the cytosol with one of two peroxisome targeting signals (PTS1 or PTS2) within their sequence. These are recognised by cytosolic receptors that initiate the import of the cargo protein into the peroxisome. The peroxisome targeting signal type 1 (PTS1) was initially described as a C-terminal motif characterised by the consensus [S/A/C]-[K/R/H]-[L/M] [8; 9; 10; 11; 12; 13] although it is now known that residues outside the tripeptide also contribute to recognition by the cycling receptor PEX5 [12; 14 13; 15]. The PEX5-cargo protein complex interacts with peroxisomal membrane proteins resulting in translocation of the cargo into the organelle matrix; the receptor is then recycled to the cytosol [16]. A second targeting signal of peroxisomal matrix proteins, the PTS2, is located near the N-terminus of cargo proteins and is recognised by a different primary receptor, PEX7. PEX7 acts as an adaptor protein that directly interacts with the so-called long isoform of PEX5 in plants and animals, enabling the two pathways to converge at the peroxisomal membrane [16; 17].

The PTS1 binds to the C-terminal tetratricopeptide repeat (TPR) domain of PEX5 whilst the N-terminal natively unstructured PEX5 domain initiates docking at the peroxisome

membrane and receptor recycling, a process that requires mono-ubiquitination at a conserved N-terminal Cys in mammals and yeast [18; 19 20]. Since this Cys is conserved in plant PEX5 a similar recycling system most likely operates across eukaryotes. High resolution structures provide molecular level information on the interaction between the C-terminal TPR domain of human and trypanosome PEX5 and model PTS1 peptides [21; 22] and full-length PTS1 cargo [23; 24].

Proteomic analyses of peroxisomes have shown that resident proteins have PTSs that can differ significantly from the simple initial consensus pattern of canonical PTS1 tripeptides[2; 3; 4; 5]. However, the technical difficulty of isolating pure peroxisomes makes direct proteomic determination of peroxisomal contents impractical for detailed insight into the variations between species, tissues and as a function of time and environmental stimuli[25]. A clear understanding of this biological system depends on the development of optimised and robust bioinformatics tools that enable the sensitive and accurate identification of functional PTSs and imported PTS1/2 proteins within sequenced genomes. High accuracy PTS1/2 protein prediction algorithms combined with large-scale gene expression analyses allow inference of the proteome of plant peroxisomes in different species, tissues and developmental stages under a variety of different abiotic and biotic stress conditions. Using machine learning methods, two prediction models were developed and evaluated [12]. The position weight matrices (PWM) gave the best results in terms of prediction specificity and sensitivity, the correct inference of novel non-canonical PTS1 tripeptides and the prediction of targeting enhancing upstream residues [12]. According to the PWM model, the 14 C-terminal aa residues of PTS1 proteins contain discriminative properties that are characteristic for plant PTS1 proteins [12]. About 1% of the Arabidopsis gene models (approx. 380 out of 33,000 gene models) have been assigned prediction scores above threshold and, hence, are predicted to be located in peroxisomes [12; 26]. Despite significant progress [12] PTS1 protein prediction is still limited by a number of parameters. For instance, the predominance of canonical PTS1 tripeptides among the >2600 positive example sequences used for model training makes the correct prediction of non-canonical PTS1 domains challenging, and many non-canonical plant PTS1 tripeptides have remained unidentified. As a result, the prediction grey-zone with both (true) PTS1 and non-PTS1 sequences has remained relatively wide (approx. 0.10-0.412).[12]

The true potential of bioinformatics lies in the combination and continuous improvement of computational predictions by experimental validations. The definition of a more precise PWM score range for peroxisome import and determining whether the predicted probabilities of PTS1 proteins for peroxisome targeting correlate with import strength and efficiency is important for model development. Novel peroxisomal candidate proteins are typically validated by *in vivo* experiments in which the full-length proteins are fused to fluorescent reporters, transiently expressed in plant cells and subsequent cellular localisation is observed [2; 3; 4; 5; 12] Potential drawbacks to this approach are; the effect of introducing a tag, which could potentially mask targeting information; non-physiological levels of expression; and the inability to generate quantitative data. A complementary approach is to explicitly measure the binding constants of putative signals with their receptor *in vitro* and to use these thermodynamic parameters to assess if the interactions are strong enough to act as the basis of cargo recognition and therefore import[27; 28]. These thermodynamic data provide rapid, robust and quantitative information about the relative affinities of different sequences to a receptor, but are limited by their reduction of protein import complexity to a simplified two-component system, namely the binding of PTS1 peptides to PEX5. Maynard and Berg[29] measured affinities of model PTS-1 binding peptides for wild type and mutant human PEX5, and deduced relative free energy contributions of binding for a range of natural human PTS1 sequences and sequences selected from a PTS1 sequence library using Hs PEX5 as bait. This study, which considered predominantly 'canonical' PTS1 signals proposed cut-off values for *in vitro* affinity that are required for functional PTS-1 signals[29]. Corresponding studies have not been performed in the plant context and systematic analysis of weaker PTS-1 signals is lacking, making the cross validation of *in vivo*, -*in silico* and *in vitro* methods hard to perform reliably.

In this work we report a systematic study of mutagenised putative PTS1 domains, validating *in silico* predictions by the two independent and complementary methods of *in vivo* targeting studies and *in vitro* determination of PTS1 peptide affinities. This data set allows; (i) more precise definition of the prediction grey-zone, (ii) validation of the predicted, position-specific strength of individual PTS1 tripeptide residues, and (iii) validation of the identity and function of targeting enhancing and inhibitory residues located in the eleven residues upstream of the PTS1 tripeptide. These results define the accuracy, dynamic range and sensitivity limits of the three methods, and advance our understanding of the function of

PTS1 targeting elements and domains. The comparative data raise intriguing questions regarding how cytosolic plant proteins are able to evolve extremely weak non-canonical PTS1s for peroxisome targeting while competing with native canonical PTS1 proteins for PEX5 binding.

Results

In the PWM-based PTS1 protein prediction model, each of the 20 possible aa residues of the C-terminal 14-aa sequence is assigned a position-specific score that indicates whether a specific residue at a particular sequence position is predicted to enhance (more positive score) or reduce peroxisome targeting (more negative score) and to what extent (Suppl. Table 1). The total prediction score represents the sum of the position-specific PWM scores of the C-terminal 14 aa residues[30]. Until now, however, quantitative experimental data validating the predicted targeting efficiency of single PTS1 domain residues (of PTS1 tripeptides or upstream residues) and of entire PTS1 domains have remained scarce, resulting in a relatively imprecise definition of the threshold for peroxisome targeting. .

To minimize secondary effects such as aa residue interdependency and secondary structure, an effect analysis of specific single and multiple point mutations introduced either into the PTS1 tripeptide or into the upstream domain is best investigated in the context of one specific constant model sequence. The *Zinnia elegans* acyl-CoA oxidase 4 (ZeACX4) sequence was considered suitable and representative because (i) the PTS1 domain construct was weakly targeted to peroxisomes in onion epidermal cells, as determined by *in vivo* subcellular targeting analyses [12] (and Fig. 1a), (ii) the sequence terminated with a non-canonical, experimentally validated PTS1 tripeptide (SRV>, ">" designates the extreme C terminus), (iii) the domain upstream to the PTS1 tripeptide contained predicted enhancer elements and (iv) the PTS1 domain had been assigned a relatively low prediction score below threshold in the prediction grey-zone.

Validation of the PTS1 protein prediction model by semi-quantitative *in vivo* subcellular targeting analyses using mutagenized PTS1 domain constructs

Enhanced yellow fluorescent protein (EYFP) fused to the C-terminal decapeptide of ZeACX4 VAKTTRPSRV> remained cytosolic after two days of expression in onion (*Allium cepa*) epidermal cells (Fig. 1a1,2). After prolonged (7 d) cold incubation the reporter fusion was

detected in organelle-like punctuate structures that coincided with DsRed-SKL labelled peroxisomes in double transformants (Fig. 1a3,b, Table 1 [12]). The peroxisome targeting efficiency of the model sequence was referred to as weak (detectable only after several days). The positive control EYFP-PTS1 (EYFP extended C-terminally by a PTS1 decapeptide terminating with CKI>, Fig. 1c) labelled peroxisomes 18-24 h post transformation (p.t.) (referred to as strong peroxisome targeting) and EYFP alone was cytosolic at all time points (Fig. 1d).

Initial experiments focused on aa mutations at position -1 (The aa residues considered for the PWM model are numbered -1 to -14 with position -1 referring to the C-terminal residue). The PWM prediction score matrix indicates that the six aa residues that have been experimentally determined to occur in plant PTS1 tripeptides at position -1 ([LMIFVY]) possess differential predicted targeting strengths, ranging from high for Leu (PWM score=0.66) and Met (0.64), followed by Ile (0.33) to weak for Phe, Val and Tyr (-0.09 to -0.016) [12], (Suppl. Table 1). Consistent with the increase of the PTS1 prediction score for the mutagenized sequence SR(V-to-I)> (from 0.216 to 0.664, Table 1), peroxisome targeting of the corresponding EYFP construct was detected at all three time points p.t. (18-24 h, 48 h and 7 d), as shown in single transformation without image modifications of brightness and contrast (Fig. 1e). Peroxisome targeting was confirmed in double labelling experiments using DsRed-SKL as peroxisomal marker (Suppl. Fig. 1a). Hence, the single point mutation V-to-I (pos. -1) converted the weak domain into a strong PTS1 domain as predicted (Fig. 1e, Table 1).

Similarly the mutation SR(V-to-M)> significantly enhanced peroxisome targeting from weak to strong efficiency (Fig. 1f, Table 1, Suppl. Fig. 1b). The significantly higher PTS1 prediction score of Met at pos. -1 (PWM score=0.66) compared to Ile (0.33) suggested that both strong PTS1 tripeptides might still differ in peroxisome targeting efficiency if investigated at sufficiently high resolution. Hence, reporter gene expression and fusion protein targeting was investigated at very early time points (4 h, 8 h, 12 h and 24 h) after biolistic bombardment. While reporter gene expression was hardly detectable until 8 h p.t., EYFP expression and fluorescence became visible 12 h p.t. for both constructs (SRM> and SRI>) without significant differences in cellular fluorescence intensity (Suppl. Fig. 1 d2 and e2). Significant differences in peroxisome targeting, however, could be resolved for both PTS1s. While the reporter fusion terminating with SRI> remained fully cytosolic in all cells investigated 12 h p.t. (Suppl. Fig. 1 e2), the corresponding fusion protein terminating with

SRM> became clearly detectable in peroxisomes against some yellow fluorescent background of newly synthesized EYFP and, hence, was assigned very strong peroxisome targeting efficiency (Suppl. Fig. 1 d2). This difference in cytosolic versus peroxisomal targeting was consistently found in nearly all transformed cells and reproducible in independent experiments.

The PTS1 tripeptide alteration SR(V-to-Y) marginally reduced the PTS1 domain prediction score from 0.216 to 0.173 (Table 1). Contrary to the expected maintenance or reduction of weak peroxisome targeting, the SRY> construct targeted peroxisomes with moderate efficiency, as indicated by the detection of peroxisome targeting 48 h p.t. (Fig. 1g, Suppl. Fig. 1c). To verify the specificity of protein import into peroxisomes in the given experimental *in vivo* system, we further investigated one predicted deleterious position -1 mutation. The point mutation SR(V-to-K)> reduced the PTS1 domain prediction score slightly by 0.1 (from 0.216 to 0.119, Table 1), and positively charged aa residues have not been identified at position -1 in plant PTS1 tripeptides. Indeed, the reporter fusion terminating with SRK> remained cytosolic even at maximum sensitivity of detecting weak peroxisome targeting (7 d p.t., Fig. 1h).

Next the effect of point mutations introduced into the model sequence terminating with SRV> at position -2 on peroxisome targeting was tested. According to present knowledge, position -2 shows highest flexibility in PTS1 tripeptides with 16 different aa residues being allowed in plant PTS1 proteins in combination with strong PTS1 residues ([SA]y[LMI]>) at the other tripeptide positions[12; 30]. Since Arg (R, 0.46) and Lys (K, 0.44) are assigned the highest PWM prediction scores at position -2 and the highest and nearly identical peroxisome targeting strength, possible differences in peroxisome targeting efficiencies were unlikely to be resolved for SRV> and SKV> by *in vivo* subcellular targeting analyses. Therefore the mutation of SRV> to SNV> (score decrease from 0.216 to -0.229, Table 1) was examined. Weak peroxisome targeting could still be detected for this reporter fusion in onion epidermal cells (Fig. 1i, Table 1), which was comparable to the weakly targeted SRV> fusion protein. To verify the specificity of PTS1 protein import and experimentally define the prediction score range limit for peroxisome targeting, SRV> was changed to STV> in the model sequence. Consistent with the very low PTS1 domain prediction score of -0.401, (Table 1)

the reporter fusion terminating with STV> was no longer targeted to peroxisomes (Fig. 1j). These experimental data confirmed the high specificity of peroxisomal protein import in the chosen *in vivo* system and assisted in defining experimentally the lower limit of PTS1 domain prediction scores for peroxisome import (PWM score=-0.2, Fig. 2).

Among all 12 possible residues allowed at pos. -3, Ser is assigned the maximum peroxisome targeting strength. The PTS1 tripeptide mutation to PRV>, which reduced the PTS1 domain prediction score from 0.216 to -0.135 (Table 1), abolished any reporter fusion targeting to peroxisomes (Fig. 1k).

Next, the effect of multiple point mutations introduced into the PTS1 tripeptide of the model sequence was investigated. The dual tripeptide mutation from SRV> to SNM> significantly enhanced peroxisome targeting from weak to moderate strength, as fluorescent peroxisomes became detectable 48 h p.t. (Fig. 1l). The result fully agreed with the significant increase in PTS1 domain prediction score (from 0.216 to 0.523, Table 1). Conversely, the dual tripeptide mutation from SRV> to SNY> abolished the weak peroxisome targeting of the model sequence (Fig. 1m). The experimental result was fully consistent with the significant decrease in PTS1 domain prediction score (from 0.216 to -0.272, Table 1).

Finally, potential enhancing function of upstream residues on peroxisome targeting was investigated. The upstream domain of the model sequence (VAKTTRP-SRV>) contained two basic residues (Lys, Arg) and one Pro residue, all of which are generally considered to act as targeting enhancing elements in plant PTS1 sequences[31; 32]. First, the two basic residues (K position -8; R position -5) and one Pro (P) residue (position -4) were exchanged to Gly (G) residues, thereby lowering the PTS1 domain prediction score slightly from 0.216 to 0.073. Similar to the original sequence, the reporter fusion terminating with the mutated decapeptide (VAGTTGG-SRV>) remained detectable in peroxisomes 7 d p.t. (Fig. 1n). In contrast, changing the two basic upstream residues to acidic residues (VAETTDP-SRV>), which further lowered the PTS1 domain prediction score (to 0.011), completely abolished peroxisome targeting (Fig. 1o). Similarly, when mutating the single Pro residue at position -4 to Asp, thereby reducing the PTS1 domain score from 0.216 to 0.045, peroxisome targeting was completely abolished (Fig. 1p, Table 1).

Posterior probabilities facilitate the interpretation of the absolute prediction scores and quantify the probability for peroxisome targeting, ranging from zero (0% probability) to one (100%), with 0.5 corresponding to the prediction threshold of 50% probability for peroxisome targeting[12]. In addition to the initial standard posterior probability[12] , a so-called balanced probability value has been calculated for the PWM model [26] by assuming an equal variance of positive (PTS1) and negative (non-PTS1) example sequence scores, which leads to a broader intermediate probability value range and higher targeting probability values for sequences differing from the majority of positive examples, i.e., non-canonical and low-abundance peroxisomal proteins. On the downside of increased sensitivity, the fraction of non-peroxisomal proteins with probability values >50% increases substantially and leads to a higher proportion of false positive predictions. To better visualise the relationship between *in vivo* targeting and *in silico* prediction of targeting signals, the experimentally tested sequences were grouped into four categories (cytosolic, weak, moderate or strong peroxisomal targeting) and plotted against the PWM score, standard posterior probability and balanced posterior probability scores (Fig. 2, Table 1 and Suppl. Tables 2 & 3). The analysis reveals a clear positive correlation between PWM score and experimentally determined strength of targeting, although there is overlap of scores between categories that can be distinguished experimentally (Fig. 2a, Suppl. Table 2 & 3). the standard posterior probability does not sensitively discriminate between sequences with different *in vivo* determined targeting strengths (Fig. 2b), the balanced post posterior probability is superior in its discrimination ability (Fig. 2c), particularly between moderate and cytosolic proteins, which are poorly distinguished using the other methods (correlation matrices are shown in Supplementary Information Table 3). Some weakly peroxisome-targeted sequences and one moderately peroxisome-targeted sequence (SRY>) fall below the 50% threshold of the balanced posterior probability (Fig. 2, Suppl. Table 2), indicating that iterative approaches combining bioinformatics and experimental research are required in the future to further improve the prediction ability of non-canonical PTS1 sequences

Determination of PTS1 peptide binding affinities to AtPEX5

In order to better understand the thermodynamics of binding between Arabidopsis (At)PEX5 and a range of potential targeting sequences, a series of *in vitro* experiments were performed. Two N- terminally His tagged versions of AtPEX5 were expressed and purified from *E. coli*:

the full length protein (aa 1-728, termed PEX5) and an N-terminal truncation comprising aa 340-728 (PEX5C) [17](Fig. 3a,b). The latter is equivalent to the human PEX5 construct used to determine the thee-dimensional structure of human (Hs)PEX5[21]. The PTS1 domains to be investigated in this study were prepared by solid phase peptide synthesis and their binding affinities to AtPEX5 were determined using a fluorescence anisotropy-based assay (Supplementary Information, Section 3)[27; 33]. The assay determines the amount of a fluorescently labelled tracer peptide (in this case the tightly binding pentapeptide YQSKL labelled at the N-terminus with Lissamine^{TM} rhodamine) associated with the receptor by virtue of the slower tumbling rate of the fluorophore when it is bound to PEX5 (higher anisotropy). The limits of anisotropy of the tracer is first determined by direct titration of the protein (e.g. PEX5C) into a fixed concentration of tracer (Suppl. Fig. 2) and the K_d of the tracer was determined by titration of the tracer solution into the protein (Fig. 3c, Suppl. Fig. 3). Fitting to the appropriate equations for a 1:1 binding model (see Methods) showed the K_d of the tracer peptide YQSKL to be virtually identical for the two receptor constructs (as 4.0 ± 0.5 nM for PEX5C and 4.5 ± 1.2 nM for PEX5, in good agreement with the value of 3.1 nM reported for the truncated human PEX5[21]). Once the affinity of the tracer to its receptor is known, the binding of a range of unlabelled sequences can be determined by using a competition assay in which unlabelled peptides compete to displace the tracer from the PTS1 binding site on PEX5. The concentration of the peptide of interest required to displace 50% of the initially bound fluorophore from PEX5 (IC_{50}) can be mathematically combined with the known affinity of the tracer for PEX5 to give the binding constant (expressed as K_i) for the sequence of interest. Example competition curves are shown in Fig. 4 for the peptide VAKTTRPSRV> and variants ending in M, I and Y binding to PEX5C. Affinity of both full length PEX5 and PEX5C for a total of 19 peptides was determined and are shown in Table 2 and Supplementary Figs 4 and 5. The peptides tested showed a range of K_i values from 100 nM to undetectable (>100 μM). No significant differences in binding affinity of individual peptides to PEX5 compared to PEX5C were observed (Table 2).

In accordance with the critical function of the most C-terminal residue in PTS1 tripeptides in peroxisome targeting *in vivo* (see above), initial studies focused on point mutations at position -1. The affinity of the original model peptide of ZeACX4 (VAKTTRPSRV>) to PEX5 and PEX5C was below the detection limit with K_i>100 μM (Table 2, Entry 1; Fig. 4). Comparing the series VAKTTRPSRX>, Leu and Met both gave rise to high affinity binding

to PEX5C (K_i=1-3 μM) with Leu marginally better (Table 2, Entries 2 & 3). The mutation to Ile in the -1 position resulted in an order of magnitude decrease in binding affinity (15-21 μM), consistent with the PWM model prediction scores (Table 2, Entry 4; Fig. 4). As predicted and consistent with the *in vivo* data, the mutation to Tyr caused a further 3-fold decrease (25-47 μM). Contrary to the PWM model predictions but fully consistent with the semi-quantitative *in vivo* peroxisome targeting analyses, the SRY> peptide showed higher PEX5 binding affinity and moderate peroxisome targeting efficiency compared to the original model peptide terminating with SRV> (undetectable PEX5 binding, weak *in vivo* peroxisome targeting, Figs. 1 and 4, Tables 1 and 2, Entry 5). The thermodynamic results demonstrate the preference of Arabidopsis PEX5 for long hydrophobic side chains at position -1 (L, M) since both the branched Ile and especially Val significantly reduced PEX5 binding. Consistent with this conclusion, V-to- L/M mutations increased PEX5 binding affinity also for VAKTTRPSN(V-to-M) and the shorter peptides YQSK(V-to-L) (Table 2, Entries 6 & 11; 17 & 18).

To investigate the effect of position -2 mutations on PEX5 binding affinity, multiple mutations were introduced into the original model peptide because the affinities of both the SRV> and the SNV> decapeptide for PEX5 were below detection limit (Table 2, Entries 1 & 6). Fully consistent with the PWM predictions, the mutation of Arg at position -2 to Lys combined with the V-to-L mutation at position -1 in the first model peptide maintained the high binding affinity of 1-2 μM (VAKTTRPSRL>, PWM score: 1.043; VAKTTRPSKL>, PWM score: 1.031), showing equivalence of Arg and Lys (position -2) in terms of PEX5 binding (Table 2, Entries 2 & 10). Asn in position -2, however, greatly decreased the binding affinity when combined with the favourable Met in the -1 position about 20-fold (VAKTTRPSRM>, K_i=1.8-3.1 μM; VAKTTRPSNM>, K_i=48-72 μM) (Table 2 , Entries 3 & 11). This result is also consistent with the reduction in PWM score (from 1.02 to 0.58) and the *in vivo* data from strong to moderate peroxisome targeting (Fig. 1, Table 1, 2). SNM (in context with different upstream residues) had been previously characterized as a functional non-canonical PTS1 in plants. [12]

Upstream residues are known to be able to enhance the function of weak PTS1 tripeptides and Pro is found reasonably frequently in positions -4 and -5 of natural plant PTS1 proteins[12]. Therefore the effect of substituting the Pro at position -4 with Gln was examined. VAKTTRQSRL> bound an order of magnitude tighter to PEX5 than VAKTTRPSRL (Table

2, Entries 12 & 2)), showing that, at least in this specific context of the strong PTS1 tripeptide SRL, Pro did not show an additional targeting enhancing effect. Substitution of Pro with Gln in the peptide VAKTTRQSRV> (Table 2, Entries 1 & 9), however, significantly increased binding from undetectable to K_i=50-53 µM, similar to VAKTTRPSNM> (K_i=48-72 µM) (Table 2, Entry 11). Substitution of the two basic residues in the peptide VAKTTRQSRL> with neutral Ala residues, singly and in combination, resulted in a decrease in affinity of binding that was additive (Table 2, Entries 13,14 & 15), confirming the importance of upstream basic residues.

Discussion

Despite molecular details on binding of peptides and cargo proteins to PEX5[21; 22; 23; 24], mutational studies of PTSs [11; 14; 15]and the availability of an increasing catalogue of peroxisomal proteins from proteomic studies [2; 3; 4; 5; 34], it remains difficult to predict reliably the identity of non-canonical PTS1 domains. Therefore, improved informatic tools that can accurately predict the peroxisomal complement of organisms from sequenced genomes would be very useful. In addition, it is desirable to predict the strength of peroxisome targeting for PTS1 proteins of interest to infer, for instance, quantitative peroxisome targeting or dual protein localization in different subcellular compartments for proteins with multiple targeting signals. Also, understanding potential variations in PTS targeting strength can give insight into regulation of the composition of the peroxisome proteome and the evolution of PTSs to endow peroxisomes with new capabilities.

Comparison of *in vivo* and *in vitro* experimental data with *in silico* predictions

Overall the three methods deployed in this study agreed remarkably well, even at highest resolution of the targeting/affinity strength of position-specific single aa residues of the PTS1 tripeptide. Fig. 5 shows a graphical representation of the relationship between the peroxisome targeting prediction by the PWM model, the measured binding affinity by fluorescence anisotropy and the strength of targeting as determined semi-quantitatively by *in vivo* assay.

Sequences that behaved as strong PTSs *in vivo* (giving rise to fluorescent peroxisomes within 24 h) such as VAKTTRPSRM> and VAKTTRPSRI> had high PWM scores and balanced post posterior probabilities and bound both PEX5 and PEX5C with micromolar affinity. Both *in vivo* and *in vitro* binding studies gave the same hierarchy of preference for residues in the

terminal (-1) position L>M>I>Y>V, and this matched well to the individual scores for these residues in the same position (Supp. Table 1), and is in good agreement with the aa residue frequency of naturally occurring Arabidopsis PTS1 proteins[12].

The targeting strength of Tyr at pos. -1 might have been underestimated due to the complete lack of Tyr at this position in any of the 2600 positive example sequences of plant PTS1 tripeptides used for model training. The first plant PTS1 protein carrying Tyr at pos. -1 was only identified relatively recently. [35] The experimental data suggest that the PWM score for this sequence should be 0.2 to 0.4 units higher to bring the result in line with sequences with similar affinities and biological import properties. This Tyr example stresses the importance of identifying novel Arabidopsis proteins carrying novel residues in their non-canonical PTS1 tripeptides since these residues are often conserved in orthologs of diverse plant species and altogether significantly improve residue representation in the large dataset of positive example sequences against predominance of canonical PTS1 triptide residues. A more precise evaluation of the effect of Tyr at pos – 1 will require further investigation of this residue in a wider range of sequence contexts.

The peptide terminating in SRV> was below the binding detection limit for the *in vitro* assay. The higher sensitivity of the *in vivo* system in detecting (weak) peroxisome targeting (Fig. 1, 2) compared to the thermodynamic assays, is remarkable and might indicate that additional components such as binding partners and/or posttranslational mechanisms enhance the affinity of non-canonical PTS1 tripeptides for PEX5 *in vivo*. Conversely, the *in vitro* binding assays were able to discriminate between strong targeting peptides that were not able to be resolved by the *in vivo* assays (Figs. 1 and 5, Table 2), revealing complementary information and an important advantage of thermodynamic binding studies.

Introduction of Pro at position -3 yielding the tripeptide PRV> was detrimental in both experimental systems, as predicted by the algorithms. At first glance, this result appears surprising because Pro is a well-known residue in functional PTS1 tripeptides such as PRL and PKL. The most likely explanation for this apparent discrepancy is that low-abundance residues such as Pro (pos. -3) only yield functional PTS1 tripeptides if combined with two high-abundance strong PTS1 residues such as Arg/Lys (pos. -2) and Leu/Met/Ile (pos. -1)[12]. Two low abundance residues such as Pro (pos. -3) and Val (pos. -1) for instance in PRV> may not possess high enough affinity to PEX5 to allow import into the peroxisome matrix.

Similarly, the STV> peptide was non-peroxisomal in this study, still consistent with the fact that Thr has been characterized as a plant PTS1 tripeptide residue for STL> [12].

Clear evidence was also obtained for the importance of aa residues upstream of the PTS1 tripeptide in modulating PEX5 affinity and peroxisome import efficiency. Pro occurs at position -4 with reasonable frequency in natural PTS1s[12; 31]. In the *in vitro* experiments the most significant effect was changing the Pro at position -4 to Gln which increased the affinity by a factor of 10 (Table 2). For shorter pentapeptides, however, only a very small effect on the binding affinities was observed (compare YQSKL> and YPSKL>, Table 2). It is possible that the cis-trans isomerisation of Pro could result in a conformation of the backbone within the longer decapeptide which does not favour receptor binding whereas the structural change does not affect binding in the shorter sequence context. Replacement of the two basic residues at position -5 and -8 reduced binding affinity in an additive fashion. *In vivo*, replacement of these residues with neutral ones had no detectable effect but acidic residues were clearly deleterious. Taken together, the results of upstream residue mutations confirmed the targeting enhancing role basic and Pro residues compared to the generally inhibitory role of acidic residues upstream of PTS1 tripeptides, as reported previously. [32]

The experimentally determined threshold for peroxisome targeting appears to be near 0.15-0.05 since both weakly peroxisomal and cytosolic constructs are located in this prediction grey-zone, which is now much better defined. Except for one apparent outliner (SNY>), four mutated model sequences with PTS1 scores below 0.05 were cytosolic, strongly suggesting that this is a realistic threshold to delineate experimentally the lower limit of the prediction grey-zone.

Implications for cargo binding to PEX5

It has recently been proposed that PEX5 from Pichia undergoes redox regulated disulfide bond formation at the conserved N terminal Cys which alters the affinity of the receptor for its cargo. [36] One surprising observation is that the binding affinity of all the peptides tested in the present study was, within experimental error, identical for both truncated and full length PEX5. Since the N terminally truncated PEX5C lacks this redox sensitive Cys such a mechanism would appear not to be relevant in the context of the binding of short peptides in our experimental system. It should be noted that there are several known cases where residues outside the targeting sequence also contribute to receptor binding affinity. For

instance, human alanine-glyoxylate aminotransferase (AGT), which has a non-canonical PTS1 (KKL>), binds much more tightly to HsPEX5C than the equivalent peptide[24]. The X-ray crystal structure of AGT in complex with HsPEX5C revealed a folded and enzymatically active dimer with each subunit bound via its PTS1 to PEX5. In addition to the interaction of the PTS1 tripeptide with the central funnel formed from the TPRs, an extended interface between the C-terminal domain of AGT and the PEX5 surface was observed. While residues immediately upstream of the PTS1 contributed to binding, there were also contributions from more distant residues. Further, residues that affected AGT folding, even to a minor extent, disrupted the interaction and therefore the import[24]. The other structure where a complex between a full length cargo protein (mSCP2 which has a canonical PTS1) and PEX5C is known [23]shows a complete lack of conservation of interactions outside the PTS1[24]. This, together with the reports that certain proteins, such as catalase, with non-canonical PTS1 make additional contacts to the region of PEX5 outside of the TPR domain [37] may make the prediction of 'weak' PTS1s by only bioinformatic or experimental analysis of the C-terminal region very challenging. The more extensive use of biophysical tools to measure quantitative binding constants for a range of recombinant peroxisomal proteins and full length and truncated PEX5 constructs may help to address these questions.

Towards mechanistic and quantitative models of import

A simple pre-equilibrium model, in which the concentration of cargo loaded PEX5 determines the likelihood of import, requires the concentrations of the cargo protein or its receptor in the cytosol to be close to the K_d for binding. It has been suggested that, consequently, proteins with lower expression levels may well have evolved stronger PEX5 binding sequences to offset their low abundance[28]. In contrast, the most abundant plant peroxisomal enzymes, generally carry canonical PTS1s of high peroxisome targeting strength[12; 31]. It is notable that both bioinformatics and *in vivo* measurements show that protein import into the peroxisome can be observed using sequences that have an affinity for the receptor that exceeds 100 µM *in vitro*. Given that it is unlikely that either PEX5 or the cargo protein generally reach this level of expression, these detailed thermodynamic insights pose interesting questions about the underlying mechanistic details of the import process.

Previous work [28; 29; 33]suggested that the C-terminal peptide motifs tend to have *in vitro* binding affinities in the sub micromolar range, although some examples with 10 fold weaker

affinity were also observed. An affinity limit of ~500 nM for import competent sequences was proposed based on the measured affinities of model peptides for human PEX5 and two pathogenic mutants, and deduced binding energies of native PTS1 sequences[29] and a dataset of PTS1 sequences selected from a yeast 2 hybrid library using human PEX5 as bait[10; 29]. These observations are markedly different to those associated with import competent systems in this study, with import being observed for protein tagged with PTS-1 sequences that have significantly weaker affinities *in vitro* than previously reported. While noting that the experimental data in the earlier studies were obtained with the human PEX5 protein, the high degree of homology between the PTS-1 import apparatus of eukaryotes means that such different observations are hard to reconcile on this basis. However it is worthy of note that some natural human PTS1 peptides have higher K_ds and correspondingly lower calculated binding energies[28] than the previously proposed threshold[29]. In the present study a range of non-canonical PTS-1 sequences with predicted weaker targeting efficiency were systematically tested for their *in vitro* binding and explicitly tested for their *in vivo* targeting. This has enabled more light to be shed on the precise limits of the targeting peptide affinities that can actually drive import. Nevertheless the weak *in vitro* binding of some of the PTS-1 sequences which are import competent still poses interesting questions about the precise mechanistic details of the import process.

One possibility is that other factors may influence the overall magnitude of the binding constants within the import system, although not the fundamental rank order for effects of individual residues. For instance, PEX5 interacts with PEX7 in the cytosol, and both PEX5 and many of its cargoes may exist as oligomers allowing for multivalent interactions to occur that might alter binding constants measured in a simplified system. In our hands (and consistent with the data reported for the truncated human PEX5[20]) at the low protein concentrations used in these assays 1:1 binding models provided good fits to the observed data, although it was noted that at much higher PEX5 concentrations deviations from idealised 1:1 binding curves started to be observed consistent with the presence of higher order oligomers affecting the equilibria being studied.

The *in vivo* data also show that proteins with weaker PTS1s take longer to accumulate in the peroxisome. The *in vivo* long-term expression studies resemble pulse chase experiments in the sense that protein synthesis primarily occurs within the first 24 h p.t. during cell incubation at room temperature whereas protein degradation is slowed down by cold

incubation and import seems to occur gradually over this time. In a model in which reversible binding to PEX5 is more rapid than import, the fraction of any given cargo bound to the receptor is determined by the ratio of the products of the individual K_ds with the individual protein concentrations. The slower import of more weakly targeted proteins is hence consistent with two possible import mechanisms: either the weaker binding affinity of the non-canonical PTS1s results in only a small fraction of the cargo being imported at any time or (under these experimental conditions) the weaker PTS1s are only imported after endogenous proteins with canonical PTS1s have been quantitatively imported and eliminated as competing cargo from the cytosol. However, in either case, the strength of the PTS1 determines the priority of the protein for import. In some situations slow import may be desirable, if a protein requires assembly and maturation steps in the cytosol as proposed for catalase.[38] Importantly, at all PTS1 tripeptide positions, single point mutations (SR(V-to-K), S(R-to-T)V and (S-to-P)RV) completely abolished peroxisome targeting *in vivo*, demonstrating that peroxisome import is specific in the experimental system.

The wide range of *in vitro* binding affinities determined for the strong targeting sequences, all of which show exclusive peroxisomal localisation, may provide further evidence for the plant's requirement to control the priority for import of peroxisomal proteins within the context of a complex and ever changing expression profile. In the physiological situation, changes in expression (for example due to circadian rhythm, tissue differentiation or stress situations) will alter the composition of the pool of proteins competing to be imported into peroxisomes. Understanding the processes governing competitive import will require detailed understanding of the kinetic parameters of the import system and measurement of the steady state pools of cargo and receptor, which will be determined by the rates of protein synthesis, cycling rate between cytosol and peroxisome, and turnover. In addition the extent to which these processes are operating at equilibrium would need to be determined to allow development of mathematical models of import. For example, one explanation for the import of weakly binding PTS1s may be that the binding of the cargo to monomeric PEX5 is not reversible within the import cycle. If subsequent steps in the cycle are fast and operationally irreversible (such as creation of protein complexes with a significantly higher affinity for the cargo, or even import itself) then a more relevant measure of PTS1 strength would be the k_{on} rate constant for binding. The correlation between binding constant and PTS1 strength could hence arise because the equilibrium constants are simply reflecting the underlying on-rate.

Evolution of peroxisome targeting signals

Genome size expansion in multicellular complex organisms also increased the absolute number of nuclear-encoded proteins targeted to subcellular organelles. While the N-terminal targeting signals for mitochondria, plastids and the secretory pathway generally evolved by exon shuffling, the relatively short C-terminal PTS1 appears to be able to evolve by random point mutations of 3'coding regions, alternative splicing and ribosomal read-through of stop codons[39; 40; 41]. Indeed, phylogenetic analysis suggested, and experimental analyses validated, that cytosolic and mitochondrial proteins of green algae and mosses can slowly evolve non-canonical and subsequently canonical PTS1s in higher plants to facilitate peroxisome targeting[26; 42]. Hence, cytosolic proteins that have entered this evolutionary track and initially possess extremely weak affinity to PEX5 must be given an opportunity of being successfully imported into peroxisomes, at least under some specific circumstances. Peroxisome import then offers a selective advantage, thereby increasing organismal fitness and propagation which, in a positive feed-forward spiral, further advances and accelerates C-termini evolution into weak non-canonical and ultimately strong canonical PTS1s. This import capability of newly evolving peroxisomal cargo of lowest PEX5 affinity is difficult to envisage in a model where proteins with strong PTS1s are constantly synthesized and saturate the import machinery. The import competency of PTS1 cargo with low PEX5 affinity *in vitro* revealed in this study provides a pathway for such evolutionary improvements even if the fundamental mechanistic details remain unclear.

Experimental Procedure

Peptides were prepared using standard Fmoc based peptide synthesis strategies [43] using 2-chloro-trityl linked solid supports which were purchased with the C-terminal residue already loaded. Standard side chain protection was employed: Arg (Pbf), Asn & Gln (Trt), Glu & Asp (OtBu), Lys (Boc), Ser, Thr & Tyr (tBu). Coupling cycles were performed in dimethylformamide, using 5 eq. of Fmoc protected aa activated with 5 eq. HCTU (O-(1H-6-chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate) and 10 eq. of di-isopropylethylamine . Fmoc deprotection was performed with 20% piperidine in DMF. Following assembly of the sequence the peptide was cleaved from the resin using a cocktail of CF_3CO_2H /H_2O, triethylsilane (95:2.5:2.5). When sequences contained Met residues an additional 1% ethanedithiol was introduced and the solution degassed with nitrogen prior to use to prevent sulfoxide formation. The cleavage solutions were concentrated and crude peptide was isolated by precipitation from diethyl ether and purified by preparative HPLC. Fluorescently labelled YQSKL was prepared by coupling the N-terminus of the peptide with lissamine sulfonyl chloride prior to cleavage. Detailed procedures and peptide characterisation are reported in the Supplementary Information.

Recombinant PEX5 and PEX5C were prepared as described in Lanyon-Hogg et al.[17]

Fluoresence Anisotropy assays we performed in 384 well microtitre plates (Black Perkin Elmer Optiplates) as follows. Five solutions were prepared [A: FA Buffer (HEPES (20 mM), NaCl (150 mM), pH 7.5); B: Blocking solution: FA buffer containing 0.32 mg/ml of porcine gelatine; C: 12 point dilution series of test peptide in FA buffer (4 mM-20 nM); D: Fluorescent Tracer solution: 120 nM solution of lissamine-YQSKL in FA buffer; E: 800 nM solution of PEX5 in FA buffer. 80 µl of Solution B was added to all wells and the plate sealed and incubated overnight. 70 µl of the blocking solution was removed from each well and the dilution series of the test peptide solution (10 µl per well) added across 6 rows of the plate. Fluorescent Tracer solution (10 µl per well) was then added to 3 rows and FA buffer added to the other 3 rows to act as blanks. Finally PEX5 solution (10 µl per well) was added to all 6 rows. The plate was incubated at 25 °C with linear shaking for 20 min and then read using a Perkin Elmer Envision Plate reader using the following optics: Excitation filter 531 nm (25 nm bandwidth) 555 nm polarised dichroic mirror, emission was detected in 2 separate

channels each with 595(60) nm filters but with orthogonal polarisation (S and P polarisers). 30 flashes were used per measurement. The instrument response factor (g value) was set to 1 on the instrument. The data were blank corrected and processed to give a blank corrected anisotropy and the data processed as detailed in the Supplementary information Section 3 to give an IC_{50} value for the competition experiment which was combined with the K_d of the Tracer to give the K_i, the binding constant for the unlabelled peptide.

Transient Import

In *in vivo* subcellular targeting analyses, the C-terminal 10 residues of the wild-type model sequence from Zinnia ACX4 and of mutagenized variants thereof were fused to the C-terminus of EYFP by PCR using an extended reverse primer (see Suppl. Table 4) and subcloned into the plant expression vector pCAT under control of a double 35S cauliflower mosaic virus promoter [44] and sequenced. For labeling of peroxisomes in double transformants, DsRed-SKL was used[45; 46]. Onion epidermal cells were transformed biolistically as described[46]. The onion slices were placed on wet paper in Petri dishes, stored at room temperature in the dark for approx. 16 h, and analyzed directly (referred to as 18-24 h p.t.) or after additional tissue incubation at 10°C in the dark for approx. 1 d (referred to as 48 h p.t.) to 6 d (referred to as 7 d p.t.). Fluorescence image acquisition was performed on a Nikon TE-2000U inverted fluorescence microscope equipped with an Exfo X-cite 120 fluorescence illumination system and single filters for YFP (exciter HQ500/20, emitter S535/30) and DsRed (exciter D560/40X, emitter D630/60M). The images were captured using a Hamamatsu Orca ER 1394 cooled CCD camera. Standard image acquisition and analysis was performed using Volocity II software (Improvision) and Photoshop.

Acknowledgements

This work was supported by a targeted priority studentship in chemical biology from the BBSRC (to NS), a YGGDRASIL IS-MOBIL fellowship from the Research Council of Norway (RCN, to GC), a FRIBIOMED grant by RCN (NFR 204822/F20 to SR) and a Leonardo Da Vinci fellowship for technicians (to MD). We also thank Dr Thomas Lingner for critical reading of the manuscript.

References

1. Hu, J., Baker, A., Bartel, B., Linka, N., Mullen, R., Reumann, S. & Zolman, B. (2012). Plant Peroxisomes: Biogenesis and Function. *Plant Cell* **24**, 2279-2303.
2. Eubel, H., Meyer, E., Taylor, N., Bussell, J., O'Toole, N., Heazlewood, J., Castleden, I., Small, I., Smith, S. & Millar, A. (2008). Novel Proteins, Putative Membrane Transporters, and an Integrated Metabolic Network Are Revealed by Quantitative Proteomic Analysis of Arabidopsis Cell Culture Peroxisomes. *Plant Physiology* **148**, 1809-1829.
3. Quan, S., Yang, P., Cassin-Ross, G., Kaur, N., Switzenberg, R., Aung, K., Li, J. & Hu, J. (2013). Proteome Analysis of Peroxisomes from Etiolated Arabidopsis Seedlings Identifies a Peroxisomal Protease Involved in beta-Oxidation and Development. *Plant Physiology* **163**, 1518-1538.
4. Reumann, S., Babujee, L., Ma, C., Wienkoop, S., Siemsen, T., Antonicelli, G., Rasche, N., Luder, F., Weckwerth, W. & Jahn, O. (2007). Proteome analysis of Arabidopsis leaf peroxisomes reveals novel targeting peptides, metabolic pathways, and defense mechanisms. *Plant Cell* **19**, 3170-3193.
5. Reumann, S., Quan, S., Aung, K., Yang, P., Manandhar-Shrestha, K., Holbrook, D., Linka, N., Switzenberg, R., Wilkerson, C., Weber, A., Olsen, L. & Hu, J. (2009). In-Depth Proteome Analysis of Arabidopsis Leaf Peroxisomes Combined with in Vivo Subcellular Targeting Verification Indicates Novel Metabolic and Regulatory Functions of Peroxisomes. *Plant Physiology* **150**, 125-143.
6. Fan, J., Quan, S., Orth, T., Awai, C., Chory, J. & Hu, J. (2005). The Arabidopsis PEX12 gene is required for peroxisome biogenesis and is essential for development. *Plant Physiology* **139**, 231-239.
7. Sparkes, I., Brandizzi, F., Slocombe, S., El-Shami, M., Hawes, C. & Baker, A. (2003). An arabidopsis *pex10* null mutant is embryo lethal, implicating peroxisomes in an essential role during plant embryogenesis. *Plant Physiology* **133**, 1809-1819.
8. Gould, S., Keller, G. & Subramani, S. (1987). Identification of a Peroxisomal Targeting Signal at the Carboxy Terminus of Firefly Luciferase. *Journal of Cell Biology* **105**, 2923-2931.
9. Gould, S., Keller, G., Hosken, N., Wilkinson, J. & Subramani, S. (1989). A Conserved Tripeptide Sorts Proteins to Peroxisomes. *Journal of Cell Biology* **108**, 1657-1664.
10. Lametschwandtner, G., Brocard, C., Fransen, M., Van Veldhoven, P., Berger, J. & Hartig, A. (1998). The difference in recognition of terminal tripeptides as peroxisomal targeting signal 1 between yeast and human is due to different affinities of their receptor Pex5p to the cognate signal and to residues adjacent to it. *Journal of Biological Chemistry* **273**, 33635-33643.
11. Lee, M., Mullen, R., Flynn, C. & Trelease, R. (1997). Characterization of the type 1 peroxisomal targeting signal (PTS1). *Plant Physiology* **114**, 1195-1195.
12. Lingner, T., Kataya, A., Antonicelli, G., Benichou, A., Nilssen, K., Chen, X., Siemsen, T., Morgenstern, B., Meinicke, P. & 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**, 1556-1572.
13. Neuberger, G., Maurer-Stroh, S., Eisenhaber, B., Hartig, A. & Eisenhaber, F. (2003). Motif Refinement of the Peroxisomal Targeting Signal 1 and Evaluation of Taxon-specific Differences. *Journal of Molecular Biology* **328**, 567-579.
14. Brocard, C. & Hartig, A. (2006). Peroxisome targeting signal 1: Is it really a simple tripeptide? *Biochimica Et Biophysica Acta-Molecular Cell Research* **1763**, 1565-1573.
15. Mullen, R., Lee, M., Flynn, C. & Trelease, R. (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 Physiology* **115**, 881-889.
16. Platta, H., Hagen, S., Reidick, C. & Erdmann, R. (2014). The peroxisomal receptor dislocation pathway: To the exportomer and beyond. *Biochimie* **98C**, 16-28.

17. Lanyon-Hogg, T., Hooper, J., Gunn, S., Warriner, S. L. & Baker, A. (2014). PEX14 binding to Arabidopsis PEX5 has differential effects on PTS1 and PTS2 cargo occupancy of the receptor. *FEBS letters* **588**, 2223-2229.
18. Grou, C., Carvalho, A., Pinto, M., Alencastre, I., Rodrigues, T., Freitas, M., Francisco, T., Sa-Miranda, C. & Azevedo, J. (2009). The peroxisomal protein import machinery - a case report of transient ubiquitination with a new flavor. *Cellular and Molecular Life Sciences* **66**, 254-262.
19. Francisco, T., Rodrigues, T., Pinto, M., Carvalho, A., Azevedo, J. & Grou, C. (2014). Ubiquitin in the peroxisomal protein import pathway. *Biochimie* **98C**, 29-35.
20. Williams, C., van den Berg, M., Sprenger, R. R. & Distel, B. (2007). A conserved cysteine is essential for Pex4p-dependent ubiquitination of the peroxisomal import receptor Pex5p. *J Biol Chem* **282**, 22534-43.
21. Gatto, G., Geisbrecht, B., Gould, S. & Berg, J. (2000). Peroxisomal targeting signal-1 recognition by the TPR domains of human PEX5. *Nature Structural Biology* **7**, 1091-1095.
22. Sampathkumar, P., Roach, C., Michels, P. & Hol, W. (2008). Structural insights into the recognition of peroxisomal targeting signal 1 by Trypanosoma brucei peroxin 5. *Journal of Molecular Biology* **381**, 867-880.
23. Stanley, W., Filipp, F., Kursula, P., Schuller, N., Erdmann, R., Schliebs, W., Sattler, M. & Wilmanns, M. (2006). Recognition of a functional peroxisome type 1 target by the dynamic import receptor Pex5p. *Molecular Cell* **24**, 653-663.
24. Fodor, K., Wolf, J., Erdmann, R., Schliebs, W. & Wilmanns, M. (2012). Molecular Requirements for Peroxisomal Targeting of Alanine-Glyoxylate Aminotransferase as an Essential Determinant in Primary Hyperoxaluria Type 1. *Plos Biology* **10**.
25. Reumann, S. (2011). Toward a definition of the complete proteome of plant peroxisomes: Where experimental proteomics must be complemented by bioinformatics. *Proteomics* **11**, 1764-1779.
26. Reumann, S., Buchwald, D. & Lingner, T. (2012). PredPlantPTS1: a web server for the prediction of plant peroxisomal proteins. *Frontiers in Plant Science* **3**, 194.
27. Gatto, G., Maynard, E., Guerrerio, A., Geisbrecht, B., Gould, S. & Berg, J. (2003). Correlating structure and affinity for PEX5 : PTS1 complexes. *Biochemistry* **42**, 1660-1666.
28. Ghosh, D. & Berg, J. (2010). A Proteome-Wide Perspective on Peroxisome Targeting Signal 1(PTS1)-Pex5p Affinities. *Journal of the American Chemical Society* **132**, 3973-3979.
29. Maynard, E. & Berg, J. (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. *Journal of Molecular Biology* **368**, 1259-1266.
30. Chowdhary, G., Kataya, A., Lingner, T. & Reumann, S. (2012). Non-canonical peroxisome targeting signals: identification of novel PTS1 tripeptides and characterization of enhancer elements by computational permutation analysis. *Bmc Plant Biology* **12**, 142.
31. Reumann, S., Ma, C., Lemke, S. & Babujee, L. (2004). AraPerox. A database of putative Arabidopsis proteins from plant peroxisomes. *Plant Physiology* **136**, 2587-2608.
32. Ma, C. & Reumann, S. (2008). Improved prediction of peroxisomal PTS1 proteins from genome sequences based on experimental subcellular targeting analyses as exemplified for protein kinases from Arabidopsis. *Journal of Experimental Botany* **59**, 3767-3779.
33. Maynard, E., Gatto, G. & Berg, J. (2004). Pex5p binding affinities for canonical and noncanonical PTS1 peptides. *Proteins-Structure Function and Bioinformatics* **55**, 856-861.
34. Arai, Y., Hayashi, M. & Nishimura, M. (2008). Proteomic analysis of highly purified peroxisomes from etiolated soybean cotyledons. *Plant and Cell Physiology* **49**, 526-539.
35. Waller, J., Dhanoa, P., Schumann, U., Mullen, R. & Snedden, W. (2010). Subcellular and tissue localization of NAD kinases from Arabidopsis: compartmentalization of *de novo* NADP biosynthesis. *Planta* **231**, 305-317.

36. Ma, C., Hagstrom, D., Polley, S. & Subramani, S. (2013). Redox-regulated Cargo Binding and Release by the Peroxisomal Targeting Signal Receptor, Pex5. *Journal of Biological Chemistry* **288**, 27220-27231.
37. Freitas, M., Francisco, T., Rodrigues, T., Alencastre, I., Pinto, M., Grou, C., Carvalho, A., Fransen, M., Sa-Miranda, C. & Azevedo, J. (2011). PEX5 Protein Binds Monomeric Catalase Blocking Its Tetrarnerization and Releases It upon Binding the N-terminal Domain of PEX14. *Journal of Biological Chemistry* **286**, 40509-40519.
38. Williams, C., Bener Aksam, E., Gunkel, K., Veenhuis, M. & van der Klei, I. J. (2012). The relevance of the non-canonical PTS1 of peroxisomal catalase. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1823**, 1133-1141.
39. Freitag, J., Ast, J. & Bolker, M. (2012). Cryptic peroxisomal targeting via alternative splicing and stop codon read-through in fungi. *Nature* **485**, 522-U135.
40. Sorhagen, K., Laxa, M., Peterhansel, C. & Reumann, S. (2013). The emerging role of photorespiration and non-photorespiratory peroxisomal metabolism in pathogen defence. *Plant Biology* **15**, 723-736.
41. Ast, J., Stiebler, A. C., Freitag, J. & Boelker, M. (2013). Dual targeting of peroxisomal proteins. *Front Physiol.* **4**, 297.
42. Xu, L., Law, S., Murcha, M., Whelan, J. & Carrie, C. (2013). The dual targeting ability of type II NAD(P)H dehydrogenases arose early in land plant evolution. *Bmc Plant Biology* **13**.
43. Chan, W. C. & White, P. D. (1999). *Fmoc solid phase peptide synthesis : a practical approach*, OUP, Oxford.
44. Fulda, M., Shockey, J., Werber, M., Wolter, F. & Heinz, E. (2002). Two long-chain acyl-CoA synthetases from Arabidopsis thaliana involved in peroxisomal fatty acid beta-oxidation. *Plant Journal* **32**, 93-103.
45. Matre, P., Meyer, C. & Lillo, C. (2009). Diversity in subcellular targeting of the PP2A B'eta subfamily members. *Planta* **230**, 935-945.
46. Ma, C., Haslbeck, M., Babujee, L., Jahn, O. & Reumann, S. (2006). Identification and characterization of a stress-inducible and a constitutive small heat-shock protein targeted to the matrix of plant peroxisomes. *Plant Physiology* **141**, 47-60.

Figure Legends

Figure 1: Analysis of the effect of point mutations in the PTS1 domain of a model sequence on semi-quantitative *in vivo* peroxisome targeting of the EYFP reporter fusion. Onion epidermal cell were biolistically transformed with EYFP fusion constructs that were C-terminally extended by the decapeptide PTS1 domain of a model sequence, acyl-CoA oxidase isoform 4 from *Zinnia elegans* (ZeACX4, VAKTTRP-SRV>) or various mutant versions. The PTS1 tripeptide alterations included four point mutations at position -1, namely V-to-I (e1-e3), V-to-M (f1-f3), V-to-Y (g1-g3) and V-to-K (h1-h3), two mutations at pos. -2, namely R-to-N (i1-i3) and R-to-T (j1-j3), and one mutation at pos. -3 (S-to-P, k1-k3). The multiple tripeptide mutations included SRV-to-SNM (l1-l3) and SRV-to-SNY (m1-m3). Furthermore, three mutations of upstream residues were investigated (VAKTTRP(SRV>)-to-VAGTTGG, n1-n3; VAKTTRP-to-VAETTDP, o1-o3, and VAKTTRP-to-VAKTTRD, p1-p3). The mutated aa are underlined. If not otherwise defined, "7aa" represents the seven upstream residues of ZeACX4 (VAKTTRP). Subcellular targeting was analyzed by fluorescence microscopy 18-24 h (a1-p1 except for b1), 48 h (a2-p2 except for b2) and 7 d (a3-p3) post transformation (p.t.). To document the efficiency of peroxisome targeting, EYFP images were not modified for brightness or contrast in single transformants (a, c-p). For each experiment at least 10-15 fluorescent cells were analysed and the results reproduced in at least 3 independent experiments. All transformed cells showed similar stages of targeting efficiency. As positive and negative controls, EYFP extended C-terminally by a PTS1 decapeptide terminating with CKI> (c) and EYFP without any extensions (d), respectively, were used. Peroxisome targeting was verified by colocalization of EYFP fluorescence with the peroxisomal marker, DsRed-SKL, for the original model sequence (b) and the PTS1 tripeptide point mutations from SRV> to SRI>, SRM> and SRY> (Suppl. Fig. 1a-c).

Figure 2: Comparison of PTS1 protein prediction scores and semi-quantitative *in vivo* peroxisome targeting of various EYFP reporter fusions with point mutations in the PTS1 domain of the model sequence. (a) PWM prediction score, (b) Standard posterior probability and (c) Balanced posterior probability. Dash lines indicate the PWM prediction thresholds of 0.412 and 0.189 for the standard and balanced posterior probabilities, respectively, (a) and of 50% (b and c). The peroxisome targeting efficiency was determined

semiquantitatively (see Fig. 1 and Table 1) and categorized as strong (peroxisomal after 18-24 h post transformation), moderate (peroxisomal only after 48 h) and weak (peroxisomal only after 7 days, i.e. 1 d RT and 6 d approx. 10°C).

Figure 3: Purification and binding activity of full length and truncated Arabidopsis PEX5. His_6-tagged versions of (a) N-terminally truncated AtPEX5C (containing the entire PTS1 binding TPR domain) and (b) full-length AtPEX5 were expressed in *E. coli* and purified by immobilised metal affinity chromatography Proteins from the indicated fractions were separated by SDS PAGE and stained with Coomassie Blue. Concentrations in the labels above the lanes refer to the concentration of imidazole used in the wash buffer and E1,E2 etc refer to elution fractions. The arrow indicates AtPEX5 and AtPEX5C**(c)** Determination of the binding constant for binding of lissamine rhodamine-YQSKL for N-terminally truncated *At*PEX5C (~100 nM). The amount of fluorescent tracer bound was calculated from the anisotropy measurements (see Supplementary Information). The curve represents a non-linear least squares fit to a single site binding equation and was generated in OrginPro.

Figure 4: Exemplar competitive binding assays between PTS1 peptides with mutations in position -1 and the tracer peptide, lissamine rhodamine-YQSKL, for binding to PEX5C. The total protein concentration was 200 nM and the tracer concentration was 30 nM. The amount of fluorescent tracer bound was calculated from the anisotropy measurements (see Supplementary Information). The curves are generated by non-linear least squares fitting to a single site competition model with the lower asymptote fixed at 0 in OrginPro. The midpoint of the transition is the IC_{50} value. The lack of inflection in the data for VAKTTRPSRV shows that the binding affinity is below the detection limit for this method.

Figure 5: Comparison of *in vitro* binding affinity of PTS1 peptides to PEX5 and PEX5C with bioinformatic predictions of targeting efficiency.

PWM score is plotted against measured K_i of the indicated peptide for competitive binding of lissamine labelled YQSKL to PEX5C. *In vivo* determined targeting strength is indicated in parenthesis where tested; w= weak, m= moderate, s=strong. Error bars represent the estimated error in K_i.

Table 1: Analysis of the effect of PTS1 domain mutations on the efficiency of *in vivo* peroxisome targeting of reporter protein fusions.

As model PTS1 domain, the C-terminal decapeptide of acyl-CoA oxidase isoform 4 (ACX4) homolog from *Zinnia elegans* (ZeACX4) was chosen. Single and multiple aa residue mutations were introduced into the model sequence and the decapeptides were attached to the C-terminal end of EYFP. The effect on the efficiency of peroxisome targeting was analysed by fluorescence microscopy. The PWM prediction scores, which are based on the C-terminal 14 aa of proteins of interest, were determined for various mutagenized decapeptides fused to EYFP by extending them N-terminally by the four C-terminal aa residues of EYFP (ELYK). The PWM score of the original 14 C-terminal aa residues of ZeACX4 (SFQL-VAKTTRP-SRV>) and those of the EYFP fusion (ELYK-VAKTTRP-SRV>) were 0.129 and 0.216, respectively. PWM prediction scores for the EYFP fusions (with a threshold of 0.412) and the standard and balanced posterior probabilities for peroxisome targeting were determined as described previously[12; 26]. The peroxisome targeting efficiency was determined semi-quantitatively and categorized as strong (peroxisomal after 18-24 h post transformation), moderate (peroxisomal only after 48 h) and weak (peroxisomal only after 7 days, i.e. 1 d RT and 6 d approx. 10^0C). For each experiment at least 10-15 fluorescent cells were analysed and the results reproduced in at least 3 independent experiments. The subcellular targeting prediction is provided according to the posterior probability and the balanced posterior probability (in parenthesis). [1] The SRM> construct could be detected in peroxisomes at very early time-points (i.e. 12 h post transformation, see Suppl. Fig. 1) and, hence, was referred to as conferring very strong peroxisome targeting to the reporter protein. C, cytosol; p, peroxisome; n.d., not determined.

Table 2: Analysis of the effect of PTS1 domain mutations on the binding affinity to AtPEX5 and *in vivo* peroxisome targeting.

The K_i for inhibition of fluorescently labelled YQSKL binding to recombinant full length Arabidopsis PEX5 or the N terminally truncated construct PEX5C was measured by fluoresence anisotropy as described in supplementary information. The K_d of the tracer peptide YQSKL was determined as 4.0 ± 0.5 nM for his_6*At*PEX5C and 4.5 ± 1.2 nM for his_6*At*PEX5 (see Suppl. Material Section 3.4). The PWM prediction score with a threshold of 0.412 and the standard and balanced posterior probabilities for peroxisome targeting were

determined as described previously[12; 26]. The peroxisome targeting efficiency was determined semi quantitatively as described for reporter fusions of the given peptides with EYFP (Table 1). The C-terminal tripeptides are printed bold. [1] The subcellular targeting prediction (column 6) is provided according to the posterior probability and the balanced posterior probability (in parenthesis). [2] For calculation of the PWM score and probabilities for peptides shorter than 14 aa residues, which is the peptide length that is considered for PWM score calculation, peptides were extended N-terminally by glycine residues. C, cytosol; p, peroxisome; n.d, not determined

Figure 1

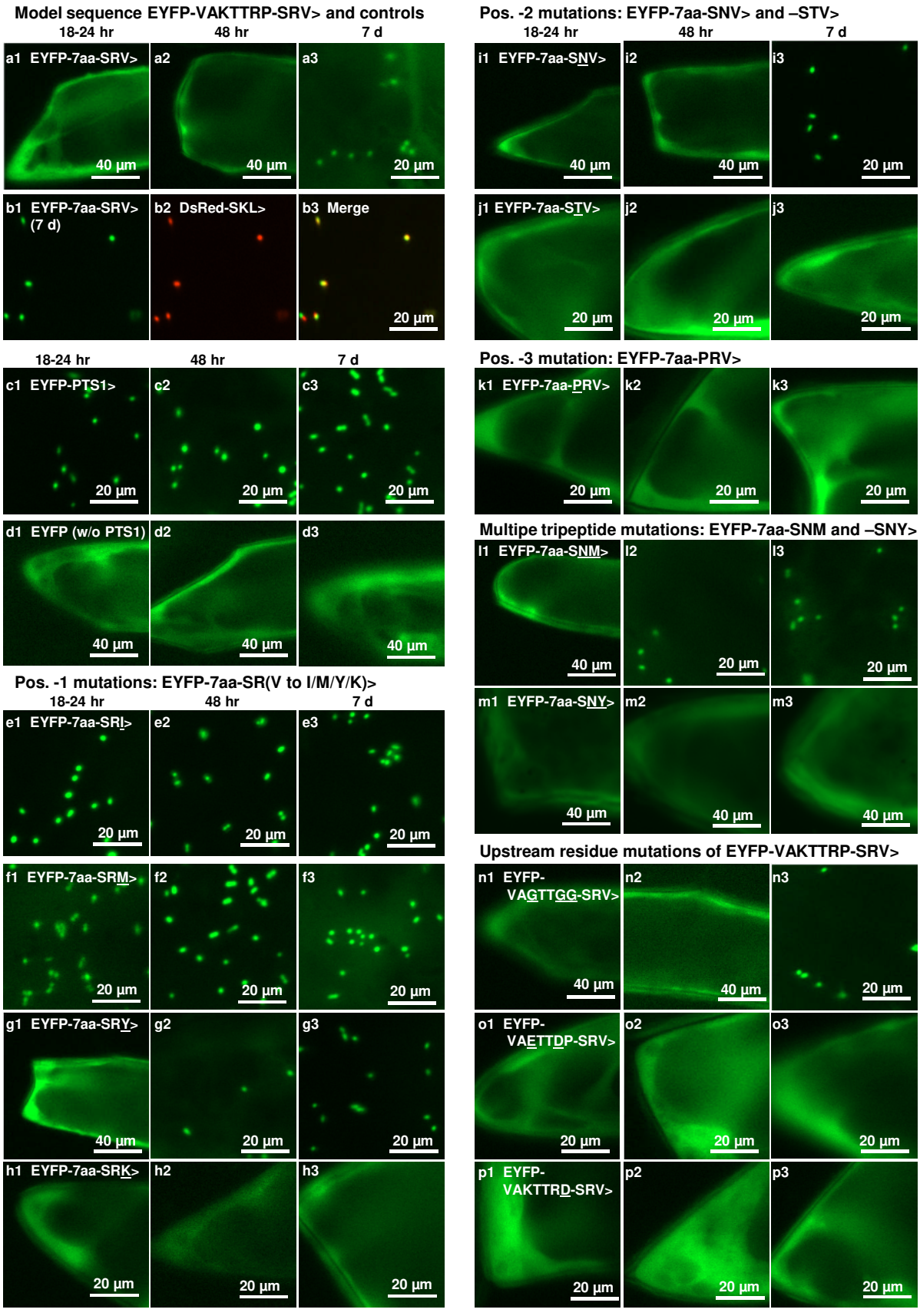
Model sequence EYFP-VAKTTRP-SRV> and controls
18-24 hr
48 hr
7 d
a1 EYFP-7aa-SRV>
a2
a3
40 µm
40 µm
20 µm
b1 EYFP-7aa-SRV> (7 d)
b2 DsRed-SKL>
b3 Merge
20 µm
18-24 hr
48 hr
7 d
c1 EYFP-PTS1>
c2
c3
20 µm
20 µm
20 µm
d1 EYFP (w/o PTS1)
d2
d3
40 µm
40 µm
40 µm
Pos. -1 mutations: EYFP-7aa-SR(V to I/M/Y/K)>
18-24 hr
48 hr
7 d
e1 EYFP-7aa-SRI>
e2
e3
20 µm
20 µm
20 µm
f1 EYFP-7aa-SRM>
f2
f3
20 µm
20 µm
20 µm
g1 EYFP-7aa-SRY>
g2
g3
40 µm
20 µm
20 µm
h1 EYFP-7aa-SRK>
h2
h3
20 µm
20 µm
20 µm
Pos. -2 mutations: EYFP-7aa-SNV> and –STV>
18-24 hr
48 hr
7 d
i1 EYFP-7aa-SNV>
i2
i3
40 µm
40 µm
20 µm
j1 EYFP-7aa-STV>
j2
j3
20 µm
20 µm
20 µm
Pos. -3 mutation: EYFP-7aa-PRV>
k1 EYFP-7aa-PRV>
k2
k3
20 µm
20 µm
20 µm
Multipe tripeptide mutations: EYFP-7aa-SNM and –SNY>
l1 EYFP-7aa-SNM>
l2
l3
40 µm
20 µm
20 µm
m1 EYFP-7aa-SNY>
m2
m3
40 µm
40 µm
40 µm
Upstream residue mutations of EYFP-VAKTTRP-SRV>
n1 EYFP-VAGTTGG-SRV>
n2
n3
40 µm
40 µm
20 µm
o1 EYFP-VAETTDP-SRV>
o2
o3
20 µm
20 µm
20 µm
p1 EYFP-VAKTTRD-SRV>
p2
p3
20 µm
20 µm
20 µm

Figure 2

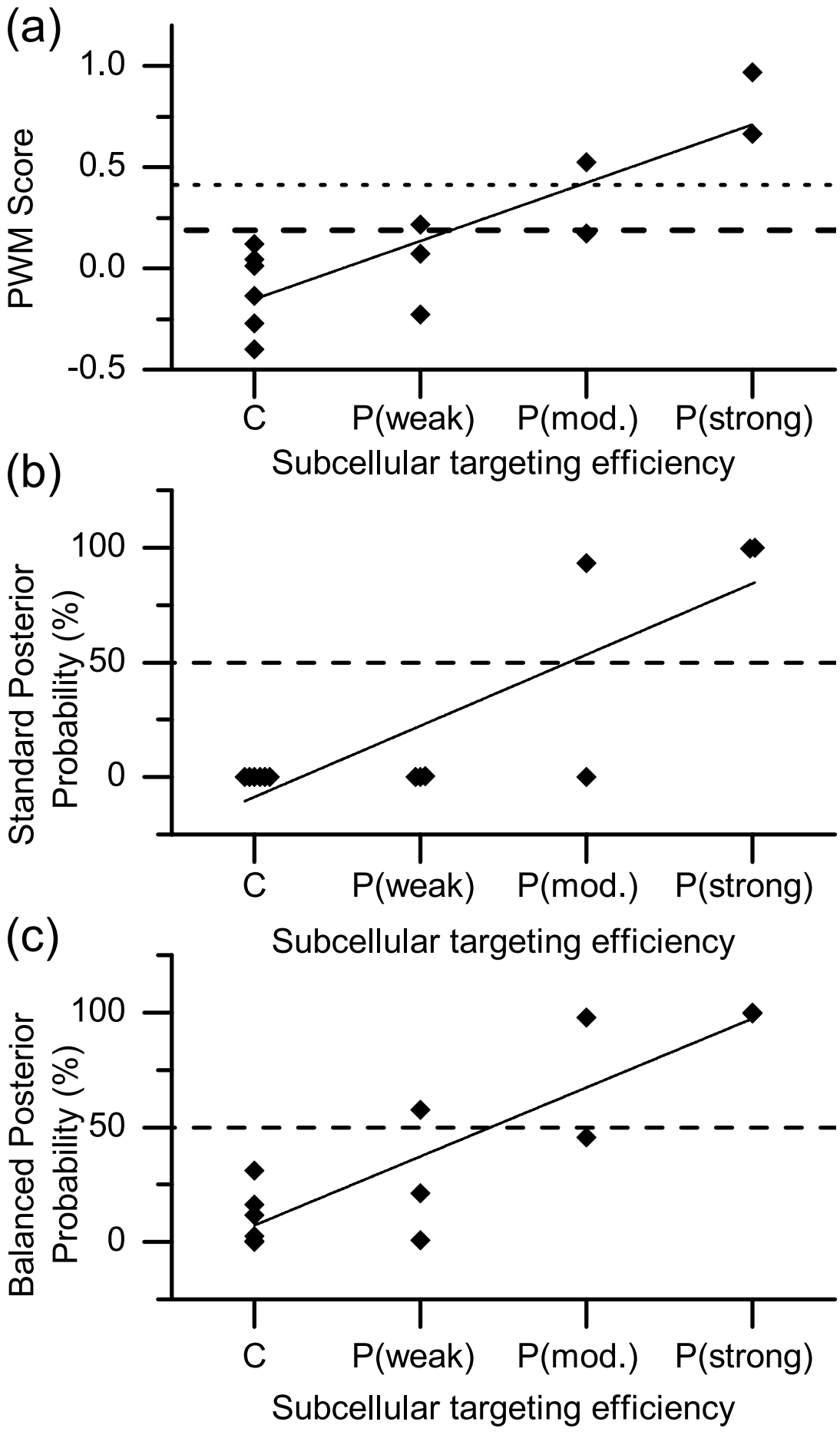
(a)
PWM Score
1.0
0.5
0.0
-0.5
C
P(weak)
P(mod.)
P(strong)
Subcellular targeting efficiency
(b)
Standard Posterior Probability (%)
100
50
0
C
P(weak)
P(mod.)
P(strong)
Subcellular targeting efficiency
(c)
Balanced Posterior Probability (%)
100
50
0
C
P(weak)
P(mod.)
P(strong)
Subcellular targeting efficiency

Figure 3

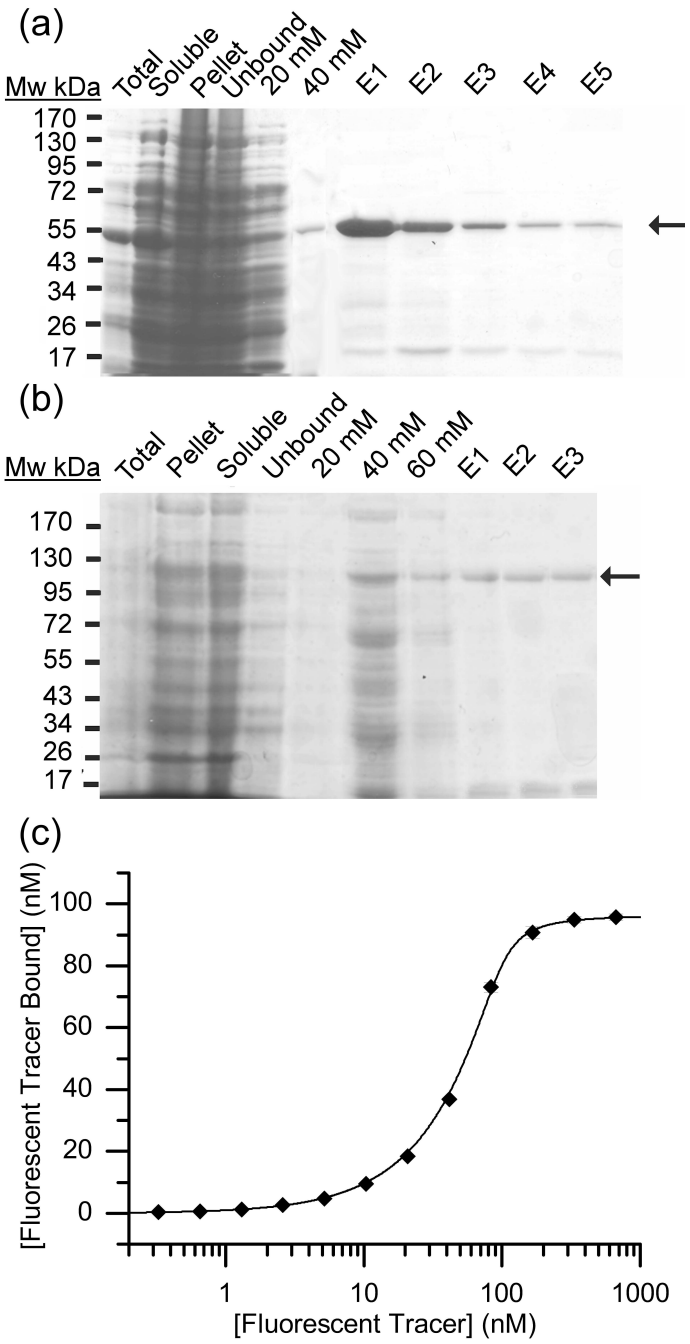
(a)
Mw kDa
Total
Soluble
Pellet
Unbound
20 mM
40 mM
E1
E2
E3
E4
E5
170
130
95
72
55
43
34
26
17
(b)
Mw kDa
Total
Pellet
Soluble
Unbound
20 mM
40 mM
60 mM
E1
E2
E3
170
130
95
72
55
43
34
26
17
(c)
[Fluorescent Tracer Bound] (nM)
100
80
60
40
20
0
1
10
100
1000
[Fluorescent Tracer] (nM)

Figure 4

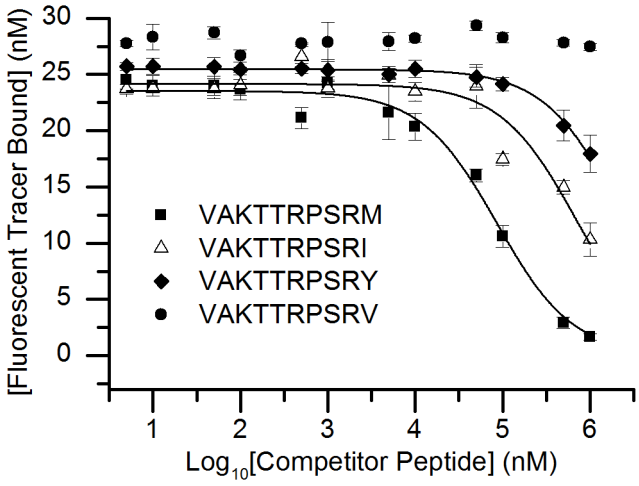
[Fluorescent Tracer Bound] (nM)
30
25
20
15
10
5
0
1
2
3
4
5
6
Log10[Competitor Peptide] (nM)
VAKTTRPSRM
VAKTTRPSRI
VAKTTRPSRY
VAKTTRPSRV

Figure 5

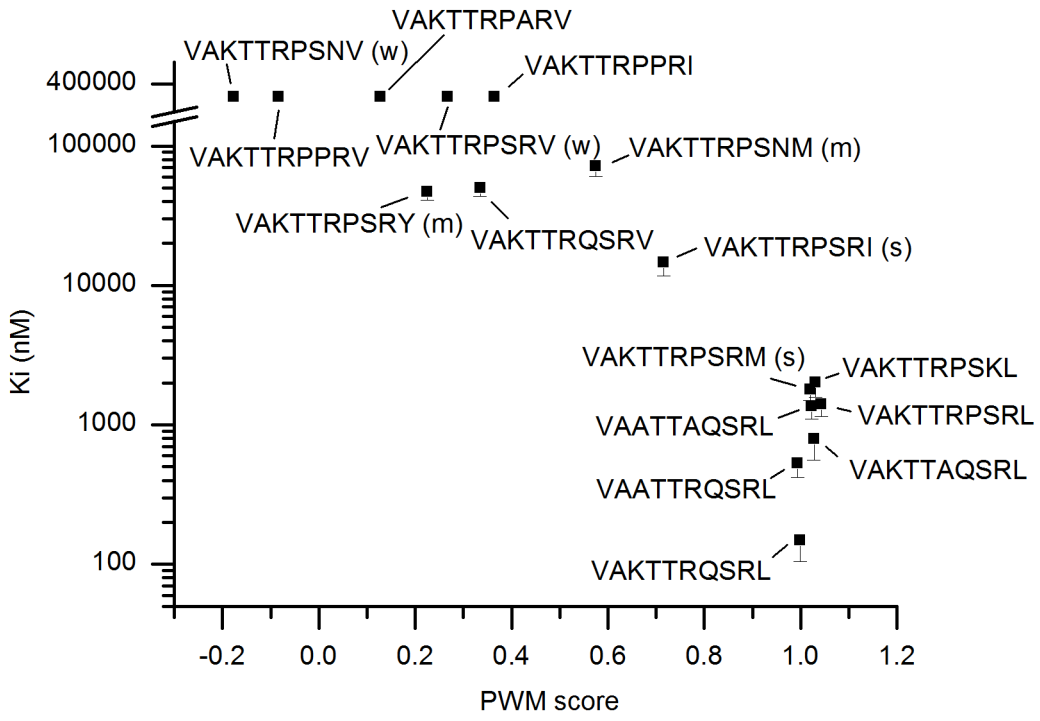
VAKTTRPARV
VAKTTRPSNV (w)
VAKTTRPPRI
VAKTTRPPRV
VAKTTRPSRV (w)
VAKTTRPSNM (m)
VAKTTRPSRY (m)
VAKTTRQSRV
VAKTTRPSRI (s)
VAKTTRPSRM (s)
VAKTTRPSKL
VAATTAQSRL
VAKTTRPSRL
VAKTTAQSRL
VAATTRQSRL
VAKTTRQSRL
400000
100000
10000
1000
100
Ki (nM)
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
PWM score

Table 1: Analysis of the effect of PTS1 domain mutations on the efficiency of *in vivo* peroxisome targeting of reporter protein fusions.

Muta-tion	**Reporter protein fusion**	**PWM pred. score**	**Post. prob. (%)**	**Bal. post. prob. (%)**	**Predict.[1]**	***In vivo* subcellular targeting**			**Perox. targ. efficiency**
						24 h	**48 h**	**7 d**	
Model protein	ZeACX4-VAKTTRP**SRV**	0.129	0.0	33.5	C (C)	n.d.	n.d.	n.d.	**n.d.**
Model EYFP constr.	EYFP-VAKTTRP**SRV**	0.216	0.4	57.6	C (P)	C	C	P (weak)	**P (weak)**
Pos. -1	EYFP-VAKTTRP**SRM**	0.969	100	100	P (P)	P (strong)	P (strong)	P (strong)	**P (very strong)[1]**
Pos. -1	EYFP-VAKTTRP**SRI**	0.664	99.6	99.5	P (P)	P (strong)	P (strong)	P (strong)	**P (strong)**
Pos. -1	EYFP-VAKTTRP**SRY**	0.173	0.1	45.5	C (C)	C	P (weak)	P (strong)	**P (mod.)**
Pos. -1	EYFP-VAKTTRP**SRK**	0.119	0.0	31.1	C (C)	C	C	C	**C**
Pos. -2	EYFP-VAKTTRP**SNV**	-0.229	0.0	0.9	C (C)	C	C	P (weak)	**P (weak)**
Pos. -2	EYFP-VAKTTRP**STV**	-0.401	0.0	0.1	C (C)	C	C	C	**C**
Pos. -3	EYFP-VAKTTRP**PRV**	-0.135	0.0	2.5	C (C)	C	C	C	**C**
Multiple (PTS1 trip.)	EYFP-VAKTTRP**SNM**	0.523	93.3	97.8	P (P)	C	P (weak)	P (strong)	**P (mod.)**
Multiple (PTS1 trip.)	EYFP-VAKTTRP**SNY**	-0.272	0.0	0.50	C (C)	C	C	C	**C**
Multiple (upstream)	EYFP-VA**G**TT**GG**SRV	0.073	0.0	21.2	C (C)	C	C	P (weak)	**P (weak)**
Multiple (upstream)	EYFP-VA**E**TT**D**PSRV	0.011	0.0	11.7	C (C)	C	C	C	**C**
Multiple (upstream)	EYFP-VAKTTR**D**SRV	0.045	0.0	16.4	C (C)	C	C	C	**C**

Table 2: Analysis of the effect of PTS1 domain mutations on the binding affinity to AtPEX5 and peroxisome targeting efficiency of reporter fusions.

	Muta-tion	**Peptide**	**PWM pred. score** [2]	**Post. prob. (%)**[2]	**Bal. post. prob. (%)**[2]	**Predict. (C/P)** [1]	**PEX5C** K_i **(nM)**	**PEX5** K_i **(nM)**	**Perox. targeting efficiency of EYFP fusions**
1	Original.	VAKTTRP**SRV**	0.268	1.8	71	C (P)	>100000	>100000	P (weak)
2	Pos. -1	VAKTTRP**SRL**	1.043	100	100	P (P)	1400 ± 250	2220 ± 630	n.d.
3	Pos. -1	VAKTTRP**SRM**	1.020	100	100	P (P)	1800 ± 310	3100 ± 980	P (very strong)
4	Pos. -1	VAKTTRP**SRI**	0.716	99.8	99.7	P (P)	14500 ± 2800	21500 ± 5900	P (strong)
5	Pos. -1	VAKTTRP**SRY**	0.225	0.5	60.0	C (P)	47100 ± 6300	25400 ± 5800	P (mod.)
6	Pos. -2	VAKTTRP**SNV**	-0.177	0.0	1.5	C (C)	>100000	n.d.	P (weak)
7	Pos. -3	VAKTTRP**ARV**	0.128	0.0	33.2	C (C)	>100000	n.d.	n.d.
8	Pos. -3	VAKTTRP**PRV**	-0.083	0.0	4.3	C (C)	>100000	n.d.	C
9	Upstream residue	VAKTTRR**QSRV**	0.335	11.3	84.0	C (P)	50200 ± 6800	53400 ± 14000	n.d.
10	Multiple	VAKTTRP**SKL**	1.031	100	100	P (P)	2000 ± 420	1300 ± 340	n.d.
11	Multiple	VAKTTRP**SNM**	0.575	97.7	98.8	P (P)	72000 ± 11500	48000 ± 13000	P (mod.)
12	Multiple	VAKTTR**QSRL**	0.999	100	100	P (P)	148 ± 44	120 ± 40	n.d.
13	Multiple	VA**A**TTR**QSRL**	0.994	100	100	P (P)	531 ± 110	375 ± 130	n.d.

14	Multiple	VAKTTAQ**SRL**	1.028	100	100	P (P)	796 ± 240	824 ± 275	n.d.
15	Multiple	VAATTAQ**SRL**	1.023	100	100	P (P)	1360 ± 260	1560 ± 450	n.d.
16	Multiple	VAKTTRP**PRI**	0.365	22.3	88	C (P)	>100000	>100000	n.d.
17	Orig. pentapeptide	YQ**SKL**	0.818	100	99.9	P (P)	166 ± 23	189 ± 52	n.d.
18	Pos. -1	YQ**SKV**	0.043	0.0	16.0	C (P)	32400 ± 4800	24100 ± 6900	n.d.
19	Upstream residue	Y**P**SKL	0.862	100	100	P (P)	515 ± 130	287 ± 84	n.d.