hfAIM: A RELIABLE BIOINFORMATICS APPROACH FOR *IN SILICO* GENOME-WIDE IDENTIFICATION OF AUTOPHAGY-ASSOCIATED ATG8-INTERACTING MOTIFS IN VARIOUS ORGANISMS

Qingjun Xie[a,1], Oren Tzfadia[a,b,c,d,1], Matan Levy[a], Efrat Weithorn[a], Hadas Peled-Zehavi[a], Thomas Van Parys[b,c,d], Yves Van de Peer[b,c,d,e] and Gad Galili[a,*]

[a]Department of Plant and Environmental Science, Weizmann Institute of Science, Rehovot 76100 Israel

[b]Department of Plant Systems Biology, VIB, 9052 Ghent, Belgium

[c]Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium

[d]Bioinformatics Institute Ghent, Ghent University, 9052 Ghent, Belgium

[e]Genomics Research Institute (GRI), University of Pretoria, 0028 Pretoria, South Africa

[1]The first two authors contributed equally to this manuscript

*Corresponding author

Keywords: AIM; Atg8; *Arabidopsis*; Autophagy; Bioinformatics; PEX; Pexophagy.

Abbreviation: ACP, AIM-containing proteins; AIM, Atg8-interacting motif; *At*, *Arabidopsis thaliana*; BiFC, Bimolecular fluorescence complementation; hfAIM, high fidelity AIM; PEX, peroxin; PSSM, position-specific scoring matrix;

ABSTRACT

Most of the proteins that are specifically turned over by selective autophagy are recognized by the presence of short Atg8 interacting motifs (AIMs) that facilitate their association with the autophagy apparatus. Such AIMs can be identified by bioinformatics methods based on their defined degenerate consensus F/W/Y-X-X-L/I/V sequences in which X represents any amino acid. Achieving reliability and/or fidelity of the prediction of such AIMs on a genome-wide scale represents a major challenge. Here, we present a bioinformatics approach, high fidelity AIM (hfAIM), which uses additional sequence requirements—the presence of acidic amino acids and the absence of positively charged amino acids in certain positions—to reliably identify AIMs in proteins. We demonstrate that the use of the hfAIM method allows for *in silico* high fidelity prediction of AIMs in AIM-containing proteins (ACPs) on a genome-wide scale in various organisms. Furthermore, by using hfAIM to identify putative AIMs in the *Arabidopsis* proteome, we illustrate a potential contribution of selective autophagy to various biological processes. More specifically, we identified 9 peroxisomal PEX proteins that contain hfAIM motifs, among which AtPEX1, AtPEX6 and AtPEX10 possess evolutionary-conserved AIMs. Bimolecular fluorescence complementation (BiFC) results verified that AtPEX6 and AtPEX10 indeed interact with Atg8 *in planta*. In addition, we show that mutations occurring within or nearby hfAIMs in PEX1, PEX6 and PEX10 caused defects in the growth and development of various organisms. Taken together, the above results suggest that the hfAIM tool can be used to effectively perform genome-wide *in silico* screens of proteins that are potentially regulated by selective autophagy. The hfAIM system is a web tool that can be accessed at link: http://bioinformatics.psb.ugent.be/hfAIM/.

INTRODUCTION

Macroautophagy (hereafter referred to as autophagy) is a highly conserved biological process in eukaryotes, which mainly functions in the degradation of macromolecules in the lytic compartment.[1-3] One of the central core proteins of the autophagy machinery is Atg8. Atg8 binding to specific target proteins is often, though not always, mediated by a conserved motif, the Atg8-interacting motif (AIM), on the target protein.[4, 5] The core AIM motif is comprised of 4 amino acids, defined as F/W/Y-X-X-L/I/V,[6] in which 'X' represents any amino acid. Notably, structural analysis suggested a striking bias towards negatively charged amino acids present within or upstream of the core AIM.[4, 7, 8] Therefore, it has been proposed that the acidic amino acid Asp (D) and Glu (E), and potentially also Ser (S) and Thr (T) that generate negative charges when phosphorylated, can improve the strength of binding of Atg8 to the AIMs.[4, 9] Furthermore, it has been hypothesized that the closer the acidic or phosphorylated amino acids are to the core AIM or their presence in the two X positions within the core AIM, the higher is the fidelity of binding of Atg8 to the AIMs.[7]

Based on the degenerate consensus sequence of AIMs, it is possible to use bioinformatics tools to look for potential Atg8-interacting proteins by searching for AIM motifs followed by verification of their binding to Atg8 by experimental methods, such as yeast two-hybrid and bimolecular fluorescence complementation (BiFC). Indeed, 2 bioinformatics tools that identify consensus AIMs in proteins were previously developed, the first being reported by our laboratory[10] and the second, iLIR, reported by Kalavari and associates.[11] Our method[10] took into account the contribution of acidic amino acids to the fidelity of binding of Atg8 to the AIM. The iLIR system defines an AIM, termed xLIR, based on a regular expression pattern that is based on the sequences of a set of verified AIMs and the 2 amino acids that precede it.[11] Additionally, iLIR scores xLIRs against a custom position-specific scoring matrix (PSSM) and identifies potentially disordered subsequences with protein interaction potential overlapping with detected xLIR-motifs (ANCHOR).[12, 13] Interestingly, the regular pattern of xLIR, that is based on experimentally determined Atg8-interacting regions, does not contain positively charged amino acids in the minus 1 and plus 1 positions upstream or downstream of the F/W/Y sequence of the AIM, strengthening the notion that the presence of positively charged amino acids in these 2 positions may hinder the binding of Atg8 to the AIM.

The peroxisome is a highly dynamic organelle involved in metabolism, development and response to stresses, and its homeostasis is regulated by selective autophagy.[14-16] Recent work indicates that peroxisomes and peroxisomal proteins accumulate in autophagy (*atg)* mutants.[17, 18] Moreover, the Atg8 protein frequently colocalizes with peroxisome aggregates, indicating that peroxisomes are selectively degraded by autophagy.[17] Therefore, it has been proposed that autophagy apparently regulates the homeostasis of peroxisomes through the degradation of certain peroxisomal proteins.[16, 19] However, it is still not clear which peroxisome proteins are selectively turned over by autophagy. Peroxin (PEX) proteins are peroxisomal proteins that serve multiple functions in the operation of this organelle.[20] Interestingly, a recent report showed that a G-to-E point mutation in the *Arabidopsis* PEX10 protein alters the shape of the peroxisome and that this mutant displays a dominant negative plant phenotype, which is highly similar to that of *AtPEX10*-knockout mutants.[21] These results suggest that the G-to-E mutation results in an irreversible degradation of AtPEX10 protein. Taking into account the key role of PEX10 in the peroxisome,[21, 22] it is reasonable to speculate that PEX10 may be one of the candidate peroxisomal proteins degraded by autophagy. However, this hypothesis, as well as the identification of other peroxisomal proteins that are regulated by autophagy, still remains to be further determined.

In the present report, we present a bioinformatics approach, termed high fidelity AIM system (hfAIM), for *in silico* genome-wide identification of AIMs. Application of the hfAIM system facilitates a rapid identification of potentially interesting proteins that are associated with autophagy, as well as studying the network regulation of autophagy. As a test case, we utilized hfAIM to identify potential AIMs in PEX proteins from multiple model organisms. Evolutionary conservation of the predicted AIMs was further used to refine the predictions. BiFC experiments were used to validate hfAIM predictions for *Arabidopsis* PEX proteins. Our results suggest that PEX6, PEX10 and likely also PEX1 contain functional AIMs and interact with Atg8, suggesting that autophagy regulates the homeostasis of peroxisomes through the degradation of specific PEX proteins. The hfAIM system is a web tool (link: http://bioinformatics.psb.ugent.be/hfAIM/), which allows users to upload FASTA files to scan for our 5-hfAIM motifs (as default) and add or remove motifs as they wish. The code is deposited in a github at https://gitlab.psb.ugent.be/thpar/hfAIM/blob/master/README.md.

RESULTS

Estimation of the contribution of acidic amino acids to functional AIMs

Atg8-interacting proteins often possess one or more functional Atg8-interacting motifs (AIMs), which are comprised of the core consensus sequence F/W/Y-X-X-L/I/V.[1, 3, 6] The presence of acidic amino acids either immediately upstream the F/W/Y sequence or at any of the 2 X positions between the F/W/Y and the L/I/V sequences appear to contribute to the fidelity of binding of Atg8 to the AIM.[7, 23] Thus, it might be useful to consider a longer 6-amino acid X_{-2}-X_{-1}-F/W/Y-X_{+1}-X_{+2}-L/I/V motif for AIMs. Based on this degenerate sequence, we have previously developed a bioinformatics tool,[10] termed "canonical AIM" (cAIM), for identifying AIMs in a group of plant exocyst subunits, whose transport to the vacuole was suggested to require the autophagy apparatus.[24] Another more recently developed tool, iLIR [10], also defines an AIM as a 6-amino acid motif, termed xLIR, based on the following degenerate amino acid sequence: [ADEFGLPRSK][DEGMSTV][WFY][DEILQTV][ADEFHIKLMPSTV][ILV]. As mentioned above, accumulating data suggested that the presence of acidic amino acids in any of the 2 "X" positions within the core AIM (namely X_{+1}, X_{+2}) or in any of the 2 "X" positions upstream to the F/W/Y sequence (namely X_{-2}, X_{-1}) may improve the binding efficiency of Atg8 to the AIM.[6, 7] While looking further into the 36 verified functional AIMs collected from the literature (see detail in **Table S1**), we found that 29 functional AIMs contain one or more acidic amino acid (**Fig. 1A, Table S1**). Of the remaining 7 AIMs, 4 AIMs contain at least one S residue, only one AIM possesses neither acidic amino acid nor S and T residues, and 2 functional AIMs are atypical AIMs (**Table S1**), which thereby were excluded from our following studies. Furthermore, the frequency of acidic amino acids at the degenerate 5 positions of the AIM motif (X_{-3}, X_{-2}, X_{-1}, X_{+1} and X_{+2}) is specifically higher than the percentage found in a random sequence of 5 amino acids (Figure 1B). These results suggest that introducing a requirement for negatively charged amino acids around the core consensus sequence of the AIM motif might improve the predictive power of bioinformatics tools.

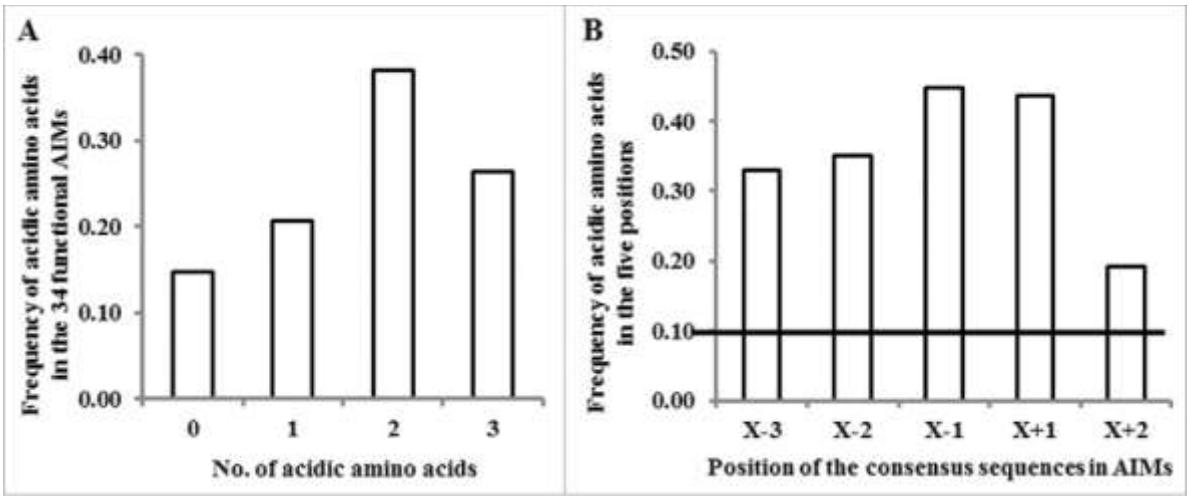

Figure 1. The distribution of acidic amino acids in known functional AIMs. (**A**) The frequency of acidic amino acids present in the 34 functional AIMs (excluding the 2 atypical AIMs). (**B**) The frequency of acidic amino acids present in each of the 5 analyzed positions in the 34 functional AIMs. The horizontal line indicates the expected frequency of acidic amino acid in a random distribution (2/20 amino acids).

Generation of the hfAIM system

Based on preference for acidic amino acids in AIM motifs we developed a bioinformatics tool, hfAIM, for the prediction of AIM motifs in proteins. An hfAIM motif was defined as a motif containing at least 2 acidic amino acids in the X_{-3}, X_{-2}, X_{-1}, X_{+1} or X_{+2} positions. Since the contribution of phosphorylation of S and T residues to AIM motif binding to Atg8 still needs to be verified experimentally, we did not introduce a bias towards these residues in our AIM prediction algorithm. This definition of the hfAIM motif resulted in ten regular patterns (**Fig. S1**). Subsequently, we examined the distribution of these ten regular patterns among the 19 functional AIMs in our data sets that contain at least 2 acidic amino acids (**Table S1**), excluding the 7 AIMs that contain only one acidic amino acid, 5 AIMs that do not contain any acidic amino acid and the 2 atypical AIMs. Interestingly, we found that 6 out of these 10 regular patterns were enough to fully cover the above 22 AIMs (A, B, C, F, H and I, see **Table S1**). In addition, only one functionally proved AIM, "PSHWPLI", out of the 34 typical verified AIMs contains a positively charged amino acid His (H) at the X_{-1} position, and none contain positively charged amino acids at the X_{+1} position (**Table S1**), supporting the notion that positively charged amino acids have a negative effect on the binding of Atg8 to AIMs. Thus, we excluded putative AIMs containing positively charged amino acids at either the X_{-1} or the X_{+1} position (see below). Finally, according to the hypothesis that the closer the

acidic amino acids are to the F/W/Y sequence of the core AIM, the higher is the fidelity of binding of Atg8 to these AIMs,[7] we excluded the "H" regular AIM pattern from the 6 regular expression patterns, resulting in only 5 regular expression patterns that meet the standard of an hfAIM motif (**Fig. 2A**).

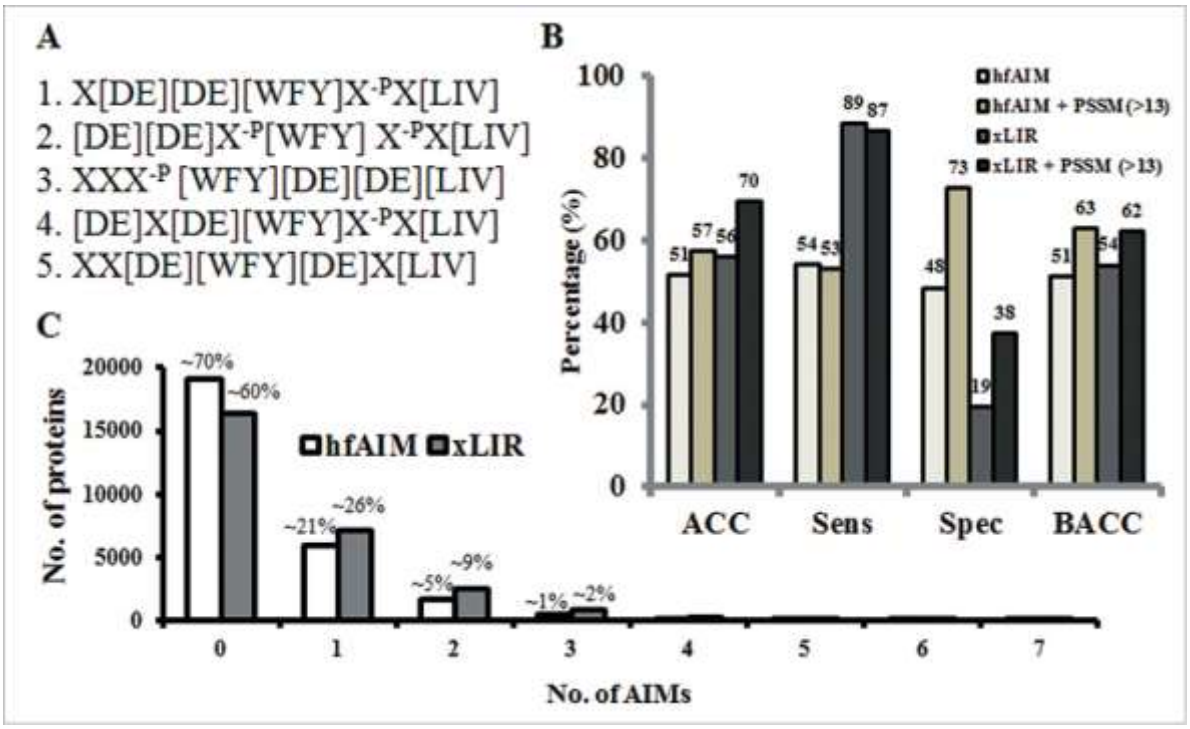

Figure 2. AIMs prediction by hfAIM. (**A**) The 5 regular patterns used for the analysis of AIMs in the hfAIM system. X, represent any amino acids. X^{-p}, represent any amino acids except for the positively charged amino acids. (**B**) Comparison of the quality of the hfAIM and iLIR systems according to the statistical validation described by Kalvari et al.[10] "ACC", accuracy of the AIM prediction; "Sens", sensitivity of the AIM prediction; "Spec", specificity of the AIM prediction; "BACC", balanced accuracy of the AIM prediction. (**C**) Distribution of the number of AIMs in *Arabidopsis* proteins as predicted by the hfAIM and iLIR systems.

To validate the quality of our prediction of potentially functional hfAIMs, we used the statistical method reported by Kalvari and associates[11] to compare our hfAIM predictions to the iLIR predictions of AIMs in the dataset of experimentally verified AIMs (**Table S1**). As shown in Figure 2B, the 2 approaches seem to give similar results in terms of both the accuracy and balanced accuracy of prediction of AIMs in this dataset that compiles mostly human proteins. Nevertheless, these 2 systems displayed different sensitivity and specificity of AIM prediction. While the hfAIM system appears more powerful in specificity, the iLIR system is better at sensitivity (**Fig. 2B**). Similarly to iLIR, the hfAIM algorithm is based on a

regular expression pattern, a sequence of symbols and characters expressing a string or pattern to be searched for within a longer piece of sequence. Though searching for a regular expression pattern is a useful method for scanning large sequence data sets for sequences of interest, it sometimes suffers from high rate of false negative sequences. Adding a stricter criterion based on position-specific scoring matrix (PSSM) can improve the specificity of AIM prediction, as it has been previously shown for the iLIR.[11] PSSM is a tool that is used to score how close any sequence is to the collected sequences used to create the scoring matrix. Based on the training sequences, a score is assigned to the presence of a residue in each position in the sequence. A higher total score represents a sequence that is closer to the training sequences relative to other sequences of similar length. Thus, a genuine AIM motif is expected to have higher PSSM scores. As the hfAIM motif is defined as a 7-amino acid long motif, while the xLIR is a 6-amino acid long motif, PSSM scores calculated for the motifs predicted by each of the approaches cannot be directly compared. Therefore, PSSM values were calculated based on a 6-amino acid motif and the Kalvari et al [11] custom PSSM for both the iLIR and hfAIM predictions of AIMs in our dataset (**Table S1**). The predictions were re-evaluated using a PSSM threshold value of 13 (i.e. only predicted sequences with PSSM value higher than 13 are taken into account[11]).

Although hfAIM still provides higher specificity relative to iLIR predictions, adding the PSSM predictor improves the specificity of both iLIR and hfAIM predictions leading to an improved balanced accuracy (**Fig. 2B**).

The dataset of verified AIMs used above (**Table S1**) contained mostly human proteins. To compare the predictive power of the iLIR and hfAIM approaches in plants, we first separately applied each of the 2 systems to identify putative AIMs in the entire *Arabidopsis* proteome (the in-house script for the hfAIM system is deposited in a github https://gitlab.psb.ugent.be/thpar/hfAIM/blob/master/README.md). Nearly 40% of the proteins of the entire *Arabidopsis* proteome contain AIMs according to the iLIR system, whereas about ~30% of the *Arabidopsis* proteins contain AIMs as defined by the hfAIM system (**Fig. 2C**). Next, we applied these 2 systems to identify AIMs in a dataset of 26 verified *Arabidopsis* Atg8-interacting proteins that were collected from the literature (see details in **Table S2**). We found differences between the predictions of the hfAIM and iLIR systems. Ten xLIR motifs derived from 9 proteins were identified by the iLIR system, while 16 hfAIM

motifs derived from 9 proteins were identified by the hfAIM system. Among these Atg8-interacting proteins, 4 proteins were recognized as containing AIMs by both systems and only 2 identical AIM motifs were recognized by both systems (**Table S2**). When a PSSM threshold value of 13 was applied to the predictions (calculated according to Kalvari et al.[11]), iLIR identified only 5 xLIR motifs in 4 proteins in the dataset of verified Atg8-interacting proteins from *Arabidopsis*, while hfAIM identified 8 AIMs in 8 of the proteins. Taken together, these results suggest that hfAIM might be somewhat better suited for the prediction of AIMs in plants.

Global analysis of the role of autophagy-associated AIMs-containing proteins in multiple biological processes in plants

To broadly investigate a potential contribution of autophagy to various biological processes in plants, we performed a gene ontology (GO) enrichment analysis for the AIMs-containing proteins (hereafter termed as ACPs) identified by hfAIM in *Arabidopsis* plants. The GO enrichment analysis was conducted on groups of proteins containing increasing numbers of AIMs (**Table S3**). Since only 4 ACPs contained 7 to 8 AIMs and > 5000 proteins contained only one AIM, these 2 groups of protein genes were discarded from the GO enrichment analysis. The GO enrichment results suggest that ACPs are involved in multiple biological processes and molecular functions (See **Table S3** for full list of GO enrichments). Notably, some GO terms associated with the ACPs were directly connected to autophagy-associated cellular catabolic process (GO:0044248), and proteolysis involved in cellular protein catabolic process (GO:0051603), suggesting that hfAIM is able to predict AIMs in proteins that are indeed likely to be involved in autophagy-related processes. The GO terms of Other ACPs were related to metabolism, like gluconeogenesis (GO:0006094) and carbohydrate biosynthetic processes (GO:0016051), which is consistent with recent reports showing the comprehensive participation of autophagy in maintaining the homeostasis of cellular metabolism[25-27]. Surprisingly, we found that > 30% of the ACPs are related to the adenyl ribonucleotide binding (GO:0032559) and nucleobase-containing compound metabolic process (GO:0006139). In addition, one of the largest groups of ACPs is involved in the regulation of transcription, implying a new, relatively poorly understood role of autophagy in transcriptional control in plants. Furthermore, ACPs were also involved in signaling transduction, stress response and

protein transport, as well as with other biological processes. Taken together, our results suggest that the hfAIM system enables an in-house genome-wide identification of ACPs, and thus facilitates the high-throughput analysis of the role of autophagy not only in plants, but also in various other organisms.

AtPEX10 interacts with Atg8 and this interaction requires a functional AIM motif predicted by hfAIM

To further verify the ability of the hfAIM system to identify functional AIMs, we used a group of peroxisome peroxin (PEX) proteins as a test case. The choice of PEX proteins was based on 4 independent reasons: (i) autophagy participates in the homeostasis of peroxisomes by a process termed pexophagy;[14-16] (ii) it has already been reported that AtPEX10-YFP fusion protein localizes to peroxisomes in tobacco leaves[28] and that a GFP-Atg8 fusion protein is a functional protein;[29] (iii) a G-to-E point mutation in the *Arabidopsis* AtPEX10 protein, which resulted in a peroxisome deficient phenotype,[21] occurs in a sequence predicted to be an AIM by the hfAIM system. The underlined G in the predicted AIM GEEYCDI sequence was mutated to E, introducing an extra acidic amino acid, that might improve the binding to Atg8 (**Fig. 3A**); and (iv) additional analysis indicated that this natural GEEYCDI AIM (amino acids 93 to 99 in AtPEX10 sequence) is evolutionary conserved (**Fig. S3**).

To look at the interaction of Atg8 with AtPEX10, we utilized the bimolecular fluorescence complementation (BiFC) assay as previously described.[30] Thus, we produced a C-terminal split YFP fusion protein with AtPEX10 (AtPEX10-YC), as well as an N-terminal split YFP fusion protein with Atg8 (YN-Atg8). Transient coexpression of AtPEX10-YC and YN-Atg8 in *N. benthamiana* leaves showed that AtPEX10 indeed interacts with Atg8 *in vivo* (**Fig. 3B**), whereas the negative controls had no signals (**Fig. S2A and S2B**). Notably, the iLIR system identified a different AIM motif in AtPEX10, "GVFLLI" (amino acids 251 to 256 in AtPEX10 sequence), with a lower PSSM score compared to the predicted hfAIM motif (**Table 1**). Therefore, we were interested in testing which of these 2 potential AIMs in AtPEX10 is needed for the interaction with Atg8. To address this issue, we eliminated the potential AIMs in AtPEX10 by substituting the Tyr residue at position 96 (Y96) and the Phe residue at position 253 (F253) to Ala (A) (**Fig. 3A**), respectively, and then generated the AtPEX10^{Y96A}-YC and AtPEX10^{F253A}-YC fusion proteins respectively. Using the BiFC assay, we transiently

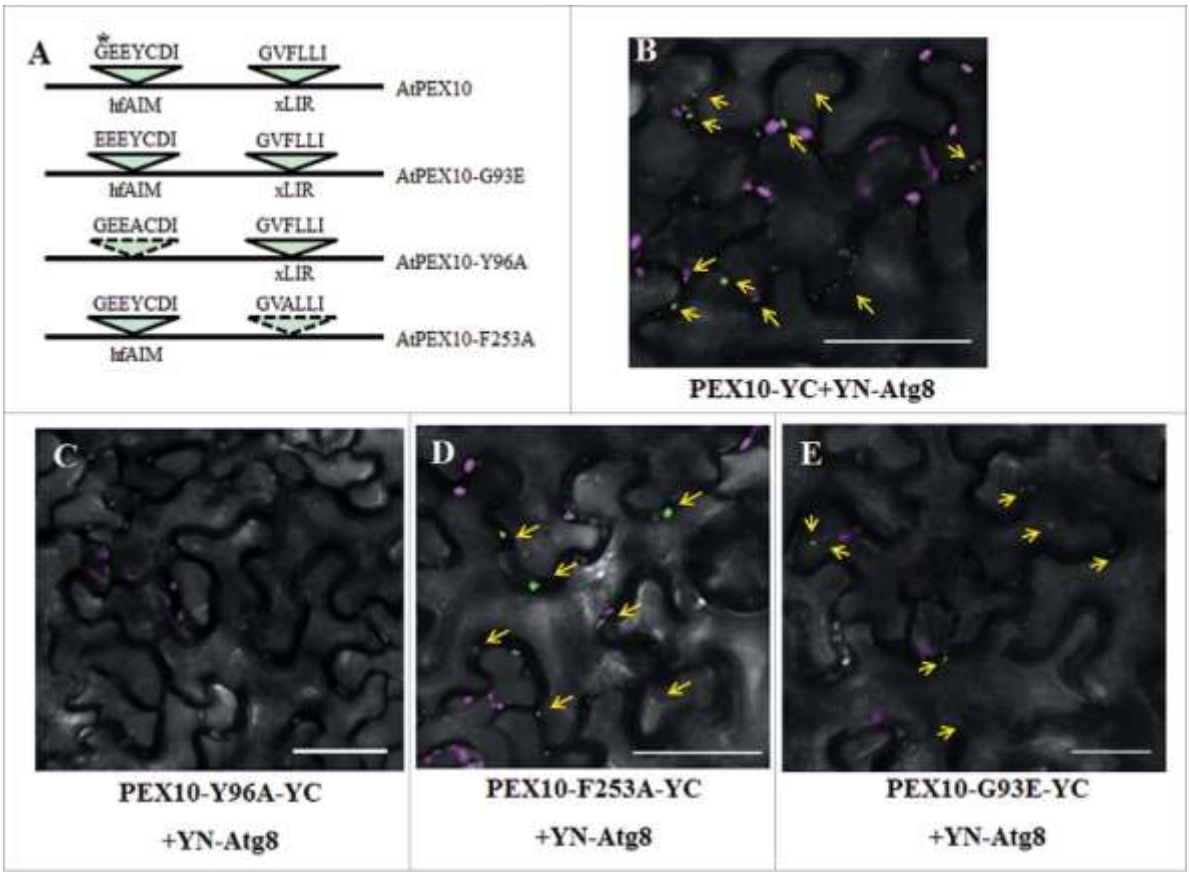

Figure 3. Identification of the functional AIM in AtPEX10. (**A**) Schematic representation of the AIM sequences predicted by either hfAIM or iLIR in AtPEX10 and their mutants. The referred G-to-E mutation is denoted by a star. (**B to E**) BiFC analysis was performed following transient coexpression of YN-Atg8f and different variants of PEX10-YC in *N. benthamiana* leaves. The interaction of wild-type PEX10 with Atg8f results in YFP fluorescence (in green) (**B**). No interaction was observed when YN-Atg8f was coexpressed with PEX10^{Y96A}-YC (**C**), demonstrating that the GEEYCDI AIM motif is necessary and sufficient for Atg8f interaction. On the other hand, mutating the AIM motif identified by iLIR did not abolish PEX10 interaction with Atg8f (**D**), suggesting that this putative AIM is not required for the interaction. The G93E mutation in the hfAIM predicted motif retained the interaction between PEX10 and Atg8f (**E**). Chloroplast autoflourescence is shown in magenta. Bar: 20 µm.

cotransformed *N. benthamiana* leaves with YN-Atg8 and either AtPEX10^{Y96A}-YC or AtPEX10^{F253A}-YC. The results indicated that while the point mutation in AtPEX10^{Y96A} abolished its interaction with Atg8, the point mutation in AtPEX10^{F253A} did not affect its Atg8 binding (**Fig. 3C and 3D**). As a G-to-E point mutation in the hfAIM predicted AIM motif in AtPEX10 was previously shown to cause a peroxisome deficient phenotype,[21] we also generated an AtPEX10^{G93E}-YC construct, and then transiently coexpressed this construct

together with YN-Atg8 in tobacco leaves. The G93E mutation in AtPEX10 did not influence the interaction between AtPEX10 and Atg8 in the BiFC assay (**Fig. 3E**). Together, we concluded that AtPEX10 interacts with Atg8 *in vivo* and that the AIM motif identified by hfAIM is the functional AIM in AtPEX10.

Identification of putative AIMs in members of the *Arabidopsis* PEX family

Triggered by the above results, we employed both the hfAIM and the iLIR systems to elucidate whether additional AtPEX proteins also contain putative AIMs. The hfAIM and iLIR systems identified 20 AIMs in 13 AtPEX proteins, including the AtPEX10 protein (**Table 1**). Further analysis demonstrated that 9 AtPEX proteins contain 12 AIMs in total based on the hfAIM system, including AtPEX1, AtPEX3-2, AtPEX5, AtPEX6, AtPEX7, AtPEX10, AtPEX14, AtPEX17 and AtPEX19-1 (**Table 1**). Utilization of the iLIR system identified 12 xLIR motifs that were present in AtPEX1, AtPEX5, AtPEX10, AtPEX11C, AtPEX11D, AtPEX11E, AtPEX12 and AtPEX17 (**Table 1**). Among these AIMs, 3 hfAIM motifs that are present in AtPEX1 as well as one hfAIM motif in AtPEX17 were also predicted by the iLIR system. The rest of the AIMs differed between the hfAIM and iLIR systems (**Table 1**). AtPEX1, AtPEX5, AtPEX10 and AtPEX17 were predicted to contain AIMs by both systems, and are therefore considered as Atg8-interacting proteins with higher confidence.

Evolutionary conservation of PEX1, PEX6 and PEX10 AIMs

We were next interested to elucidate whether any of the 20 AIMs mentioned above have been evolutionary conserved. To address this issue, we compared the sequences of the PEX proteins that were predicted to contain AIMs from multiple organisms represented in the Peroxisome DB 2.0 (www.peroxisomedb.org/).[31] Sequence alignment revealed that only PEX1, PEX6 and PEX10 contained highly conserved AIMs (**Table S4**). The hfAIM predicted sequence "GEEYCDI" in AtPEX10 (hfAIM pattern 1, see **Fig. 2A**) was found in the conserved PEX10 regions of 92% organisms that were analyzed (**Fig. S3**). The hfAIM predicted sequences "EDDWEVL" and "FEDFDSI" in AtPEX1 (hfAIM pattern 1, 2, 4 and 5, see **Fig.**

Table 1. Identification of putative AIMs in AtPEX proteins by hfAIM and iLIR.

Protein	Gene ID	Description	hfAIM[1]	xLIR[2]	PSSM[3]
AtPEX1	AT5G08470	Peroxisomal AAA-ATPases	EDDWEVL* (1, 2, 4, 5)	DDWEVL*	26 (1.2e-03)
			ND	STYVDV	12 (1.1e-01)
			RLGWEDV* (3)	LGWEDV*	16 (3.0e-02)
			FDEFDSI* (1, 5)	DEFDSI*	15 (4.2e-02)
AtPEX6	AT1G03000	Peroxisomal AAA-ATPases	VIFFDEL (3)	ND	5 (1.0e+00)
AtPEX10	AT2G26350	ZN RING proteins	ND	GVFLLI	6 (7.4e-01)
			GEEYCDI (1)	ND	9 (2.8e-01)
AtPex3-2	AT1G48635	Peroxisome membrane assembly	LQLWDEL (3)	ND	14 (5.7e-02)
AtPex5	AT5G56290	Peroxisome targeting sequence binding	GAAWDEV (3)	ND	15 (4.2e-02)
			ND	PMFEPV	10 (2.0e-01)
AtPex7	AT1G29260	Peroxisome targeting sequence binding	AHDFEIL (5)	ND	14 (5.7e-02)
			STGWDEL (3)	ND	18 (1.6e-02)
AtPex14	AT5G62810	Peroxisome docking	RKYFEDL (3)	ND	7 (5.3e-01)
AtPex17	AT4G18197	Peroxisome docking	ND	PSFTTL	11 (1.5e-01)
			SGEWETL* (5)	GEWETL*	20 (8.4e-03)
AtPex19-1	AT3G03490	Peroxisome membrane assembly	LKQFEDL (3)	ND	9 (2.8e-01)
AtPEX2	AT1G79810	ZN RING proteins	ND	ND	-
AtPEX3-1	AT3G18160	Peroxisome membrane assembly	ND	ND	-
AtPex4	AT5G25760	Peroxisome matrix protein import	ND	ND	-
AtPEX11A	AT1G47750	Peroxisome division-proliferation	ND	ND	-
AtPEX11B	AT3G47430	Peroxisome division-proliferation	ND	ND	-
AtPEX11C	AT1G01820	Peroxisome division-proliferation	ND	STFLFL	12 (1.1e-01)
AtPEX11D	AT2G45740	Peroxisome division-proliferation	ND	STFLFL	12 (1.1e-01)
AtPEX11E	AT3G61070	Peroxisome division-proliferation	ND	STFLFL	12 (1.1e-01)
AtPEX12	AT3G04460	ZN RING proteins	ND	FTYQLL	12 (1.1e-01)
AtPEX13	AT3G07560	Peroxisome docking	ND	ND	-
AtPEX16	AT2G45690	Peroxisome membrane assembly	ND	ND	-
AtPex19-2	AT5G17550	Peroxisome membrane assembly	ND	ND	-
AtPEX22	AT3G21865	Peroxisome matrix protein import	ND	ND	-

[1]hfAIM, high fidelity AIM prediction. The numbers in brackets correspond to the hfAIM patterns indicated in Figure 2A; [2]xLIR was predicted by iLIR online website; [3]PSSM values were derived from the iLIR system. *, stars indicate that the AIMs were predicted by both hfAIM and iLIR; ND, no AIMs predicted by either hfAIM or iLIR.

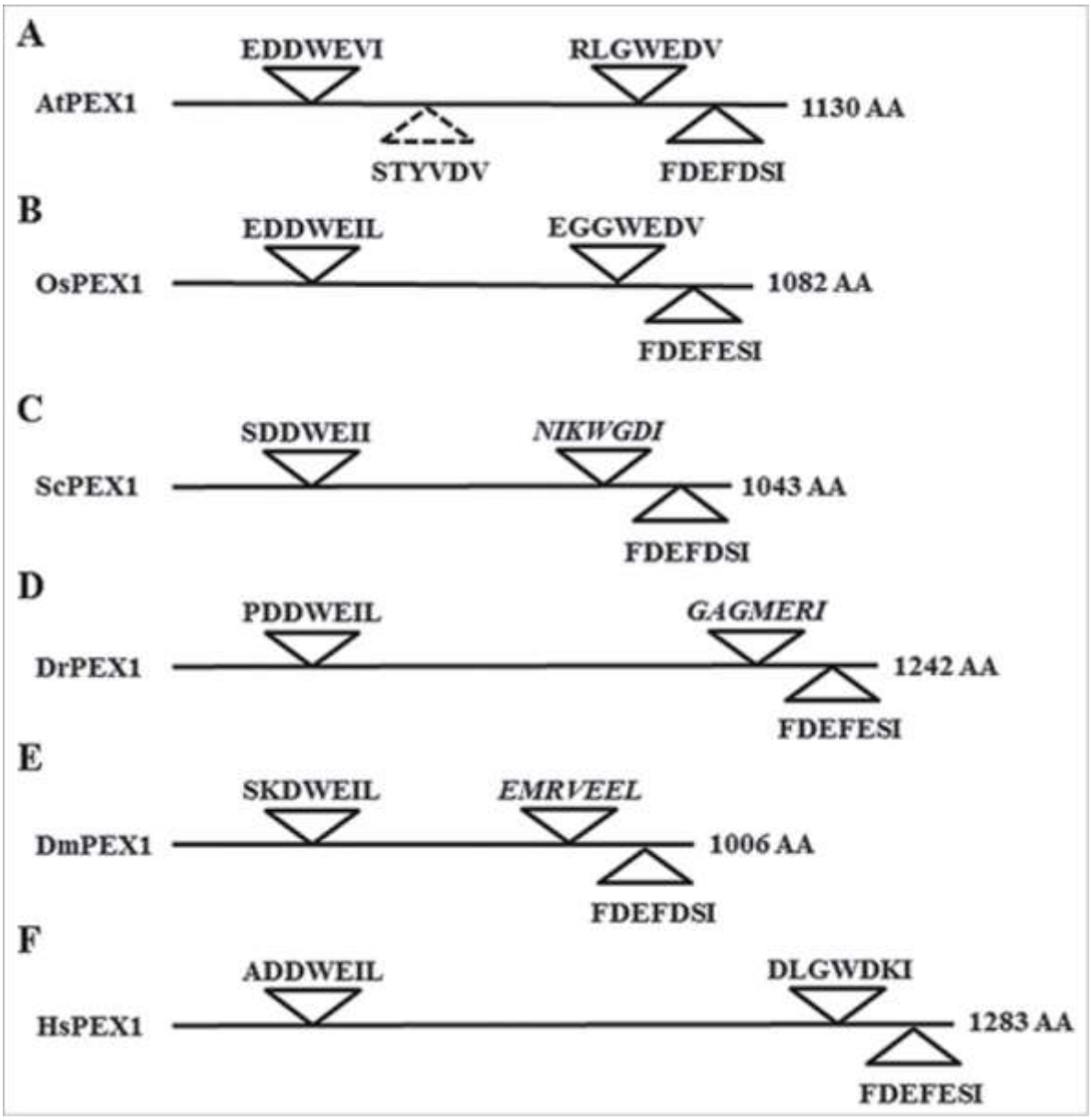

Figure 4. Evolutionarily conserved AIMs predicted in PEX1 by either hfAIM or iLIR. Schematic representation of the PEX1 proteins from 6 representative organisms and their AIMs sequences as predicted by either hfAIM or iLIR. The AtPEX1 AIM that was predicted only by iLIR is indicated by a dashed angle. The sequences in italics indicate non conserved AIM sequences. At, the dicot plant *Arabidopsis* (*Arabidopsis thaliana*); Os, the monocot plant rice (*Oryza sativa*); Sc, yeast (*Saccharomyces cerevisiae*); Dm, *Drosophila* (*Dorosophila melanogaster*); Dr, Zebrafish (*Danio rerio*); Hs, human (*Homo sapiens*). AA, amino acid.

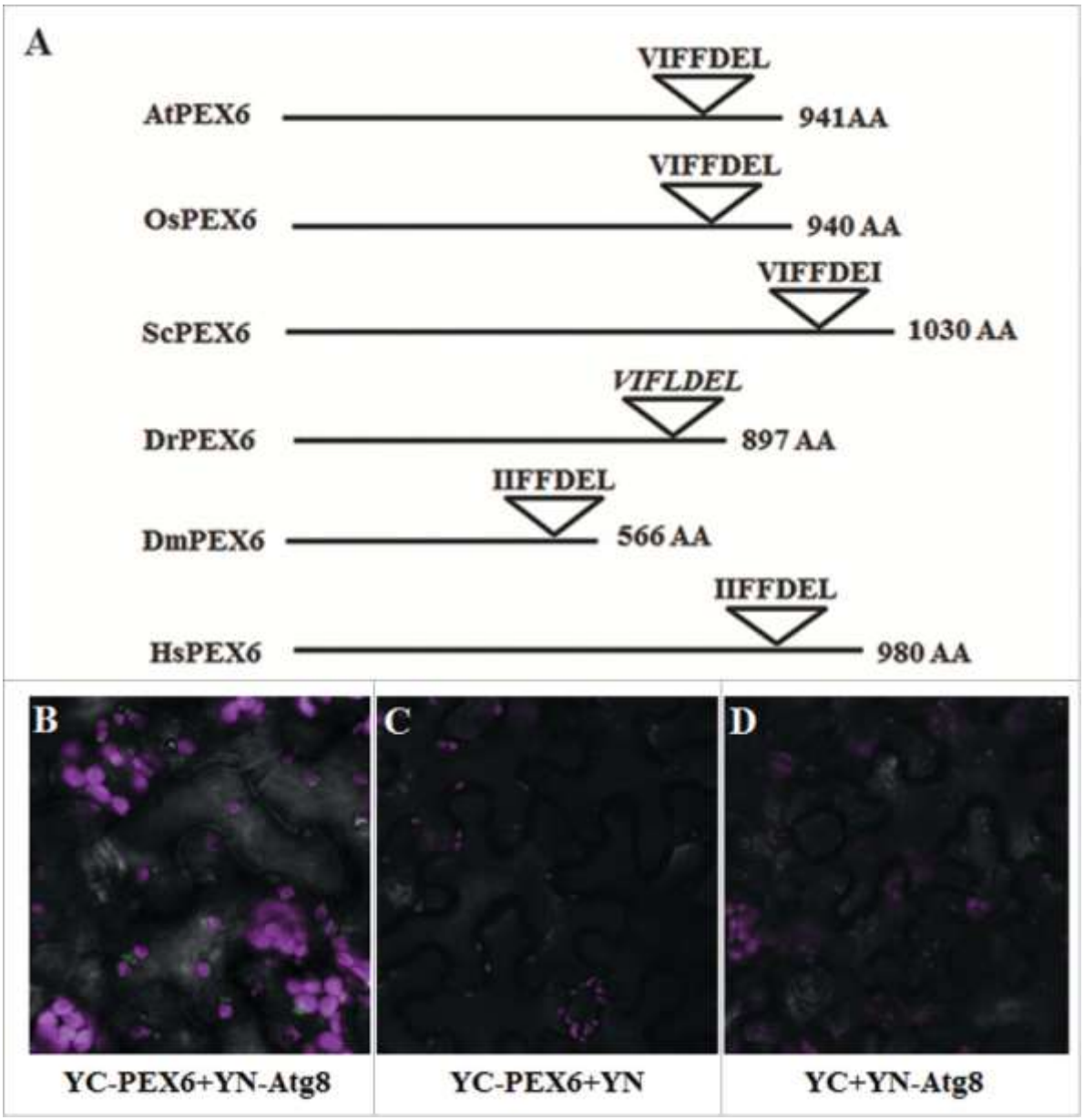

Figure 5. AtPEX6 interacts with Atg8 *in vivo*. (**A**) Schematic representation of the PEX6 proteins from 6 representative organisms and the AIMs sequences present in them as predicted by hfAIM. The sequence in italics indicates AIM that was not recognized by hfAIM. At, the dicot plant *Arabidopsis* (*Arabidopsis thaliana*); Os, the monocot plant rice (*Oryza sativa*); Sc, yeast (*Saccharomyces cerevisiae*); Dm, *Drosophila* (*Dorosophila melanogaster*); Dr, Zebra fish (*Danio rerio*); Hs, human (*Homo sapiens*). AA, amino acid. (**B**) BiFC analysis was performed following transient coexpression of YN-Atg8f and YC-PEX6 in *N. benthamiana* leaves. YFP fluorescence (green bodies indicated by arrows) demonstrates the interaction of PEX6 with Atg8f. Transient coexpression of YC-PEX6 with unfused YN (**C**) or coexpression of YN-Atg8f with unfused YC (**D**) did not result in any YFP fluorescence demonstrating the specificity of the interaction. Chloroplast autoflourescence is shown in magenta. Bar: 20 µm.

2A) were found in the conserved PEX1 regions of 76% or 89% organisms that were analyzed, respectively (**Fig. S4**), and the predicted hfAIM "VIFFDEL" sequence in AtPEX6 (hfAIM pattern 3, **Fig. 2A**) was found in the conserved PEX6 regions of 93% organisms that were analyzed (**Figure S5**). Interestingly, 2 out of the 3 AIMs in AtPEX1 that were identified both by hfAIM and iLIR were evolutionary conserved, while the third predicted AIM was not conserved among various species (**Fig. 4**, schematic representation of the sequences of only 6 representative organisms is presented). Moreover, the conserved AIM in AtPEX6 was only identified by hfAIM, while the iLIR detected no such motif in AtPEX6 (**Table 1 and Fig. 5A**). To determine the accuracy of the hfAIM prediction, we performed a BiFC assay to look at the potential interaction of Atg8 with AtPEX6 *in vivo*. As shown in Figure 5B, transient coexpression of YN-Atg8 and YC-AtPEX6 results in YFP fluorescence that is visualized as punctate spherical structures similar in size to plant peroxisomes, while neither the cotransformation of YC-AtPEX6 with YN nor YN-Atg8 and YC yielded any signal (Figure 5C and 5D). These results indicate that AtPEX6 indeed binds to Atg8 *in vivo* as predicted by hfAIM. In addition, a "LQLWDEL" sequence in AtPEX3-2 was also predicted by hfAIM as a potential AIM motif (hfAIM pattern 3, **Fig. 2A**), and this pattern appeared in 56% organisms that were analyzed (**Fig. S6**). In respect to the rest of the AIMs identified in the other PEX proteins by either the hfAIM or iLIR systems, all of these AIMs were not conserved in evolution (**Fig. S7 to S11 and Table S4**). Taken together, our results suggest that AtPEX6 and AtPEX10, and likely also AtPEX1 and AtPEX3-2 interact with Atg8 *in planta*. Furthermore, the autophagy mechanism underlying the degradation of these AtPEX proteins is apparently highly evolutionary conserved.

Developmental phenotypes of *pex1*, *pex6* and *pex10* mutants are apparently associated with mutations occurring within or nearby their AIMs

A number of *atg* mutants and their corresponding genes have already been isolated and well characterized in plants.[3] Interestingly, recent studies revealed that peroxisome degradation is noticeably attenuated in the backgrounds of autophagy mutants,[15, 18, 32] implying that autophagy is involved in the degradation of peroxisomes, possibly through the interactions of Atg8 with peroxisome proteins possessing AIMs. Based on this assumption, we attempted to figure out whether mutations occurring in AIMs of PEX proteins influence the interaction of

Table 2. Mutations associated with or nearby AIMs in *Arabidopsis* and human PEX1, PEX6 and PEX10.

Protein	Species	Function	Mutation	AIM Association	Phenotype	Refs.
PEX10	At[1]	Peroxisome elongation	G_{98} to E	TL**G**EEYCDIIQV	Vermiform peroxisome	22
PEX10	Hs[2]	Peroxisome elongation	T_{68} to N	**T**LGEEYVSIIQV	Possibly damaging	34
PEX10	Hs[2]	Peroxisome elongation	G_{70} to R	TL**G**EEYVSIIQV	Possibly damaging	34
PEX10	Hs[2]	Peroxisome elongation	E_{71} to K	TLG**E**EYVSIIQV	Possibly damaging	34
PEX10	Hs[2]	Peroxisome elongation	Q_{78} to R	TLGEEYVSII**Q**V	Possibly damaging	34
PEX1	Hs[2]	Pex5 dislocation	R_{949} to W	FDEFESIAPR**R**	Possibly damaging	34
PEX6	At[1]	Pex5p dislocation	R_{766} to Q	FDELDSLAPA**R**	PTS2 processing defect	35

[1]*Arabidopsis thaliana*; [2]*Homo sapiens*. The mutated amino acid is indicated in bold.

Atg8 with these proteins, eventually leading to abnormal peroxisome phenotypes. According to the Polyphen prediction method,[33] a R949W mutation in the *homo sapiens* PEX1 protein has been suggested to cause a "probably damaging" phenotype.[34] Intriguingly, this mutation is located close to an hfAIM predicted AIM (**Table 2**). In AtPEX6 protein, a mutation of the conserved R766 residue, also located close to an hfAIM predicted AIM motif, to Q (*pex6-1)* (**Table 2**), has been reported to cause peroxisomal targeting signaling (PTS2) processing defect in *Arabidopsis*.[35] Notably, a distinct mutation in AtPEX6 (*pex6-2*) results in similar physiological responses, such as resistance to inhibition by 2,4-dichlorophenoxybutyric acid (2, 4-DB) and to the promotive effects by protoauxin indole-3-butyric acid (IBA), but has no effect on peroxisomal matrix proteins import.[36] Thus, it is possible that the effect on matrix proteins import is related to changes in the autophagic degradation of *Atpex6-1*. Yet, the most striking findings were several mutations that occur within the hfAIM predicted AIM in PEX10 (**Table 2**). Prestele and associates[21] identified a dominant negative G93E mutant in PEX10 that exhibited vermiform peroxisome shapes. Since it is proposed that the presence of acidic amino

acids within or nearby the core AIM would increase the strength of binding of Atg8 to this AIM,[4, 7] the G93E mutation might increase the Atg8-mediated turnover of AtPEX10. Indeed, the quantity of peroxisomes was reduced in the PEX10^{G93E} mutant-supporting our hypothesis that the G93E mutation enhances the binding efficiency of Atg8 to the mutated AtPEX10, thus increasing the turnover of peroxisomes by pexophagy. Additionally, 4 other mutations that occurred within or nearby this AIM cause possibly damaging phenotypes according to the Polyphen prediction.[34] Among these 4 mutations, an E71K mutation is expected to reduce the strength of Atg8 binding to AtPEX10 due to the conversion of the negatively charged E residue to a positively charged K residue (E71K; **Table 2**). Taken together, the information described above supports our above mentioned results, suggesting that PEX1, PEX6 and PEX10 are turned over by autophagy via the interactions with Atg8.

DISCUSSION

Autophagy is an evolutionarily conserved process in eukaryotic organisms, including animals and plants. Accumulating evidence suggests that most of the proteins that are selectively turned over by autophagy contain one or multiple Atg8-interacting motif (AIM).[3, 7, 23] The core consensus of the AIM motif is F/W/Y-X-X-L/I/V.[4] Using this degenerate consensus sequence, it is possible to screen and identify AIM-containing proteins (ACPs) on a genome wide scale by bioinformatics approaches.[10,11,24] But, as the consensus AIM motif is short and degenerate, a simple search for this motif will likely generate multiple false positive results. Thus, generating reliable, high fidelity bioinformatics tools will minimize the experimental work required to verify the predictions and therefore it is highly desirable.

The binding of Atg8 to the AIMs present in various proteins was shown to be enhanced by negative charge. Thus the presence of aspartate or glutamate, or phosphorylated Serine and Threonine residues, either immediately upstream or within the core AIM, strengthen the interaction between Atg8 and the AIM motif.[7] This finding is supported by our sequence analysis of experimentally verified AIM motifs that indicated that these motifs are highly enriched in D and E residues (**Fig. 1**). Based on this information we developed a bioinformatics tool, hfAIM, that uses a definition of an AIM motif as a sequence of 7 amino acids, X-X-X-F/W/Y-X-X-L/I/V, that contains at least 2 acidic amino acids. Although there is

no direct evidence to support the negative influence of positively charged amino acids on the binding strength of Atg8 to the AIM, we still excluded these amino acids in the regular patterns of our hfAIM system in an attempt to improve the reliability of AIM prediction (**Fig. 2**). Note that in the present report, we only considered the contribution of the acidic amino acids to Atg8 binding in our prediction scheme. This restrain might enhance the fidelity of the AIM prediction with the expense of reducing the sensitivity, as the contribution of the S and T residues is not taken into account.

Another recently developed bioinformatics tool for the identification of AIM motifs in proteins, the iLIR tool,[11] is also based on regular expression patterns. iLIR defines an AIM motif as a sequence of 6 amino acids, X-X-F/W/Y-X-X-L/I/V, where the permitted residues at any given "X" position are based on multiple sequence alignment of verified AIM motifs. A comparison of iLIR and hfAIM predictions using a dataset of verified AIM motifs suggested that while iLIR has a better sensitivity, hfAIM is more stringent (**Fig. 2**). Furthermore, hfAIM was somewhat better at identifying AIM motifs in verified *Arabidopsis* Atg8-interacting proteins (**Table S2**). Unfortunately, experimental information about the specific sequences needed for the Atg8 interaction of the verified Atg8 interacting proteins from *Arabidopsis* is still scarce. However, our BiFC analysis suggested that the hfAIM motif identified in AtPEX10 is indeed a functional AIM and is necessary for the interaction with Atg8, while the iLIR identified AIM is not (**Table S2 and Fig. 3**). Furthermore, a single AIM was identified solely by hfAIM in AtPEX6, and BiFC analysis indeed verified AtPEX6 - Atg8 interaction (**Table S2 and Fig. 5**). It is possible that as iLIR regular expression patterns were defined using mostly non-plant verified AIMs, it does not represent well the composition of amino acids in plant AIM motifs and therefore it is not best suited to identify AIMs in plants.

Recently, the role of autophagy in multiple biological processes has been characterized in *Arabidopsis* by omics analysis.[25, 26] However, the identification of the specific and/or novel components regulated by autophagy from the large-scale data is still not entirely resolved. Our hfAIM system can be used as a complementary approach to identify potential ACPs and may provide new insight into the regulation of autophagy-mediated degradation processes. For instance, differentially expressed proteins (DEPs) identified by proteomic analysis in *Arabidopsis* can be further analyzed by the hfAIM system to look at their potential regulation by the autophagy apparatus. Those DEPs containing hfAIM motifs could be considered as

potential Atg8 interacting candidates, and then selected for further analysis. However, it is important to remember that Atg8 binding to target proteins in not always mediated by a typical AIM motif.[5] For example, the verified Atg8 binding proteins Calcium-binding and coiled-coil domain-containing protein 2 (UniProtKB accession: CACO2_HUMAN) and Tax1-binding protein 1 (UniProtKB accession: TAXB1_HUMAN) do not contain a typical AIM motif (**Table S1**). Therefore, the characterization of additional verified functional AIMs by experimental methods is essential for the development of more accurate prediction tools.

The fidelity of AIMs prediction by hfAIM can be improved by several approaches. One possible approach could be to combine hfAIM predictions with iLIR predictions, and to consider as more promising AIM motifs that were also predicted by iLIR, or had PSSM scores according to Kalvari et al above a defined cut-off value.[11] As demonstrated using the verified AIM motifs dataset (**Table S1**), the use of a PSSM cut-off value improved the specificity for both hfAIM and iLIR (**Fig. 2B** and Kalvari et al.[11]). However, the use of this higher level of stringency can lead to false negative results, as demonstrated with AtPEX10. Though the PSSM score of the predicted hfAIM in AtPEX10 is quite low (PSSM=9), and therefore it would be regarded as a low confidence AIM, the protein was experimentally shown to interact with Atg8 (**Fig. 3**) through this predicted hfAIM.

Prediction of potential AIMs by bioinformatics tools can also be strengthened by additional lines of evidence such as evolutionary conservation of the AIMs.[11, 24] This approach assumes that a functional AIM will be conserved across species in proteins that are involved in selective autophagy-mediated degradation processes or whose homeostasis is regulated by selective autophagy. The peroxisomal PEX proteins were used as a case study to evaluate this approach. Peroxisomes are highly dynamic organelles functioning in multiple biological processes.[20] Furthermore, pexophagy, the process of selective degradation of peroxisomes by autophagy, is an essential mechanism regulating the homeostasis of the peroxisomes. Although pexophagy in plants was demonstrated by several recent reports,[14, 17, 18, 32] the regulatory mechanisms underlying the degradation of peroxisomes still await additional studies. Indeed, the colocalization of Atg8 and peroxisomes, specifically aggregated peroxisomes,[14, 17, 18] suggests that Atg8 interacts with certain peroxisome PEX proteins, leading to their specific autophagy-mediated transport, or the transport of the entire peroxisome to the vacuole for degradation. Therefore, we used both the hfAIM and the iLIR tools to identify potential AIMs

in the *Arabidopsis* family of PEX proteins (**Table 1**), and then compared the sequences of PEX proteins from 38 different organisms and used them to analyze the evolutionary conservation of the predicted AIMs in these proteins (**Fig. S3 to S11**). Out of the 22 AtPEX proteins, 13 proteins contain AIM motifs according to either hfAIM or iLIR (**Table 1**), but only the AIM motifs in PEX1, PEX6 and PEX10 are highly conserved across species (**Fig. S3 to S5**). PEX6 encodes a peroxisomal AAA-ATPase, and forms a complex with PEX1 that participates in peroxisomal matrix proteins import.[35, 37] Only a single, highly conserved AIM was detected in PEX6 by our hfAIM system, and this AIM is located in the Walker B domain of PEX6 (**Table 1**). Indeed, our BiFC results verified that PEX6 does interact with Atg8 *in planta* (**Fig. 5**), supporting the functional role of the conserved AIM motif. Therefore, we propose that this conserved amino acid sequence in PEX6 may have a dual function, either serving as a Walker B domain and/or serving as an AIM, allowing the binding of Atg8 to this protein. AtPEX1 contains 3 hfAIMs, which are also recognized by iLIR and an additional motif that is recognized solely by iLIR (**Table 1**). Two of these 4 motifs are highly conserved across species (**Fig. 4 and Fig. S4**) implying that PEX1 interacts with Atg8 at a reasonable confidence.

PEX10 is involved in both peroxisome formation and matrix protein import.[21, 22] The evolutionary highly conserved amino acid sequences in PEX10 are thought to be essential for peroxisome biogenesis and plant development.[21] Looking further into PEX10 proteins from various organisms, we found that all of them contain a highly conserved AIM (GEEYCDI) recognized by our hfAIM method, but not by the iLIR system (**Table 1**). A G93E mutation in PEX10 that causes vermiform peroxisome shape [21] occurs in the minus 3 position of the core hfAIM (**Fig. 3A**). Moreover, this mutation also leads to a lower number of peroxisomes,[21] as expected from a mutation that strengthen Atg8-PEX10 interaction and therefore will lead to increase turnover by autophagy. Indeed, BiFC experiments confirmed the functional role of this conserved AIM in PEX10 interaction with Atg8 *in planta* (**Fig. 3**). Interestingly, an additional AIM, located in another position of the PEX10 protein is predicted by the iLIR system. Yet, this xLIR motif is not evolutionary conserved in PEX10 proteins of the various lower and higher organisms that have been studied (**Fig. S3**). Moreover, our BiFC analysis also verified that, unlike the conserved AIM, this xLIR motif is not necessary for Atg8 binding to PEX10 (**Fig. 3**). The rest of the AIMs present in the other PEX proteins are not conserved in

evolution (**Fig. S6 to S11**). Therefore, we propose that PEX6, PEX10 and possibly also PEX1, are likely to interact with Atg8 through their evolutionary conserved AIMs, and that selective autophagy in probably involved in the turnover of these proteins.

In the present study, 9 AtPEX proteins were identified by our hfAIM approach as potential Atg8-interacting proteins (**Table 1**). Interestingly, the iLIR system also identified 8 AtPEX proteins as potential Atg8-interacting proteins. However, AIMs in PEX3 and PEX14, which interact with Atg8 in mammalian cells,[38, 39] were recognized only by the hfAIM system (**Table 1**). Though the predicted AIM motif in AtPEX14 does not seem to be very well conserved among other organisms (**Fig. S9**), it was recently shown that Atg8 colocalized with AtPEX14 in peroxisome aggregates, suggesting that the interaction between Atg8 and PEX14 is conserved. Furthermore, though fission events are suggested to be involved in the degradation of yeast peroxisomes following protein aggregation,[40] no AIM motifs were predicted by hfAIM in AtPEX family members involved in peroxisome division-proliferation (**Table 1**). Interestingly, the iLIR tool did predict xLIR motifs in several division-proliferation AtPEX proteins and these motifs contained S and T residues, strengthening the notion that S and T might contribute to Atg8 binding of AIMs. Alternatively, as recent studies demonstrate that some proteins do not require a typical AIM motif to bind Atg8,[5, 41, 42] these proteins might interact with Atg8 in an AIM-independent manner.

In summary, we have generated a high fidelity bioinformatics tool, termed hfAIM, available as a web tool (http://bioinformatics.psb.ugent.be/hfAIM/) for *in silico* genome-wide prediction of AIMs in proteins. Using hfAIM it is possible to perform fast and reliable genome-wide screening of AIM-containing proteins that may be regulated by autophagy, and to select candidates for further studies using experimental approaches. This bioinformatics approach can facilitate a better understanding of the contribution of autophagy to multiple biological processes in various organisms. Using PEX proteins as a test case, our investigations indicate that PEX1, PEX6 and PEX10 are selectively turned over by autophagy. More specifically, our results may also shed a new light on the regulatory mechanism(s) underlying how Atg8 coordinates the homeostasis of specific PEX proteins as well as the operation of pexophagy.

MATERIALS and METHODS

Identification of AIMs within proteins

To identify AIMs that meet the standards determined in Figure 2A within proteins, we adapted the stand-alone version of the PatMatch software.[43] This program is available for download at The *Arabidopsis* Information Resource (TAIR) at ftp://ftp.arabidopsis.org/home/tair/Software/Patmatch/). In house scripts were tailored to calculate the percentage of *Arabidopsis* proteins possessing either zero, 1, 2, 3 or more AIMs in the entire *Arabidopsis* proteome. Position specific scoring matrices (PSSM), were calculated using the iLIR web interface provided by Kalvari et al (http://repeat.biol.ucy.ac.cy/iLIR/).[11]

Plasmids construction

To generate PEX6-YC, PEX10-YC and YN-Atg8, we used the pSAT vector system for BiFC assays.[44] Cloning was done with the In-Fusion kit (Clontech, 639649) according to the manual instructions by using the corresponding primers (**Table S3**). The resulting clones were finally introduced into the pPZP vectors as previously described.[30] To generate the PEX10^{E95A}-YC, PEX10^{E95H}-YC, PEX10^{Y96A}-YC and PEX10^{F253A}-YC, we also used the pSAT vector system (https://www.arabidopsis.org/abrc/catalog/vector_2.html) for BiFC assays.[44] Mutagenesis of AtPEX10 was generated via substituting the corresponding amino acids by specific primers (**Table S5**) and the corresponding binary vector was produced as described above.

Multiple alignments

To test the homology of the protein sequences among various organisms, PEX protein sequences derived from the PeroxisomeDB (http://www.peroxisomedb.org/home.jsp) were aligned by the ClustalW method in the MEGA 6 software with defaults settings.[45] To simplify the presentation, only the alignment of a window of 10 to 30 amino acids that includes the predicted AIM motifs is shown in Figures S3 to S11.

BiFC assay and confocal microscopy

To verify the interactions of ATG8f with either AtPEX6 or AtPEX10, we used *Agrobacterium* strains harboring each of the following plasmids separately: YN-Atg8 with either PEX6-YC or PEX10-YC were transiently cotransformed in *Nicotiana benthamiana* leaves as previously described.[30] For analysis of the interactions of ATG8f with the AtPEX10 mutants PEX10^{E95A}-YC, PEX10^{E95H}-YC, PEX10^{Y96A}-YC or PEX10^{F253A}-YC, we employed the same approach as above. Confocal microscopy analysis was performed using Olympus Fluoview 1000 IX81 (Olympus Life Science, Tokyo, Japan) and the Nikon A1 (Nikon, Japan) systems as previously described.[30] Briefly, samples were put between 2 microscope glass cover slips. Images were taken from a single focal plane unless otherwise indicated. GFP fluorescence images were taken using 488-nm laser excitation and the emission was collected via the 525-nm filter. Chlorophyll autofluorescence was taken with the 640-nm laser and collected with the 700-nm filter. Acquired images were analyzed by either Olympus Fluoview 1000 viewer or the NIS-Elements AR imaging software.

ACKNOWLEDGEMENTS

The authors thank Omrit Zemach and Dana Averbuch for excellent technical assistance. Our research was supported by grants from The Israeli Ministry of Agriculture, The Israel Science Foundation (grant No.395/11), and the J & R center for scientific research at the Weizmann Institute of Science. GG is an incumbent of the Bronfman Chair of Plant Science at the Weizmann Institute of Science.

REFERENCES

1. Avin-Wittenberg T, Honig A, Galili G. Variations on a theme: plant autophagy in comparison to yeast and mammals. Protoplasma 2012; 249:285-99.
2. Yoshimoto K, Takano Y, Sakai Y. Autophagy in plants and phytopathogens. FEBS Lett 2010; 584:1350-8.
3. Li F, Vierstra RD. Autophagy: a multifaceted intracellular system for bulk and selective recycling. Trends in plant science 2012; 17:526-37.
4. Noda NN, Kumeta H, Nakatogawa H, Satoo K, Adachi W, Ishii J, et al. Structural basis of target recognition by Atg8/LC3 during selective autophagy. Genes to cells : devoted to molecular & cellular mechanisms 2008; 13:1211-8.

5. Behrends C, Sowa ME, Gygi SP, Harper JW. Network organization of the human autophagy system. Nature 2010; 466:68-76.
6. Noda NN, Ohsumi Y, Inagaki F. Atg8-family interacting motif crucial for selective autophagy. FEBS letters 2010; 584:1379-85.
7. Schreiber A, Peter M. Substrate recognition in selective autophagy and the ubiquitin-proteasome system. Biochimica et biophysica acta 2014; 1843:163-81.
8. Kumeta H, Watanabe M, Nakatogawa H, Yamaguchi M, Ogura K, Adachi W, et al. The NMR structure of the autophagy-related protein Atg8. Journal of biomolecular NMR 2010; 47:237-41.
9. Wild P, Farhan H, McEwan DG, Wagner S, Rogov VV, Brady NR, et al. Phosphorylation of the autophagy receptor optineurin restricts Salmonella growth. Science 2011; 333:228-33.
10. Tzfadia O, Galili G. The Arabidopsis exocyst subcomplex subunits involved in a golgi-independent transport into the vacuole possess consensus autophagy-associated atg8 interacting motifs. Plant signaling & behavior 2013; 9.
11. Kalvari I, Tsompanis S, Mulakkal NC, Osgood R, Johansen T, Nezis IP, et al. iLIR: A web resource for prediction of Atg8-family interacting proteins. Autophagy 2014; 10:913-25.
12. Meszaros B, Simon I, Dosztanyi Z. Prediction of protein binding regions in disordered proteins. PLoS computational biology 2009; 5:e1000376.
13. Dosztanyi Z, Meszaros B, Simon I. ANCHOR: web server for predicting protein binding regions in disordered proteins. Bioinformatics 2009; 25:2745-6.
14. Yoshimoto K, Shibata M, Kondo M, Oikawa K, Sato M, Toyooka K, et al. Organ-specific quality control of plant peroxisomes is mediated by autophagy. Journal of cell science 2014; 127:1161-8.
15. Shibata M, Oikawa K, Yoshimoto K, Goto-Yamada S, Mano S, Yamada K, et al. Plant autophagy is responsible for peroxisomal transition and plays an important role in the maintenance of peroxisomal quality. Autophagy 2014; 10:936-7.
16. Lee HN, Kim J, Chung T. Degradation of plant peroxisomes by autophagy. Frontiers in plant science 2014; 5:139.
17. Shibata M, Oikawa K, Yoshimoto K, Kondo M, Mano S, Yamada K, et al. Highly oxidized peroxisomes are selectively degraded via autophagy in Arabidopsis. The Plant cell 2013; 25:4967-83.
18. Kim J, Lee H, Lee HN, Kim SH, Shin KD, Chung T. Autophagy-related proteins are required for degradation of peroxisomes in Arabidopsis hypocotyls during seedling growth. The Plant cell 2013; 25:4956-66.
19. Kim J, Lee HN, Chung T. Plant cell remodeling by autophagy: switching peroxisomes for green life. Autophagy 2014; 10:702-3.
20. Hu J, Baker A, Bartel B, Linka N, Mullen RT, Reumann S, et al. Plant peroxisomes: biogenesis and function. The Plant cell 2012; 24:2279-303.
21. Prestele J, Hierl G, Scherling C, Hetkamp S, Schwechheimer C, Isono E, et al. Different functions of the C3HC4 zinc RING finger peroxins PEX10, PEX2, and PEX12 in peroxisome formation and matrix protein import. Proc Natl Acad Sci USA 2010; 107:14915-20.
22. U. S, Prestele J, Henriette O, Brueggeman R, Wanner G, Gietl C. Requirement of the C3CH4 zing RING finger of the Arabidopsis PEX10 for photorespiration and leaf

peroxisome contact with chloroplasts. Proceedings of the National Academy of Sciences of the United States of America 2007; 104:1069-74.
23. Birgisdottir AB, Lamark T, Johansen T. The LIR motif - crucial for selective autophagy. Journal of cell science 2013; 126:3237-47.
24. Cvrckova F, Zarsky V. Old AIMs of the exocyst: evidence for an ancestral association of exocyst subunits with autophagy-associated Atg8 proteins. Plant signaling & behavior 2013; 8:e27099.
25. Avin-Wittenberg T, Bajdzienko K, Wittenberg G, Alseekh S, Tohge T, Bock R, et al. Global Analysis of the Role of Autophagy in Cellular Metabolism and Energy Homeostasis in Arabidopsis Seedlings under Carbon Starvation. The Plant cell 2015.
26. Masclaux-Daubresse C, Clement G, Anne P, Routaboul JM, Guiboileau A, Soulay F, et al. Stitching together the Multiple Dimensions of Autophagy Using Metabolomics and Transcriptomics Reveals Impacts on Metabolism, Development, and Plant Responses to the Environment in Arabidopsis. The Plant cell 2014; 26:1857-77.
27. Ren C, Liu J, Gong Q. Functions of autophagy in plant carbon and nitrogen metabolism. Frontiers in plant science 2014; 5:301.
28. Sparkes IA, Brandizzi F, Slocombe SP, El-Shami M, Hawes C, Baker A. An Arabidopsis pex10 null mutant is embryo lethal, implicating peroxisomes in an essential role during plant embryogenesis. Plant physiology 2003; 133:1809-19.
29. Contento AL, Xiong Y, Bassham DC. Visualization of autophagy in Arabidopsis using the fluorescent dye monodansylcadaverine and a GFP-AtATG8e fusion protein. The Plant journal : for cell and molecular biology 2005; 42:598-608.
30. Michaeli S, Honig A, Levanony H, Peled-Zehavi H, Galili G. Arabidopsis ATG8-INTERACTING PROTEIN1 Is Involved in Autophagy-Dependent Vesicular Trafficking of Plastid Proteins to the Vacuole. The Plant cell 2014; 26:4084-101.
31. Schluter A, Real-Chicharro A, Gabaldon T, Sanchez-Jimenez F, Pujol A. PeroxisomeDB 2.0: an integrative view of the global peroxisomal metabolome. Nucleic acids research 2010; 38:D800-5.
32. Farmer LM, Rinaldi MA, Young PG, Danan CH, Burkhart SE, Bartel B. Disrupting autophagy restores peroxisome function to an Arabidopsis lon2 mutant and reveals a role for the LON2 protease in peroxisomal matrix protein degradation. The Plant cell 2013; 25:4085-100.
33. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Current protocols in human genetics / editorial board, Jonathan L Haines [et al] 2013; Chapter 7:Unit7 20.
34. Ebberink MS, Mooijer PAW, Gootjes J, Koster J, Wanders RJA, Waterham HR. Genetic Classification and Mutational Spectrum of More Than 600 Patients with a Zellweger Syndrome Spectrum Disorder. Human mutation 2011; 32:59-69.
35. Zolman BK, Bartel B. An Arabidopsis indole-3-butyric acid-response mutant defective in PEROXIN6, an apparent ATPase implicated in peroxisomal function. Proc Natl Acad Sci USA 2004; 101:1786-91.
36. Burkhart SE, Lingard MJ, Bartel B. Genetic dissection of peroxisome-associated matrix protein degradation in Arabidopsis thaliana. Genetics 2013; 193:125-241.
37. Goto S, Mano S, Nakamori C, Nishimura M. Arabidopsis ABERRANT PEROXISOME MORPHOLOGY9 is a peroxin that recruits the PEX1-PEX6 complex to peroxisomes. The Plant cell 2011; 23:1573-87.

38. Hara-Kuge S, Fujiki Y. The peroxin Pex14p is involved in LC3-dependent degradation of mammalian peroxisomes. Experimental cell research 2008; 314:3531-41.
39. Farre JC, Manjithaya R, Mathewson RD, Subramani S. PpAtg30 tags peroxisomes for turnover by selective autophagy. Developmental cell 2008; 14:365-76.
40. Manivannan S, de Boer R, Veenhuis M, van der Klei IJ. Lumenal peroxisomal protein aggregates are removed by concerted fission and autophagy events. Autophagy 2013; 9:1044-56.
41. Sawa-Makarska J, Abert C, Romanov J, Zens B, Ibiricu I, Martens S. Cargo binding to Atg19 unmasks additional Atg8 binding sites to mediate membrane-cargo apposition during selective autophagy. Nature cell biology 2014; 16:425-33.
42. Lin L, Yang P, Huang X, Zhang H, Lu Q, Zhang H. The scaffold protein EPG-7 links cargo-receptor complexes with the autophagic assembly machinery. The Journal of cell biology 2013; 201:113-29.
43. Yan T, Yoo D, Berardini TZ, Mueller LA, Weems DC, Weng S, et al. PatMatch: a program for finding patterns in peptide and nucleotide sequences. Nucleic acids research 2005; 33:W262-6.
44. Tzfira T, Tian GW, Lacroix B, Vyas S, Li J, Leitner-Dagan Y, et al. pSAT vectors: a modular series of plasmids for autofluorescent protein tagging and expression of multiple genes in plants. Plant molecular biology 2005; 57:503-16.
45. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular biology and evolution 2013; 30:2725-9.
46. Alemu EA, Lamark T, Torgersen KM, Birgisdottir AB, Larsen KB, Jain A, et al. ATG8 family proteins act as scaffolds for assembly of the ULK complex: sequence requirements for LC3-interacting region (LIR) motifs. The Journal of biological chemistry 2012; 287:39275-90.
47. Svenning S, Lamark T, Krause K, Johansen T. Plant NBR1 is a selective autophagy substrate and a functional hybrid of the mammalian autophagic adapters NBR1 and p62/SQSTM1. Autophagy 2011; 7:993-1010.
48. Zhou J, Wang J, Cheng Y, Chi YJ, Fan B, Yu JQ, et al. NBR1-mediated selective autophagy targets insoluble ubiquitinated protein aggregates in plant stress responses. PLoS genetics 2013; 9:e1003196.
49. Arabidopsis Interactome Mapping C. Evidence for network evolution in an Arabidopsis interactome map. Science 2011; 333:601-7.
50. Chen J, Lalonde S, Obrdlik P, Noorani Vatani A, Parsa SA, Vilarino C, et al. Uncovering Arabidopsis membrane protein interactome enriched in transporters using mating-based split ubiquitin assays and classification models. Frontiers in plant science 2012; 3:124.
51. Kim DY, Scalf M, Smith LM, Vierstra RD. Advanced proteomic analyses yield a deep catalog of ubiquitylation targets in Arabidopsis. The Plant cell 2013; 25:1523-40.
52. Li F, Chung T, Vierstra RD. AUTOPHAGY-RELATED11 plays a critical role in general autophagy- and senescence-induced mitophagy in Arabidopsis. The Plant cell 2014; 26:788-807.
53. Zhuang X, Wang H, Lam SK, Gao C, Wang X, Cai Y, et al. A BAR-domain protein SH3P2, which binds to phosphatidylinositol 3-phosphate and ATG8, regulates autophagosome formation in Arabidopsis. The Plant cell 2013; 25:4596-615.

54. Yamaguchi M, Matoba K, Sawada R, Fujioka Y, Nakatogawa H, Yamamoto H, et al. Noncanonical recognition and UBL loading of distinct E2s by autophagy-essential Atg7. Nature structural & molecular biology 2012; 19:1250-6.
55. Klopffleisch K, Phan N, Augustin K, Bayne RS, Booker KS, Botella JR, et al. Arabidopsis G-protein interactome reveals connections to cell wall carbohydrates and morphogenesis. Molecular systems biology 2011; 7:532.
56. Honig A, Avin-Wittenberg T, Ufaz S, Galili G. A new type of compartment, defined by plant-specific Atg8-interacting proteins, is induced upon exposure of Arabidopsis plants to carbon starvation. The Plant cell 2012; 24:288-303.

SUPPLEMENTARY DATA

Table S1. The experimentally verified AIMs used for this study.

UNIPROT ID	ACC.	Sequence	Verified	xLIR	hfAIM	PSSM score (e-value)
Alemu et al[46]						
Atg1_YEAST	P53104	REYVVV	Yes	REYVVV	EREYVVV (C)	14 (5.7e-02)
ATG13_HUMAN	O75143	EGFQTV	No	EGFQTV	-	11 (1.5e-01)
		DDFVMI	Yes	DDFVMI	HDDFVMI (A)	20 (8.4e-03)
Atg19_YEAST	P35193	-	No	-	YHDYERL (B)	13 (7.9e-02)
		LTWEEL	Yes	LTWEEL	ALTWEEL (H, I, J)	18 (1.6e-02)
Atg3_YEAST	P40344	-	No	-	EQMFEDI (I)	10 (2.0e-01)
		GDWEDL	Yes	GDWEDL	VGDWEDL (B, D, I)	22 (4.4e-03)
Atg32_YEAST	P40458	KEYQSL	No	KEYQSL	-	12 (1.1e-01)
		LGYILL	No	LGYILL	-	10 (2.0e-01)
		GSWQAI	Yes	GSWQAI	-	17 (2.2e-02)
ATG4B_HUMAN	Q9Y4P1	EDFEIL	No	EDFEIL	DEDFEIL (A-C, E, H)	17 (2.2e-02)
		PMFELV	No	PMFELV	-	10 (2.0e-01)
		LT YDTL	Yes	LT YDTL	-	12 (1.1e-01)
BNI3L_HUMAN	O60238	AEFLKV	No	AEFLKV	-	10 (2.0e-01)
		SSWVEL	Yes	SSWVEL	-	20 (8.4e-03)
CALR_HUMAN	P27797	DDWDFL	No	DDWDFL	EDDWDFL (A-C, E, H)	26 (1.2e-03)
		GGYVKL	No	GGYVKL	-	12 (1.1e-01)
		DEFTHL	Yes	DEFTHL	DDEFTHL (A, C, F)	14 (5.7e-02)
CBL_HUMAN	P22681	LT YDEV	No	LT YDEV	FLTYDEV (I)	11 (1.5e-01)
		DTYQHL	No	DTYQHL	-	14 (5.7e-02)
		REFVSI	No	REFVSI	-	13 (7.9e-02)
		-	No	-	ENEYFRV (C)	6 (7.4e-01)
		FGWLSL	Yes	FGWLSL	-	18 (1.6e-02)
CLH1_HUMAN	Q00610	EDYQAL	No	EDYQAL	EEDYQAL (A, C, F)	16 (3.0e-02)

		GMFTEL	No	GMFTEL	-	11 (1.5e-01)
		-	No	-	SGNWEEL (I)	18 (1.6e-02)
		-	No	-	RGYFEEL (I)	8 (3.9e-01)
		PDWIFL	Yes	PDWIFL	-	22 (4.4e-03)
DVL2_HUMAN	O14641	-	No	-	DQDFGVV (C)	9 (2.8e-01)
		RMWLKI	Yes	RMWLKI	-	18 (1.6e-02)
FUND1_HUMAN	Q8IVP5	GGFLLL	No	GGFLLL	-	10 (2.0e-01)
		DSYEVL	Yes	DSYEVL	-	16 (3.0e-02)
FYCO 1_HUMAN	Q9BQS8	ADYQAL	No	ADYQAL	-	15 (4.2e-02)
		AVFDII	Yes	AVFDII	-	8 (3.9e-01)
NBR1_HUMAN	Q14596	LSFELL	No	LSFELL	-	10 (2.0e-01)
		EDYIII	Yes	EDYIII	SEDYIII (A)	17 (2.2e-02)
OPTN_HUMAN	Q96CV9	DSFVEI	Yes	DSFVEI	EDSFVEI (D, F, G)	15 (4.2e-02)
Q8MQJ7_DRO ME	Q8MQJ7	ADYLSV	No	ADYLSV	-	14 (5.7e-02)
		DDFVLV	Yes	DDFVLV	SDDFVLV (A)	17 (2.2e-02)
Q9SB64_ARATH	Q9SB64	RVWVLI	No	RVWVLI		15 (4.2e-02)
		SEWDPI	Yes	SEWDPI	VSEWDPI (B)	20 (8.4e-03)
RBCC 1_HUMAN	Q8TDY2	-	No	-	LCDFEPL (B)	12 (1.1e-01)
		-	No	-	IQEFEKV (B)	10 (2.0e-01)
		FDFETI	Yes	FDFETI	TFDFETI (B)	17 (2.2e-02)
SQSTM_HUMAN	Q13501	DDWTHL	Yes	DDWTHL	DDDWTHL (A, C, F)	24 (2.3e-03)
STBD1_HUMAN	O95210	EEWEMV	Yes	EEWEMV	HEEWEMV (A, B, E)	21 (6.1e-03)
T53I1_HUMAN	Q96A56	DEWILV	Yes	DEWILV	DDEWILV (A, C, F)	20 (8.4e-03)
T53I2_HUMAN	Q8IXH6	DGWLII	Yes	DGWLII	-	21 (6.1e-03)
TBC25_HUMAN	Q3MII6	EVYLSL	No	EVYLSL	-	8 (3.9e-01)
		-	No	-	SREYEQL (B)	13 (7.9e-02)
		EDWDII	Yes	EDWDII	LEDWDII (A, B, E)	24 (2.3e-03)
TBCD5_HUMAN	Q92609	DDFILI	No	DDFILI	SDDFILI (A)	17 (2.2e-02)
		SGFTIV	Yes	SGFTIV	-	11 (1.5e-01)
		KEWEEL	Yes	KEWEEL	RKEWEEL (B, D, I)	20 (8.4e-03)

ULK1_HUMAN	O75385	DDFVMV	Yes	DDFVMV	TDDFVMV (A)	19 (1.2e-02)
ULK2_HUMAN	Q8IYT8	DDFVLV	Yes	DDFVLV	TDDFVLV (A)	17 (2.2e-02)
Birgisdottir et al[23]						
ATG34_YEAST	Q12292	KVYEKL	No	KVYEKL	-	8 (3.9e-01)
		FTWEEI	Yes	FTWEEI	PFTWEEI (I)	20 (8.4e-03)
BNIP3_HUMAN	Q12983	AEFLKV	No	AEFLKV	-	10 (2.0e-01)
		GSWVEL	Yes	GSWVEL	-	19 (1.2e-02)
C0H519_PLAF7	C0H519	NDWLLP	Yes	-	-	12 (1.2e-02)
		-	No	-	EEIFEDI (E-J)	8 (3.9e-01)
CACO 2_HUMAN	Q13137	DILVV	Yes	-	-	N/A
		FMWVTL	No	FMWVTL	-	20 (8.4e-03)
CTNB1_HUMAN	P35222	SHWPLI	Yes	-	-	11 (1.5e-01)
MK15_HUMAN	Q8TD08	RVYQMI	Yes	RVYQMI	-	10 (2.0e-01)
TAXB1_HUMAN	Q86VP1	DMLVV	Yes	-	-	N/A
		ADFDIV	No	ADFDIV	EADFDIV (B, C, H)	15 (4.2e-02)
Svenning et al[47]						
C0Z2C5_ARATH	C0Z2C5	EREYVLV	Yes	REYVLV	EREYVLV (C)	13 (7.9e-02)
TSPO_ARATH	O82245	NALYLYL	Yes	-	-	6 (7.4e-01)
		-	No	LVWDPV	-	12 (1.1e-01)

The publication from which the data was obtained is indicated above the corresponding entries. ACC, UniProt accession code; Verified, experimentally verified functional AIM motifs; xLIR, iLIR predicted motifs; hfAIM, hfAIM predicted motifs. The letters in brackets following the hfAIM predicted sequences correspond to the regular expression patterns shown in Figure S1.

Table S2. Experimentally identified Atg8-interacting proteins in *Arabidopsis* and their predicted AIM motifs**.**

ATG8 isoforms	Interacting protein	Experimental Evidence	Ref	WxxL	xLIR	hfAIM[1]	PSSM score (e-value)
ATG8A, ATG8B, ATG8C, ATG8D, ATG8F, ATG8G	At4g24690	PCA; Reconstituted complex	47, 48	VSYGGV			2 (2.6e+00)
				LKFLKI			7 (5.3e-01)
				KVYMDL			5 (1.0e+00)
				DYWTSL			16 (3.0e-02)
				APFTKI			7 (5.3e-01)
				PIYSEL			6 (7.4e-01)
				NEFHGL			5 (1.0e+00)
				SEFLGI			10 (2.0e-01)
				SSFNMV			9 (2.8e-01)
				MGFKEI			7 (5.3e-01)
					RVWVLI		15 (4.2e-02)
					SEWDPI	VSEWDPI (5)	**20 (8.4e-03)**
ATG8B	At1g13030	Two-hybrid	49	NVYNAI			3 (1.9e+00)
				IRFKPL			3 (1.9e+00)
				IDYEQL		PIDYEQL (5)	13 (7.9e-02)
				IAYRLI			3 (1.9e+00)
				IEFSAL		EIEFSAL (4)	7 (5.3e-01)
				SPWEEL		VSPWEEL (3)	**17 (2.2e-02)**
				WSYKAL			6 (7.4e-01)
					RSWVVL		20 (8.4e-03)
	At2g20570	Two-hybrid		CGFTII			8 (3.9e-01)
				IDFDDI		IIDFDDI (3, 5)	13 (7.9e-02)
				VFFLKV			2 (2.6e+00)
				HHFRPL			2 (2.6e+00)
				RPWLPL			15 (4.2e-02)
ATG8B, ATG8F	At5g40690	Two-hybrid		RTYRIL			10 (2.0e-01)
				GLFTKV			5 (1.0e+00)
				VTYKML			9 (2.8e-01)
				MSFCDV			3 (1.9e+00)
				EGWYLV		DEGWYLV (2)	**14 (5.7e-02)**
ATG8B, ATG8D	At5g65480	Two-hybrid		VSFPRL			3 (1.9e+00)
				PLYKKV			4 (1.4e+00)
					GGFVVL		12 (1.1e-01)
ATG8D	At5g24930	Two-hybrid		AAFLCL			5 (1.0e+00)
				PFYDSV			6 (7.4e-01)
				CYFSDI			1 (3.6e+00)
				DPYLDL			11 (1.5e-01)
				GGYNCI			6 (7.4e-01)

				SGFGLV			6 (7.4e-01)
					ASWLLL		17 (2.2e-02)
						DLDYGNV (4)	12 (1.1e-01)
	At1g19450	PCA		SVFGSL			3 (1.9e+00)
				IGWLSI			15 (4.2e-02)
				VPWRIL			10 (2.0e-01)
			50	VRFVDL			6 (7.4e-01)
				RYYFPL			5 (1.0e+00)
				NMYNIL			8 (3.9e-01)
				TLYALV			2 (2.6e+00)
				CGFTVV			7 (5.3e-01)
	At1g29800	Two-hybrid		MGWDQI			15 (4.2e-02)
				TMWDVI			18 (1.6e-02)
				VRFHPI			2 (2.6e+00)
			49	VCFVRL			4 (1.4e+00)
				IDFIIV			12 (1.1e-01)
				ALYRAL			4 (1.4e+00)
					GVWIPV		12 (1.1e-01)
	At4g23030	PCA		LGFANI			4 (1.4e+00)
				TGYSLL			8 (3.9e-01)
				KRFKLL			5 (1.0e+00)
				ILWLNI			9 (2.8e-01)
				SAFFAV			4 (1.4e+00)
				INYLLV			7 (5.3e-01)
				AIWTNV			10 (2.0e-01)
			50	IVFSGV			-1 (6.9e+00)
				KGWRSL			14 (5.7e-02)
				WWWYEI			7 (5.3e-01)
				AMFFAL			7 (5.3e-01)
				NCWARL			8 (3.9e-01)
				CCFYFV			0 (5.0e+00)
				FDFKGL			11 (1.5e-01)
					LGFLII		10 (2.0e-01)
	At4g39050	Two-hybrid		HEYNPL			7 (5.3e-01)
				YAFDKV			5 (1.0e+00)
				TPFSLV			3 (1.9e+00)
			49	DLFFNL			4 (1.4e+00)
				PLWYLL			10 (2.0e-01)
				DVFSII			8 (3.9e-01)
				VSYLEI			10 (2.0e-01)
				NNFNLL			5 (1.0e+00)

				DEYDGV		GDEYDGV (1, 5)	13 (7.9e-02)
				DKFDSL			12 (1.1e-01)
				NMWVLV			16 (3.0e-02)
				FFYQRV			5 (1.0e+00)
	At5g06780	Two-hybrid		QSYPSI			6 (7.4e-01)
				SFYEAV			7 (5.3e-01)
				ETWEWV			15 (4.2e-02)
				KFFGEI			2 (2.6e+00)
	At1g01340	PCA		RVFKNV			2 (2.6e+00)
				QNWNKI			12 (1.1e-01)
				FLFACV			1 (3.6e+00)
				PLFFYI			2 (2.6e+00)
				DAFYII			8 (3.9e-01)
				SSYFII			10 (2.0e-01)
			50	LKFSII			6 (7.4e-01)
				PLYTEV			7 (5.3e-01)
				HVFGAL			0 (5.0e+00)
				CFWWGL			3 (1.9e+00)
				VLFALL			1 (3.6e+00)
				GFFNAV			1 (3.6e+00)
				SQFRRL			5 (1.0e+00)
				FRFYSV			4 (1.4e+00)
	At5g16150	PCA		AGYKSL			9 (2.8e-01)
				MSFSSV			5 (1.0e+00)
				LPFVGV			4 (1.4e+00)
				QLFICI			3 (1.9e+00)
				AGWFDL			14 (5.7e-02)
				SRYWKV			4 (1.4e+00)
				FLFQQL			6 (7.4e-01)
				FTWKAL			16 (3.0e-02)
				LYFLSV			4 (1.4e+00)
			50	LGFAGV			3 (1.9e+00)
	At2g02860	PCA		VPYRNL			1 (3.6e+00)
				SSFIWL			8 (3.9e-01)
				QPFVGI			5 (1.0e+00)
				RPFILV			7 (5.3e-01)
				IGFWLL			4 (1.4e+00)
				CLWMAI			8 (3.9e-01)
				EWFPFL			3 (1.9e+00)
				VVFLTI			5 (1.0e+00)
				IKYERV			6 (7.4e-01)

				IVFALL			1 (3.6e+00)
				GPWDQL			14 (5.7e-02)
ATG8D, ATG8F	At2g35900	Two-hybrid	49, 51	MGWIWI			12 (1.1e-01)
				LNFPGL			2 (2.6e+00)
	At2g46550	Two-hybrid		AGFMPL			8 (3.9e-01)
				PNFPDL			5 (1.0e+00)
				QEFIEL			11 (1.5e-01)
				SPWNLL			12 (1.1e-01)
				KQWLQL			14 (5.7e-02)
				ALYLQI			6 (7.4e-01)
	At5g45900	Two-hybrid		LQFAPL			4 (1.4e+00)
				HSFSSL			6 (7.4e-01)
				ESFNKL			8 (3.9e-01)
				KIWEDI		NKIWEDI (3)	**14 (5.7e-02)**
				PRFLVI			8 (3.9e-01)
				VPFFLV			1 (3.6e+00)
				RNYLAL			11 (1.5e-01)
				RGFADL			8 (3.9e-01)
				MRWRAL			9 (2.8e-01)
				VDYGKV			8 (3.9e-01)
				GEFKAV			8 (3.9e-01)
				SQFSQI			6 (7.4e-01)
ATG8E	At4g30790	PCA; Two-hybrid	52	EQYYRV			5 (1.0e+00)
				QMFVEV			10 (2.0e-01)
				ACYNSI			4 (1.4e+00)
				DLFADL		DDLFADL (2)	6 (7.4e-01)
				DNFDDI		KDNFDDI (3)	13 (7.9e-02)
				SQYTAL			10 (2.0e-01)
				ISFGRL			3 (1.9e+00)
				GHYEAI			9 (2.8e-01)
				CEYFIV			6 (7.4e-01)
					FEYEVL	QFEYEVL (5)	**15 (4.2e-02)**
ATG8E, ATG8F	At4g34660	Two-hybrid	53	FPYHGV			3 (1.9e+00)
				TLWFFV			9 (2.8e-01)
					GEYVVV		13 (7.9e-02)
ATG8F	At5g03240	Affinity capture-MS	51	-			N/A
	At5g07730	Two-hybrid	49	DWWEPL			16 (3.0e-02)
				RLYSKL			6 (7.4e-01)
				RSFYPV			7 (5.3e-01)
				WWWMRV			5 (1.0e+00)

				VVFPPI			0 (5.0e+00)
				SPFEQI			8 (3.9e-01)
	At5g61500	Reconstituted complex	54	DKYLFL			13 (7.9e-02)
	At2g03670	Two-hybrid	55	GRFDAL			8 (3.9e-01)
				VTWDDV		KVTWDDV (3)	**15 (4.2e-02)**
				ASFFSL			7 (5.3e-01)
				GRFDLV			7 (5.3e-01)
				ARFEIL			9 (2.8e-01)
	At4g15930	Two-hybrid	49	ATWHCI			12 (1.1e-01)
				FVYFYL			4 (1.4e+00)
	At2g45980	PCA	56	NEWEVV		GNEWEVV (5)	**18 (1.6e-02)**
				ERWQVL			16 (3.0e-02)
	At4g00355	PCA		NDWEVV		GNDWEVV (5)	**21 (6.1e-03)**
				DDFGGL		SDDFGGL (1)	11 (1.5e-01)
				HVYEPV			5 (1.0e+00)
				ERWQIL			16 (3.0e-02)
				PPWYDV			10 (2.0e-01)
ATG8G	At2g02040	PCA	50	LAYYGI			2 (2.6e+00)
				GRYWTI			5 (1.0e+00)
				ACFSGI			2 (2.6e+00)
				MFFGGL			-3 (1.3e+01)
				WFYFSI			0 (5.0e+00)
				TVFMGL			1 (3.6e+00)
				IIFSAV			1 (3.6e+00)
				IIWVPL			12 (1.1e-01)
				RKFTGV			6 (7.4e-01)
				GLFVSV			6 (7.4e-01)
				PQYFIL			7 (5.3e-01)
				EVFYFI			5 (1.0e+00)
				DYFFWL			3 (1.9e+00)
					KGFTEI		11 (1.5e-01)

The different ATG8 isoforms interacting proteins were derived from the UNIPROT database. PCA, a protein-fragment complementation assay, e.g. the split-ubiquitin assay. Reconstituted complex, an interaction is detected between purified proteins *in vitro*. Two-hybrid, yeast two-hybrid assay. Affinity capture-western, an interaction is inferred when a bait protein is affinity captured from cell extracts by either polyclonal antibody or epitope tag and the associated interaction partner identified by western blot with a specific polyclonal antibody or second epitope tag.

[1] The numbers in brackets correspond to the regular expression pattern shown in Figure 2A.

Table S3. Gene Ontology enrichment in hfAIM predicted AIM-containing proteins (ACPs).

No. of ACPs[1]	GO term	# of genes	raw p-value	corrected p-value	Frequency in set (%)
11 (6)	regulation of transcription, DNA-dependent - GO:0006355	5	5.50E-04	0.02	63
24 (5)	M phase - GO:0000279	4	1.91E-04	0.024	20
	adenyl ribonucleotide binding - GO:0032559	8	2.64E-04	0.032	40
101 (4)	adenyl ribonucleotide binding - GO:0032559	28	4.67E-09	0.001	32
	nucleoside-triphosphatase activity - GO:0017111	15	3.85E-07	0.001	17
	cellular catabolic process - GO:0044248	21	2.21E-05	0.011	24
	organic substance metabolic process - GO:0071704	18	3.77E-05	0.013	20
	gluconeogenesis - GO:0006094	6	5.84E-05	0.02	6.8
	carbohydrate biosynthetic process - GO:0016051	12	1.28E-04	0.047	14
	proteolysis involved in cellular protein catabolic process - GO:0051603	9	1.28E-04	0.047	10
449 (3)	adenyl ribonucleotide binding - GO:0032559	109	3.47E-25	0.001	28
	nucleoside-triphosphatase activity - GO:0017111	62	3.58E-23	0.001	16
	ADP binding - GO:0043531	24	2.92E-16	0.001	6.1
	organophosphate metabolic process - GO:0019637	67	1.98E-13	0.001	17

	carbohydrate derivative metabolic process - GO:1901135	68	2.34E-13	0.001	17
	hydrolase activity - GO:0016787	103	6.96E-13	0.001	26
	defense response - GO:0006952	65	1.92E-12	0.001	16
	nucleobase-containing compound metabolic process - GO:0006139	100	7.72E-10	0.001	25
	regulation of organelle organization - GO:0033043	21	1.02E-09	0.001	5.3
	helicase activity - GO:0004386	17	3.11E-09	0.001	4.3
	cellular catabolic process - GO:0044248	70	2.50E-08	0.001	18
	reproductive process - GO:0022414	73	4.74E-08	0.001	18
	organelle organization - GO:0006996	69	1.41E-07	0.001	17
	system development - GO:0048731	80	1.65E-07	0.001	20
	ligase activity, forming carbon-sulfur bonds - GO:0016877	8	1.69E-07	0.001	2
	tissue development - GO:0009888	41	2.45E-07	0.001	10
	regulation of biological quality - GO:0065008	45	2.64E-07	0.001	11
	sister chromatid cohesion - GO:0007062	14	2.74E-07	0.001	3.5
	chromosome organization - GO:0051276	34	4.91E-07	0.001	8.6
	trichome morphogenesis - GO:0010090	14	5.45E-07	0.001	3.5
	actin filament-based process - GO:0030029	18	8.21E-07	0.001	4.6

	meiotic cell cycle - GO:0051321	17	2.20E-06	0.002	4.3
	vegetative to reproductive phase transition of meristem - GO:0010228	23	3.63E-06	0.002	5.8
	cellular response to stimulus - GO:0051716	85	3.80E-06	0.003	22
	negative regulation of biological process - GO:0048519	43	4.86E-06	0.003	11
	cell cycle - GO:0007049	32	4.97E-06	0.003	8.1
	flavin adenine dinucleotide binding - GO:0050660	13	5.04E-06	0.003	3.3
	cell communication - GO:0007154	70	5.16E-06	0.003	18
	M phase - GO:0000279	19	6.79E-06	0.005	4.8
	UDP-N-acetylmuramate dehydrogenase activity - GO:0008762	7	8.47E-06	0.007	1.8
	cytoskeleton organization - GO:0007010	24	1.04E-05	0.009	6.1
	gene silencing by RNA - GO:0031047	18	1.12E-05	0.009	4.6
	regulation of cell cycle process - GO:0010564	12	1.59E-05	0.013	3
	meristem structural organization - GO:0009933	15	2.51E-05	0.022	3.8
	macromolecule modification - GO:0043412	80	3.51E-05	0.034	20
	cellular localization - GO:0051641	50	4.10E-05	0.038	13
	anatomical structure morphogenesis - GO:0009653	51	5.38E-05	0.05	13

	hydrogen peroxide biosynthetic process - GO:0050665	8	5.56E-05	0.05	2
1620 (2)	adenyl ribonucleotide binding - GO:0032559	247	3.52E-24	0.001	18
	purine nucleotide binding - GO:0017076	267	8.31E-24	0.001	20
	hydrolase activity - GO:0016787	265	1.04E-12	0.001	19
	transferase activity - GO:0016740	260	1.06E-12	0.001	19
	hydrolase activity, acting on acid anhydrides - GO:0016817	105	2.06E-12	0.001	7.7
	ADP binding - GO:0043531	32	4.95E-10	0.001	2.3
	developmental process involved in reproduction - GO:0003006	175	1.35E-08	0.001	13
	organic substance metabolic process - GO:0071704	153	1.49E-08	0.001	11
	cofactor binding - GO:0048037	59	4.17E-07	0.002	4.3
	xyloglucan:xyloglucosyl transferase activity - GO:0016762	12	5.23E-07	0.002	0.88
	UDP-N-acetylmuramate dehydrogenase activity - GO:0008762	13	7.66E-07	0.003	0.95
	organelle organization - GO:0006996	178	8.23E-07	0.003	13
	protein modification process - GO:0036211	213	1.04E-06	0.004	16
	post-embryonic development - GO:0009791	168	1.48E-06	0.004	12
	chromosome organization - GO:0051276	78	2.12E-06	0.005	5.7

	anatomical structure development - GO:0048856	248	3.20E-06	0.006	18
	cellular response to stimulus - GO:0051716	239	3.84E-06	0.006	17
	regulation of cellular component organization - GO:0051128	42	3.86E-06	0.006	3.1
	embryo development ending in seed dormancy - GO:0009793	59	4.21E-06	0.006	4.3
	cellular catabolic process - GO:0044248	170	4.37E-06	0.006	12
	catabolic process - GO:0009056	205	5.60E-06	0.008	15
	defense response - GO:0006952	127	6.63E-06	0.009	9.3
	xyloglucan endotransglucosylase activity - GO:0080039	5	6.80E-06	0.009	0.37
	glycine hydroxymethyltransferase activity - GO:0004372	5	6.80E-06	0.009	0.37
	ATPase activity, coupled to movement of substances - GO:0043492	26	8.32E-06	0.01	1.9

[1]No. of ACPs with the no. of hfAIM motifs in each protein given in brackets.

Table S4. Conservation of AIMs in AtPEX proteins predicted by hfAIM.

Protein	hfAIM motif (pattern)[1]	Conservation (%)[2]
AtPEX1	EDDWEVL (1, 2, 4, 5)	29/38 (76%)
	RLGWEDV (3)	2/38 (5%)
	FEDFDSI (1, 5)	34/38 (89%)
AtPEX3-2	LQLWDEL (3)	18/32 (56%)
AtPEX5	GAAWDEV (3)	1/46 (2%)
AtPEX6	VIFFDEL (3)	40/43 (93%)
AtPEX7	AHDFEIL (5)	1/31 (3%)
	STGWDEL (3)	2/31 (6%)
AtPEX10	GEEYCDI (1)	35/38 (92%)
AtPEX14	RKYFEDL (3)	1/37 (3%)
AtPEX17	SGEWETL (5)	1/7 (14%)
AtPEX19-2	LKQFEDL (3)	1/35 (3%)

[1]The regular expression patterns correspond to the hfAIM patterns indicated in Figure 2A. [2]Conservation is calculated as the percentage of organisms with the conserved hfAIM out of the total number of analyzed organisms.

Table S5. Primers used in this study.

Name	Primers (5'-3')	Annotation
YC_PEX6 _F	ggactcagatctcgaATGGTGGAGAGACGGAATC	For YC-PEX6
YC_PEX6 _R	gatcccgggcccgcgGCTCGAACGGCCTTGA	For YC-PEX6
PEX10_YC_F	ggactcagatctcgaATGAGGCTTAATGGGGATTC	For PEX10-YC
PEX6_YC_R	gatcccgggcccgcgAAAATCAGAATGATACAAACA	For PEX10-YC
YN_Atg8_F	ggactcagatctcgaATGGCAAAAAGCTCGTTCA	For YN-Atg8
YN_Atg8_R	gatcccgggcccgcggtTTATGGAGATCCAAATCC	For YN-Atg8
PEX10_G93E_YC_F	CAAACTTTAGaAGAGGAATAT	For PEX10-G93E-YC
PEX10_G93E_YC_F	ATATTCCTCTtCTAAAGTTTG	For PEX10-G93E-YC
PEX10_Y96A_YC_F	GGAGAGGAAgcTTGTGACAT	For PEX10-Y96A-YC
PEX10_Y96A_YC_R	ATGTCACAAgcTTCCTCTCC	For PEX10-Y96A-YC
PEX10_F253A_YC_F	CTTGGGGTTgcCCTTCTAATC	For PEX10-F253A-YC
PEX10_ F253A _YC_R	GATTAGAAGGgcAACCCCAAG	For PEX10-F253A-YC

Figure S1. The ten regular patterns that define the hfAIM motif.

Figure S2. Additional controls for PEX10 and Atg8 interaction assays.

Figure S3. Multiple sequence alignment of the predicted AIM containing regions of AtPEX10.

Figure S4. Multiple sequence alignment of the predicted AIM containing regions of AtPEX1.

Figure S5. Multiple sequence alignment of the predicted AIM containing regions of AtPEX6.

Figure S6. Multiple sequence alignment of the predicted AIM containing regions of AtPEX3-2.

Figure S7. Multiple sequence alignment of the predicted AIM containing regions of AtPEX5.

Figure S8. Multiple sequence alignment of the predicted AIM containing regions of AtPEX7.

Figure S9. Multiple sequence alignment of the predicted AIM containing regions of AtPEX14.

Figure S10. Multiple sequence alignment of the predicted AIM containing regions of AtPEX17.

Figure S11. Multiple sequence alignment of the predicted AIM containing regions of AtPEX19-2.

A	X[DE][DE][WFY]XX[LIV]
B	XX[DE][WFY][DE]X[LIV]
C	[DE]X[DE][WFY]XX[LIV]
D	XX[DE][WFY]X[DE][LIV]
E	X[DE]X[WFY][DE]X[LIV]
F	[DE][DE]X[WFY]XX[LIV]
G	X[DE]X[WFY]X[DE][LIV]
H	[DE]XX[WFY][DE]X[LIV]
I	XXX[WFY][DE][DE][LIV]
J	[DE]XX[WFY]X[DE][LIV]

Figure S1. The ten regular patterns that define the hfAIM motif. Each pattern contains at least two acidic amino acids. X represents any amino acid.

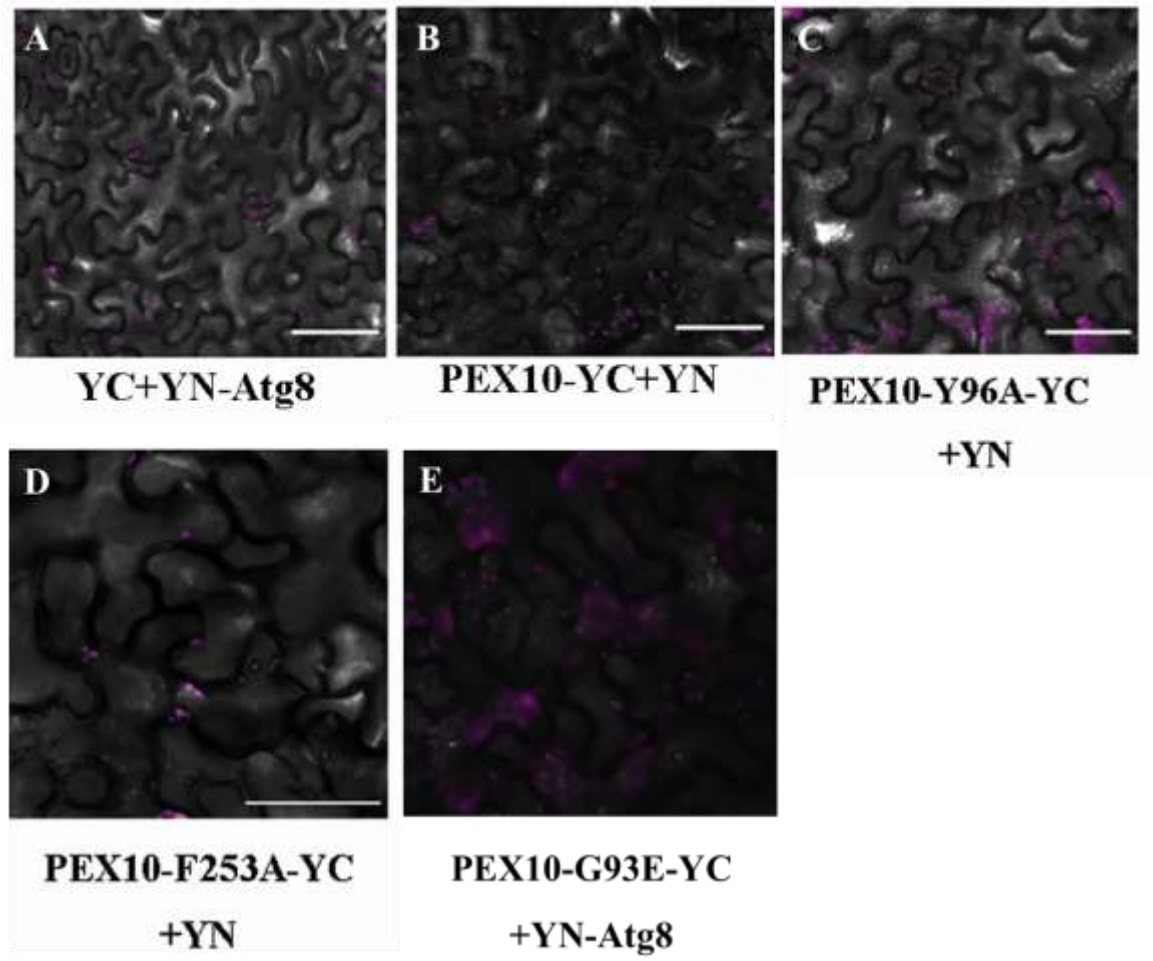

Figure S2. Additional controls for PEX10 and Atg8 interaction assays. No YFP fluorescence was observed following transient coexpression of the unfused YC with YN-Atg8 **(A)**, the unfused YN with PEX10-YC (**B**), PEX10^{Y96A}-YC with YN (**C**), PEX10^{F253A}-YC with YN (**D**) or PEX1^{G93E}-YC with YN-Atg8 (**E**). Bar: 20 µm.

Figure S3. Multiple sequence alignment of the predicted AIM containing regions of AtPEX10. The AtPEX10 regions containing either hfAIM or iLIR predicted AIMs were aligned with the corresponding regions from PEX10 proteins from various species. The location of the AIMs identified in AtPEX10 by either hfAIM or iLIR is indicated on the top. The PEX10 protein sequences derived from the PeroxisomeDB (http://www.peroxisomedb.org/home.jsp) were aligned using ClustalW.

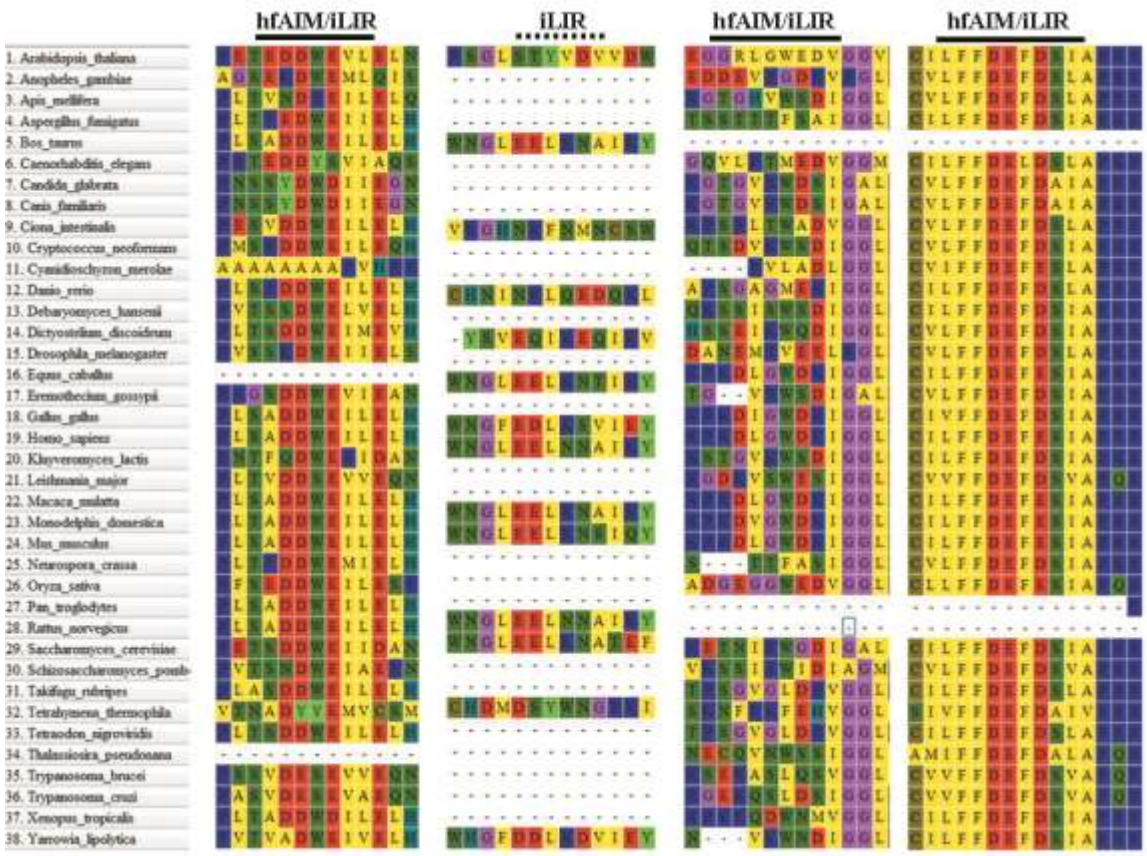

Figure S4. Multiple sequence alignment of the predicted AIM containing regions of AtPEX1. The AtPEX1 regions containing either hfAIM or iLIR predicted AIMs were aligned with the corresponding regions from PEX1 proteins from various species. The location of the AIMs identified in AtPEX1 by either hfAIM or iLIR is indicated on the top. The PEX1 protein sequences derived from the PeroxisomeDB (http://www.peroxisomedb.org/home.jsp) were aligned using ClustalW.

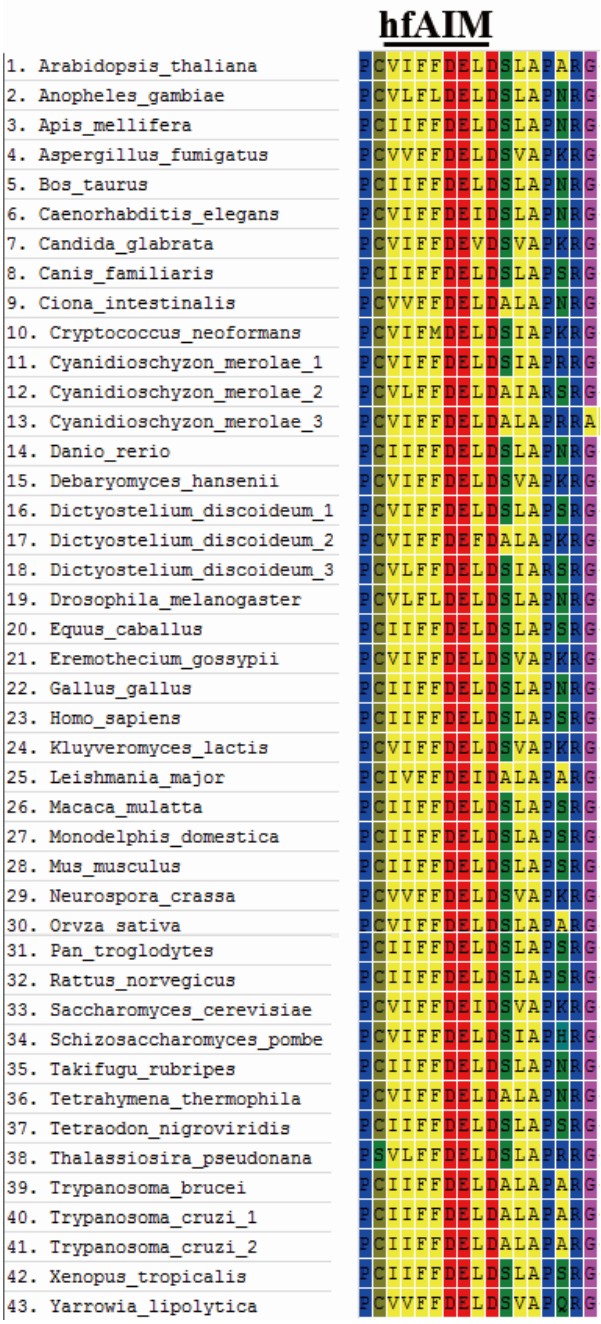

Figure S5. Multiple sequence alignment of the predicted AIM containing regions of AtPEX6. The AtPEX6 regions containing either hfAIM or iLIR predicted AIMs were aligned with the corresponding regions from PEX6 proteins from various species. The location of the AIMs identified in AtPEX6 by either hfAIM or iLIR is indicated on the top. The PEX6 protein sequences derived from the PeroxisomeDB (http://www.peroxisomedb.org/home.jsp) were aligned using ClustalW.

Figure S6. Multiple sequence alignment of the predicted AIM containing regions of AtPEX3-2. The AtPEX3-2 regions containing either hfAIM or iLIR predicted AIMs were aligned with the corresponding regions from PEX3-2 proteins from various species. The location of the AIMs identified in AtPEX3-2 by either hfAIM or iLIR is indicated on the top. The PEX3-2 protein sequences derived from the PeroxisomeDB (http://www.peroxisomedb.org/home.jsp) were aligned using ClustalW.

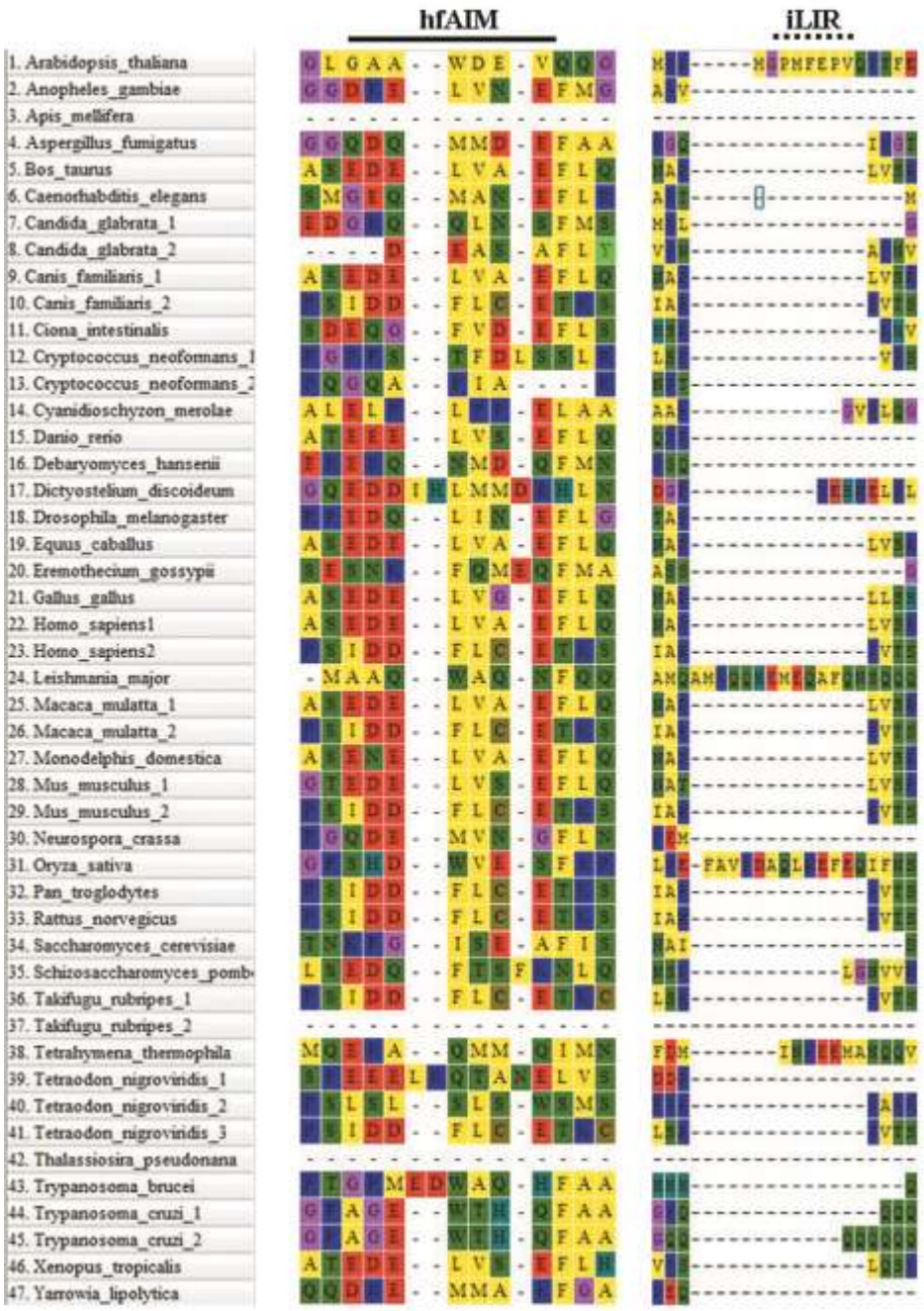

Figure S7. Multiple sequence alignment of the predicted AIM containing regions of AtPEX5. The AtPEX5 regions containing either hfAIM or iLIR predicted AIMs were aligned with the corresponding regions from PEX5 proteins from various species. The location of the AIMs identified in AtPEX5 by either hfAIM or iLIR is indicated on the top. The PEX5 protein sequences derived from the PeroxisomeDB (http://www.peroxisomedb.org/home.jsp) were aligned using ClustalW.

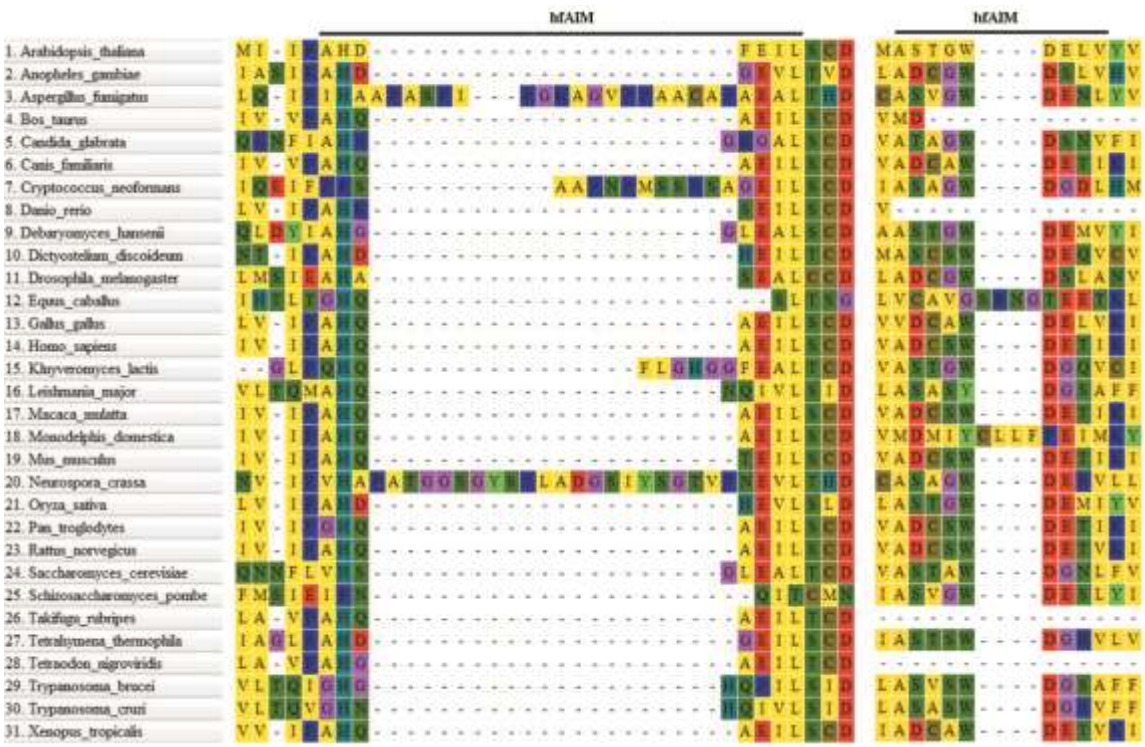

Figure S8. Multiple sequence alignment of the predicted AIM containing regions of AtPEX7. The AtPEX7 regions containing either hfAIM or iLIR predicted AIMs were aligned with the corresponding regions from PEX7 proteins from various species. The location of the AIMs identified in AtPEX7 by either hfAIM or iLIR is indicated on the top. The PEX7 protein sequences derived from the PeroxisomeDB (http://www.peroxisomedb.org/home.jsp) were aligned using ClustalW

Figure S9. Multiple sequence alignment of the predicted AIM containing regions of AtPEX14. The AtPEX14 regions containing either hfAIM or iLIR predicted AIMs were aligned with the corresponding regions from PEX14 proteins from various species. The location of the AIMs identified in AtPEX14 by either hfAIM or iLIR is indicated on the top. The PEX14 protein sequences derived from the PeroxisomeDB (http://www.peroxisomedb.org/home.jsp) were aligned using ClustalW

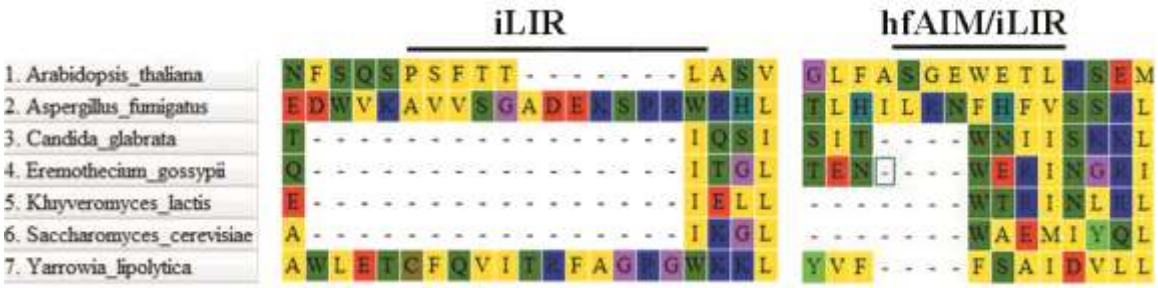

Figure S10. Multiple sequence alignment of the predicted AIM containing regions of AtPEX14. The AtPEX17 regions containing either hfAIM or iLIR predicted AIMs were aligned with the corresponding regions from PEX17 proteins from various species. The location of the AIMs identified in AtPEX17 by either hfAIM or iLIR is indicated on the top. The PEX17 protein sequences derived from the PeroxisomeDB (http://www.peroxisomedb.org/home.jsp) were aligned using ClustalW.

Figure S11. Multiple sequence alignment of the predicted AIM containing regions of AtPEX19-2. The AtPEX19-2 regions containing either hfAIM or iLIR predicted AIMs were aligned with the corresponding regions from PEX14 proteins from various species. The location of the AIMs identified in AtPEX19-2 by either hfAIM or iLIR is indicated on the top. The PEX19-2 protein sequences derived from the PeroxisomeDB (http://www.peroxisomedb.org/home.jsp) were aligned using ClustalW.