THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 52, pp. 45048–45062, December 30, 2011

Structural Requirements for Interaction of Peroxisomal Targeting Signal 2 and Its Receptor PEX7*[S]

Received for publication, September 15, 2011, and in revised form, November 3, 2011 Published, JBC Papers in Press, November 5, 2011, DOI 10.1074/jbc.M111.301853

Markus Kunze[‡1], Georg Neuberger[§2], Sebastian Maurer-Stroh[§¶2], Jianmin Ma[§], Thomas Eck[‡], Nancy Braverman[‖], Johannes A. Schmid[**], Frank Eisenhaber[§‡‡§§2], and Johannes Berger[‡]

From the [‡]Center for Brain Research, Medical University of Vienna, Spitalgasse 4, 1090 Vienna, Austria, the [§]Bioinformatics Institute, Agency for Science, Technology and Research, 30 Biopolis Street, Singapore 138671, the [‡‡]Department of Biological Sciences, National University of Singapore, 8 Medical Drive, Singapore 117597, the [§§]School of Computer Engineering, Nanyang Technological University, 50 Nanyang Drive, Singapore 637553, the [‖]Departments of Human Genetics and Pediatrics, McGill University-Montreal Children's Hospital Research Institute, Montreal, Quebec H3Z 2Z3, Canada, the [**]Department of Vascular Biology and Thrombosis Research, Medical University of Vienna, Schwarzspanierstrasse 17, 1090 Vienna, Austria, and the [¶]School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551

Background: Type 2 peroxisomal targeting signals (PTS2) tag proteins for import into peroxisomes.
Results: Characterization of structural properties of PTS2 allows the prediction of novel PTS2 and identification of the binding site on the receptor PEX7.
Conclusion: PTS2 forms helical structures that bind to a groove on PEX7.
Significance: Understanding the recognition of PTS2 by its receptor is a critical step in peroxisomal protein transport.

The import of a subset of peroxisomal matrix proteins is mediated by the peroxisomal targeting signal 2 (PTS2). The results of our sequence and physical property analysis of known PTS2 signals and of a mutational study of the least characterized amino acids of a canonical PTS2 motif indicate that PTS2 forms an amphipathic helix accumulating all conserved residues on one side. Three-dimensional structural modeling of the PTS2 receptor PEX7 reveals a groove with an evolutionarily conserved charge distribution complementary to PTS2 signals. Mammalian two-hybrid assays and cross-complementation of a mutation in PTS2 by a compensatory mutation in PEX7 confirm the interaction site. An unstructured linker region separates the PTS2 signal from the core protein. This additional information on PTS2 signals was used to generate a PTS2 prediction algorithm that enabled us to identify novel PTS2 signals within human proteins and to describe KChIP4 as a novel peroxisomal protein.

Peroxisomes are single membrane-bound organelles, which are found in all nucleated cells. They host a variety of metabolic functions such as detoxification of hydrogen peroxide (H_2O_2), the degradation of very long and branched chain fatty acids or D-amino acids, and the synthesis of plasmalogens, docosahexaenoic acid, or bile acids (1).

Soluble peroxisomal proteins contain cis-acting peroxisomal targeting signals that mediate their recognition and import into peroxisomes. These signals reside either at the extreme C terminus (PTS1) (2, 3) or in proximity to the N terminus (PTS2) (4, 5). PTS1 is recognized by its receptor, the peroxin 5 (PEX5) (6, 7), and similarly, PTS2 is specifically bound by PEX7 (8, 9). These soluble receptors mediate the transport of their cargo proteins to the peroxisomal surface. There they bind to a multimeric protein complex (docking complex) initiating the transfer of the proteins across the peroxisomal membrane (10). In contrast to many other transport mechanisms, the import machinery of peroxisomes can transport fully folded and even oligomerized proteins across the membrane (11, 12).

Most of the proven peroxisomal matrix proteins of yeast and mammals harbor a PTS1, but in *Arabidopsis thaliana* 30% of the known peroxisomal proteins are transported via the PTS2 pathway (13). The PTS2 motif was originally inferred from the analysis of the first 40 amino acids of yeast (4) and rat thiolase (5). More detailed studies on the thiolase PTS2 of yeast (14), rat (15), and tobacco (16) identified relevant positions of the core nonapeptide, and the motif (R/K)(L/V/I)X_5(Q/H)(L/A) was established as a canonical consensus sequence (14). Recent investigations took advantage of the increasing number of available sequence data and tried to extract a more restrictive consensus sequence based on sequence comparison (13, 17), which finally led to the suggestion of R(L/V/I/Q)X_2(L/V/I/H)(L/S/G/A)X(H/Q)(L/A) for the most common PTS2 variants and (R/K)(L/V/I/Q)X_2(L/V/I/H/Q)(L/S/G/A/K)X(H/Q)(L/A/F) comprising essentially all known possibilities (17). The binding of PTS2 to PEX7 is mediated by a conserved WD-40 domain of PEX7 usually folding into a propeller-like structure, which is often found in peptide-binding proteins (18). The whole transport process appears saturable and can be inhibited by antibodies blocking chaperones of the Hsp70 and Hsp40 family (19). In most species the N-terminal part of the protein including the PTS2 (transit peptide) is processed inside peroxisomes (20), in mammals by the protease TYSND1 (21).

* This work was supported in part by European Union Project "Peroxisomes" LSHG-C/2004-512018, by Austrian Science Fund (FWF) Projects P15510 and P21950-B20, and by the Austrian Genomforschung in Österreich: Bioinformatik Integrationsnetzwerk (until summer 2007).
[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Tables 1 and 2 and additional references.
[1] To whom correspondence should be addressed. Tel.: 43-1-40160-34091; Fax: 43-1-40160-934203; E-mail: markus.kunze@meduniwien.ac.at.
[2] Supported by the Research Institute of Molecular Pathology until summer 2007.

In humans, the selective defect in the PTS2-dependent import pathway due to mutations in PEX7 leads to the severe disease rhizomelic chondrodysplasia punctata type 1 (RCDP1)[3] (22). Patients suffer from congenital cataracts, growth, and mental retardation, shortening of the upper extremities (rhizomelia), and stippled foci of calcification in epiphyseal cartilage (chondrodysplasia punctata) (23). In mammals, three enzymes are known to harbor a PTS2, namely acyl-CoA thiolase exerting the last step of fatty acid β-oxidation, alkylglycerone-phosphate synthase (alkyldihydroxyacetone phosphate synthase) (24) participating in plasmalogen biosynthesis, and phytanoyl-CoA hydroxylase (25) exerting the first step of the α-oxidation of branched chain fatty acids. Mevalonate kinase, which participates in the synthesis of cholesterol, has been reported to be peroxisomal (26) and harbors a PTS2-like sequence (27), yet no interaction with PEX7 could be found (28).

The quality of algorithms evaluating putative targeting signals based on their similarity to naturally occurring signals depends on a large learning set. This can be compensated by implementation of structural characteristics. However, the quality of the prediction can also serve as criterion for the relevance of the implemented parameters. Recent investigations primarily analyzed the amino acid frequencies at each position of known PTS2 motifs and of putative PTS2 signals encoded in orthologues of PTS2-carrying proteins (13, 17) without evaluation of physical property patterns of side chains or sequence segment-based properties.

Using biochemical, cell biological, and computational methods, we have revealed structural requirements for functional PTS2 signals that are important for their interaction with PEX7. This allowed us to generate a prediction algorithm that identified functional PTS2 signals and a novel peroxisomal protein demonstrating the relevance of the identified criteria.

EXPERIMENTAL PROCEDURES

Cell Culture and Immunofluorescence Microscopy—The green monkey kidney cell line COS7 was purchased from ATCC, and human fibroblasts from RCPD1 patients carrying mutations H39P/W206X have been previously described (29). Cells were cultivated in DMEM (COS7) or RPMI (fibroblasts) supplemented with 10% fetal calf serum (FCS), 2 mM L-glutamine, 50 units/ml penicillin, and 100 μg/ml streptomycin (BioWhittaker). Cells were transfected using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions or electroporation. 48 h after transfection, cells were fixed for 15 min with 4% paraformaldehyde in phosphate-buffered saline (PBS). Cells were washed, permeabilized (5 min with 0.1% Triton X-100 in PBS), and blocked in blocking solution (PBS with 10% FCS and 5% bovine serum albumin (BSA, Roche Applied Science)). After incubation with primary antibodies from different species (rabbit, α-PMP70 (1:2000, ABR, Golden, CO); α-EGFP (1:2000, polyclonal α-EGFP antibody, kindly provided by Prof. Werner Sieghart, Medical University of Vienna); mouse, α-EGFP (1:800, Roche Applied Science), α-ATPase (1:400, A21350, Molecular Probes Eugene, OR), slides were washed with PBS several times and exposed to compatible secondary antibodies (Cy2- and Cy3-labeled goat α-rabbit IgG and goat α-mouse IgG, 1:200, Jackson ImmunoResearch, West Grove, PA). Finally, cells were mounted in PBS/glycerol (1:9) with 3% DABCO (Sigma). For microscopic analysis, transmission microscope BX51 and invert microscope IX71 equipped with CCD cameras (Olympus DP50 and CAM-XM10) and appropriate filter sets were used together with analysis and C-M-cell software (Olympus). During the analysis of subcellular distribution, cells were avoided that showed extremely high expression levels, because in these usually a cytosolic distribution of the reporter protein was observed, probably due to saturation of the PTS2-dependent import pathway.

Cross-complementation—Plasmids encoding $PTS2_{thiolase}$-EGFP HS_3E and myc-hPEX7 variants (ratio 1:3) were transfected into COS7 cells by electroporation, and cells were processed for immunofluorescence microscopy and Western blot analysis as described above.

DNA Cloning—For details on DNA cloning, see supplemental material.

Western Blot Analysis—COS7 were transfected by electroporation, and after growing for 2 days on plastic dishes the cells were harvested. Equal amounts of protein of total cellular extracts were separated by SDS-PAGE and transferred to a nitrocellulose membrane (Schleicher & Schnell). After blocking with 4% skimmed milk in Tris-buffered saline with Tween 20 (TBS-T) (25 mM Tris, pH 7.5, 150 mM NaCl, and 0.05% (w/v) Tween 20), proteins were detected by subsequent incubation of the membrane first with primary antibodies (α-MYC (4A6, mouse, 1:2000, Upstate) and α-β-actin (mouse, 1:15000, Chemicon)) and, after several washings with TBS-T, with horseradish peroxidase-coupled secondary α-mouse IgG antibodies (Dako).

Luciferase Assay—COS7 cells were transfected in 24-well plates using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions with the following plasmids: the appropriate combinations of 0.35 μg of bait (pM-GAL4 encoding plasmids) and 0.35 μg of prey (VP16-DNA-BD encoding plasmids) together with 0.1 μg of luciferase reporter plasmid pFRluc (P1383, Stratagene) and 0.05 μg of pCMV-β-Gal (P204, Promega) for normalization. After 48 h, the cells were washed once with PBS and incubated with 50 μl of lysis buffer (100 mM phosphate buffer, pH 7.8, 0.5% Triton X-100, complete protease inhibitor mixture (Roche Applied Science)) for 20 min. The extracts were centrifuged for 20 min at 15,300 × *g*, and the supernatant was measured. The luciferase assay was performed according to the protocol of the Matchmaker™ system (Clontech) using pRF-Luc vector (Stratagene) for detection of interaction by luminescence measurements.

Sequence Analysis of PTS2 Segments and Three-dimensional Structural Modeling—cDNA sequences of proteins were derived from the NCBI-based GenBank™ data base (30). For comparison of the proteins within the cordata lineage, the Ensembl data base (31) was used.

Sequence Sets—For the generation of the positive set, only soluble proteins were considered that required the PTS2 signal

[3] The abbreviations used are: RCDP1, rhizomelic chondrodysplasia punctata type 1; ER, endoplasmic reticulum; EGFP, enhanced GFP; PDB, Protein Data Bank.

for their import into peroxisomes (*i.e.* the PTS2 is either sufficient to target a reporter protein to peroxisomes or mutations in the PTS2 signal destroyed the peroxisomal targeting signal or the encoding protein was found in the cytosol of PEX7-deficient cells). In contrast, PTS2 signals encoded in membrane proteins, such as rat PEX11 (32) or mouse stearoyl-CoA desaturase (SCD1) (33), were not considered. Thus, in summary, 14 evolutionary independent protein families were identified, namely acyl-CoA thiolase, alkylglycerone-phosphate synthase, phytanoyl-CoA hydroxylase, mevalonate kinase, malate dehydrogenase, citrate synthase, acyl-CoA oxidase, heat shock protein 26 (Hsp26), heat shock protein 70 (Hsp70), transthyretin-like protein, long chain acyl-CoA synthetase, aspartate aminotransferase, amine oxidase, and fructose-1,6-bisphosphate aldolase. If one were to take the whole pool of sequence data from these families, a bias would arise because thiolases are widely conserved in eukaryotic evolution, whereas the majority of the other proteins with PTS2 signals are only found in the plant kingdom (eight families). Metazoa (three families), fungi (one family), or protozoa (one family) together contribute five independent protein families. Moreover, the number of available protein sequences differed between the protein families. To produce an evolutionarily balanced and unbiased set of PTS2 proteins, we selected (if possible) three proteins from each protein family, except for thiolase from which three proteins from each eukaryotic kingdom were selected (supplemental Table 1). Within the kingdoms, the chosen proteins originate from evolutionarily distant species such as fish, amphibians, and mammals from metazoa or monocotyledons and dicotyledons from plant species to cover the whole width of the respective kingdom. Finally, the resulting set of 43 selected sequences was aligned according to their PTS2 nonapeptide motif together with the 15 preceding and 25 succeeding amino acids. The maximal pairwise sequence identity in the motif region was determined to be below 70%.

A negative or background set was created to judge statistical significance of enrichment of amino acids in the PTS2 motif positions. It was derived by random selection of eukaryotic N termini out of the IPI proteomes (34) from *Homo sapiens, Mus musculus, Rattus norvegicus, Danio rerio, Bos taurus, Gallus gallus,* and *Arabidopsis thaliana*, after removing sequences with greater than 98% sequence identity from each proteome (with cd-hit (35)). To obtain more stable background frequencies, the negative set chosen was 10 times bigger than the positive set. Special care was taken so that the length distribution was identical in both sets to replicate the varying distances of PTS2 motifs from the N terminus.

Sequence Logo—The sequence logo in Fig. 2*C* was created with the twosamplelogo webserver (36). Only amino acids are shown at the respective positions that are over-represented in PTS2 motifs with a statistical significance of $p < 0.005$ (*t* test). The coloring is according to amino acid type. The height of amino acid letters and position columns in general are proportional to their level of enrichment.

Entropy Difference Analysis—Significance of positional amino acid restrictions (Fig. 2*A*) was further evaluated with randomized entropy difference analysis as implemented in the HCV database (65). In short, Shannon entropies are calculated for each position of a positive/query and a negative/background alignment (same sets as described above for sequence logo). Next, using a Monte Carlo procedure, the two sets were mixed randomly with replacement resulting in two random sets with the same size as the original two. Positions marked in *red* (Fig. 2*A*) show significance with a p value <0.001 if the random sets would obtain higher entropy difference than the original sets in a maximum of 1 out of 1000 set randomizations.

Physical Property Single Position Deviation—The 20-dimensional vector of amino acid frequencies on each position of the alignment of 43 selected PTS2 motifs was tested for correlation with physical properties from a data base of roughly 700 parameter sets (37–39). The best correlating representatives of nonredundant physical properties for positive charge (40), negative charge (40), bulkiness (41), and aliphatic side chains (42) were selected. Next, the selected physical property parameters were normalized between 0 and 1, and the average was calculated for each position in the PTS2 motif alignment compared with the average of the same physical property in the UniRef50 data base (43). If the absolute value of the difference of the averages at one motif position is higher than twice the absolute value of the median of the physical property over all motif positions, the physical property deviation at the respective position is shown in Fig. 2*B*.

Physical Property Window Deviation—Averages of windows of physical properties (~700 parameter sets from the data base described above) with length 1–12 were evaluated for maximal deviation between the set of 43 selected PTS2 motifs and the Uniref50 data base (43). The influence of different window sizes is balanced by deriving the average not by dividing by the number of positions but by the square root of the number of positions. The resulting property window averages are ranked by their difference from the UniRef50 average, and among sets of redundant properties (with R-value >0.4), only the highest deviating instance is kept. In Fig. 3, we show the identified characteristic physical properties "normalized frequency of α-helix" (44), "flexibility parameter with no rigid neighbors" (45), and "information measure of coil" (46). Only deviations are shown that are consistently above or below the data base average for a window length of at least four positions.

Structural Modeling—The three-dimensional structure of PEX7_HUMAN was modeled according to multiple structural templates identified by the consensus structure prediction server three-dimensional jury (47) by using the stand-alone version of MODELLER (version 9.5) (48). The templates used are histone-binding protein RBBP7 (PDB code 3cfs, chain B) from *H. sapiens* and chromatin assembly factor 1 p55 subunit (PDB code 3c9c, chain A) from *Drosophila melanogaster*. The modeling process was performed in two steps. Step 1 is the building of a three-dimensional structural model according to multitemplates. Dynamic programming-based structural alignment was performed to the aforementioned templates, and the amino acid sequence of PEX7_Human. This process was performed by using the salign class of MODELLER (48), and then 100 structural models were built based on this alignment and the structures of the templates by using the automodel class of MODELLER. At the same time, the discrete optimized protein energy score of each model was calculated, and the one with the

lowest energy was selected for further loop refinement. Step 2 is the loop refinement. According to the alignment, there are some amino acids in PEX7 that are corresponding to gaps in the templates in the alignment. These loop regions can be further refined by using the loop model class of MODELLER. Because the alignment is available, the refinement can be carried out automatically. During this process, 200 models were built; the discrete optimized protein energy score of each model was calculated, and the one with the lowest energy was selected as the final model for further amino acid conservation value mapping.

Mapping of Amino Acid Conservation Values—In total, 41 orthologue sequences of PEX7 were retrieved from the website of OMA (49), which were aligned together with PEX7_Human by using the multiple sequence alignment toolkit of MAFFT (L-INS-I settings) (50). The conservation values of each amino acid of PEX7_Human were calculated by using the method of real valued evolutionary trace (51), excluding positions with more than 30% gaps. Those values were then mapped to the *B*-factor column of the PDB file of the structural model built above. The structure and conservation mapping were then visualized in Yasara (52). To confirm that the observed conservation site in PEX7 is protein family-specific rather than fold-specific, we have repeated the procedure for the RBBP7 protein, which has the same fold as PEX7 but is from a different family and found a distinct pattern of conservation on the side that interacts with the histone helix (data not shown).

Helix Docking—To evaluate possible binding conformations of the putative PTS2 helix, multiple orientations were tried through manual placement, and one tentative candidate orientation was chosen that satisfied the complementary pattern of charge and hydrophobicity. The complex of PEX7 with the PTS2 helix was then energy-minimized through short simulated annealing molecular dynamics simulations using the AMBER03 force field as implemented in Yasara (52).

PTS2 Signal in Silico Screening—The methods and detailed description of the basic PTS2 *in silico* screening algorithm used in this study are summarized in the supplemental material.

RESULTS

The experimentally verified consensus sequence for PTS2 signals is dominated by the characterizations of positions S_1, S_2, S_3, and S_4 in various species.[4] In contrast, restrictions of the central five positions of the signal (X_1–X_5) are hardly understood and primarily extrapolated from amino acid frequencies in naturally occurring PTS2 signals. However, mammalian proteins are usually under-represented in such comparisons.

Mutational Analysis of Canonical PTS2 Signal—Thus, we performed a mutational analysis of these central five amino acids, using human thiolase as model PTS2 peptide. A reporter construct was generated, in which the first 30 amino acids of rat thiolase B were cloned in front of EGFP, and the PTS2 nonapeptide was flanked by two restriction sites (PstI and EcoRI) allowing the simple exchange of nonapeptides (Fig. 1*A*). When the plasmid encoding the reporter protein with the human thiolase PTS2 (RLQVVLGHL) was transfected into COS7 cells and the subcellular localization of EGFP was analyzed by immunofluorescence microscopy, we obtained a punctate staining pattern. EGFP was found to be colocalized with the peroxisomal membrane protein PMP70 (Fig. 1*B*), indicating peroxisomal targeting of the fusion protein. In contrast, when the reporter protein harbored an arbitrary tripeptide (-RSL) instead of a PTS2, EGFP was found evenly distributed across the cell, indicating a cytosolic and nuclear distribution (Fig. 1*C*). This proved that the import of the reporter protein was dependent on a functional PTS2. Using this reporter construct, we analyzed the effect of single amino acid substitutions of the central five amino acids (X_1–X_5) of human thiolase by either acidic (aspartate), basic (lysine or arginine), or bulky hydrophobic (leucine) amino acids (Table 1).

We found that the introduction of a negative charge (aspartate) at position X_2 (VX_2D) and X_3 (VX_3D) destroyed the PTS2, but it was well tolerated at positions X_1 (QX_1D), X_4 (LX_4D), and X_5 (GX_5D). Similarly, the introduction of a positive charge at position X_3 (VX_3K) destroyed the PTS2, but at all other positions the import was retained. Interestingly, the reporter protein was found in peroxisomes and mitochondria when the positive charge was introduced at position X_2 (VX_2K, Fig. 1, *D* and *E*) or X_5 (GX_5R, Fig. 1, *F* and *G*). This was indicated by colocalization with the mitochondrial marker protein ATPase. The introduction of the hydrophobic amino acid leucine at position X_1 (QX_1L) (Fig. 1*H*) and X_5 (GX_5L) did not destroy the PTS2, but the mutation QX_1L introduced an additional ER targeting signal as indicated by colocalization with the ER marker protein-disulfide isomerase (Fig. 1*I*).

In summary, these experiments demonstrate that in an evolutionarily optimized PTS2 (human thiolase), the majority of mutations in the central five amino acids is well tolerated. However, charged residues at position X_3 or a negatively charged residue at position X_2 destroy the PTS2 signal, revealing new restrictions for functional PTS2 signals. Nonetheless, these restrictions are still too loose to explain the low number of functional PTS2 signals and the strong conservation of PTS2 signals across evolution.

Sequence and Physical Property Analysis of Naturally Occurring PTS2 Signals—To elucidate further restrictions for functional PTS2 signals, we performed a detailed sequence and physical property analysis of naturally occurring PTS2 signals. In contrast to previous investigations, we compensated for the over-representation of plant proteins, which contribute more than 50% of protein families harboring PTS2. Therefore, a positive set of PTS2 carrying proteins was compiled, in which each protein family and each phylum are represented by three evolutionarily distant proteins. Moreover, the deviation of amino acid frequencies found in sequences proximal to the N terminus compared with frequencies of overall proteins was taken into account (details on the selection of PTS2 carrying proteins and the determination of amino acid frequency are summarized under "Experimental Procedures").

[4] In this paper, we use a nomenclature that is derived from the published consensus sequence of PTS2 signals (R/K)(L/V/I)X_5(Q/H)(L/A), namely $S_1S_2X_1X_2X_3X_4X_5S_3S_4$, in which S indicates conserved positions of the minimal PTS2 consensus signal, and X_1–X_5 indicate the positions without clear conservation of amino acid types. Moreover, the analyzed motif was extended by the 15 residues preceding (y_{15}–y_1) and 25 residues succeeding (z_1–z_{25}) the core PTS2.

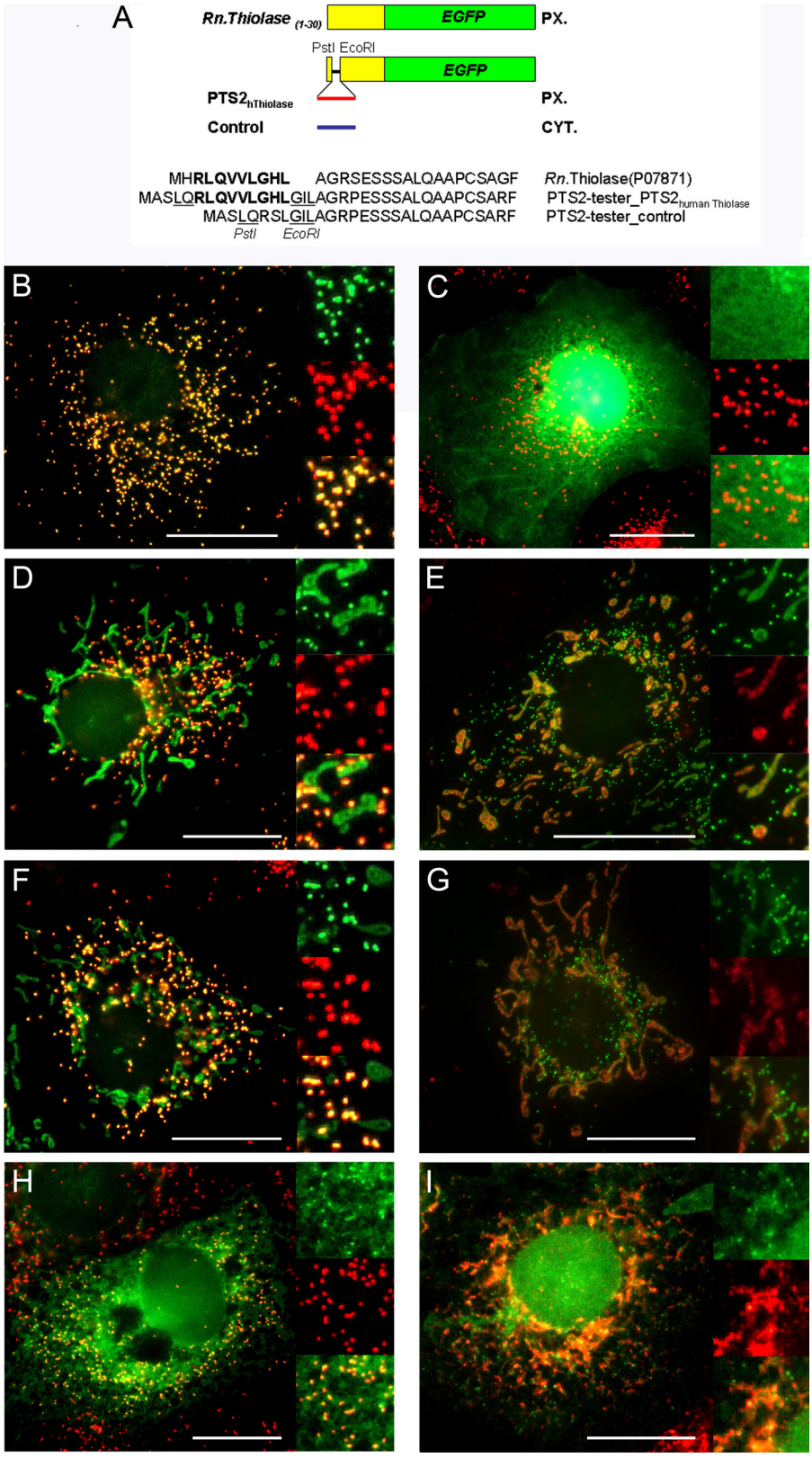


FIGURE 1. **Mutational analysis of the central five amino acids of the human thiolase PTS2.** *A,* schematic representation of the reporter system consisting of the first 30 amino acids of rat thiolase B, in which the PTS2 signal is flanked by restriction sites (*CYT*, cytosolic; *PX*, peroxisomal). The subcellular distribution of the reporter protein encoding the human thiolase PTS2 (*B*) or an arbitrary tripeptide (negative control) (*C*) was investigated by immunofluorescence microscopy using α-EGFP (*green*) and α-PMP70 (*red*) antibodies. Analogously, the subcellular distribution of the reporter protein encoding variants of the human thiolase PTS2, namely V(X_2)K (*D* and *E*), G(X_5)R (*F* and *G*), and Q(X_1)L (*H* and *I*) was investigated either by α-EGFP and α-PMP70 as before or by α-EGFP (*green*) and α-ATPase (*red*) (*E* and *G*) or by α-EGFP (*green*) and α-protein-disulfide isomerase (*PDI*) (*red*) (*I*). *Scale bars,* 20 μm.

TABLE 1
Mutational analysis of the human thiolase PTS2

Wild type	S_1	S_2	X_1	X_2	X_3	X_4	X_5	S_3	S_4	Import[a]
	R	L	Q	V	V	L	G	H	L	+
QX_1D	R	L	**D**[b]	V	V	L	G	H	L	+
VX_2D	R	L	Q	**D**	V	L	G	H	L	−
VX_3D	R	L	Q	V	**D**	L	G	H	L	−
LX_4D	R	L	Q	V	V	**D**	G	H	L	+
GX_5D	R	L	Q	V	V	L	**D**	H	L	+
QX_1K	R	L	**K**	V	V	L	G	H	L	+
VX_2K	R	L	Q	**K**	V	L	G	H	L	+
VX_3K	R	L	Q	V	**K**	L	G	H	L	−
LX_4K	R	L	Q	V	V	**K**	G	H	L	+
GX_5R	R	L	Q	V	V	L	**R**	H	L	+
QX_1L	R	L	**L**	V	V	L	G	H	L	+
GX_5L	R	L	Q	V	V	L	**L**	H	L	+

[a] + indicates import; − indicates no import of the reporter protein harboring the human thiolase PTS2.
[b] Boldface type indicates residues that were introduced into human thiolase PTS2 within the reporter protein context.

When the extended PTS2 motif alignment was analyzed for the information density (Shannon entropy) (Fig. 2*A*) at each position, significant differences between the positive set comprising the PTS2 harboring sequences and the background set were found at all positions of the core PTS2 nonapeptide. Maximum differences were obtained at the characteristic positions of the consensus sequence (S_1–S_4) and at position X_3. However, several positions outside of the core PTS2 signal also showed significant differences between the sets suggesting further restrictions for PTS2 signals.

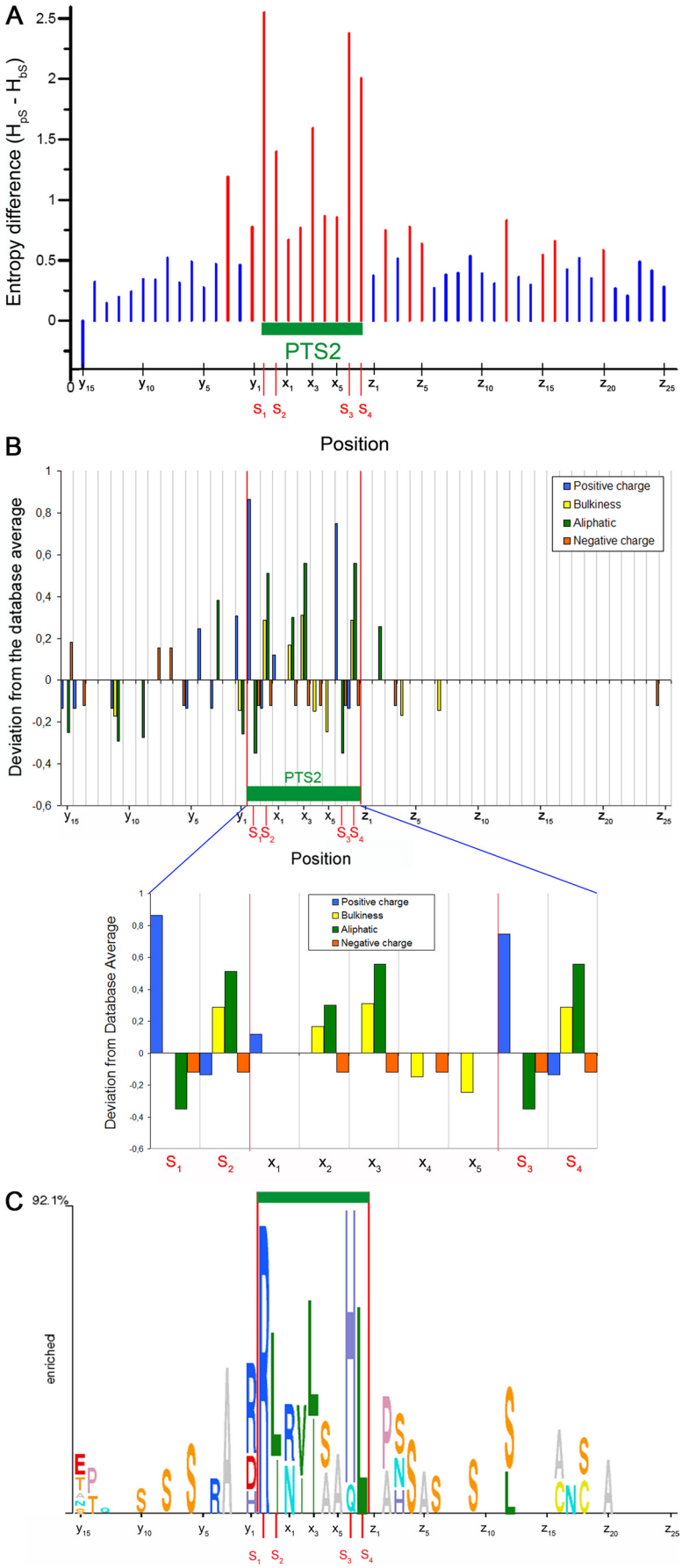
A similar pattern of positions, with significant differences between positive and background set, was found when the relative abundance of classes of amino acids sharing physical properties of their side chains such as charge or bulkiness was compared (Fig. 2*B*, *upper part*). Within canonical PTS2 sequences (Fig. 2*B*, *lower part*), the known characteristics of conserved positions are well reflected in that basic residues are found over-represented at S_1 and S_3 and large hydrophobic residues are preferred at positions S_2 and S_4. The properties of residues preferentially found at position X_3, and to a lesser extent at X_2, resemble those at positions S_2 and S_4. Minor but significant differences between positive and background sets were also found, namely an over-representation of basic residues at position X_1 an under-representation of acidic residues at positions X_2, X_3, and X_4 and of bulky hydrophobic residues at position X_5. Furthermore, significant deviations of the positive set were again found at several positions outside of the core PTS2, but information density and amino acid properties together were significantly different only at positions y_3 (aliphatic), y_1 (basic), and z_2 (aliphatic).

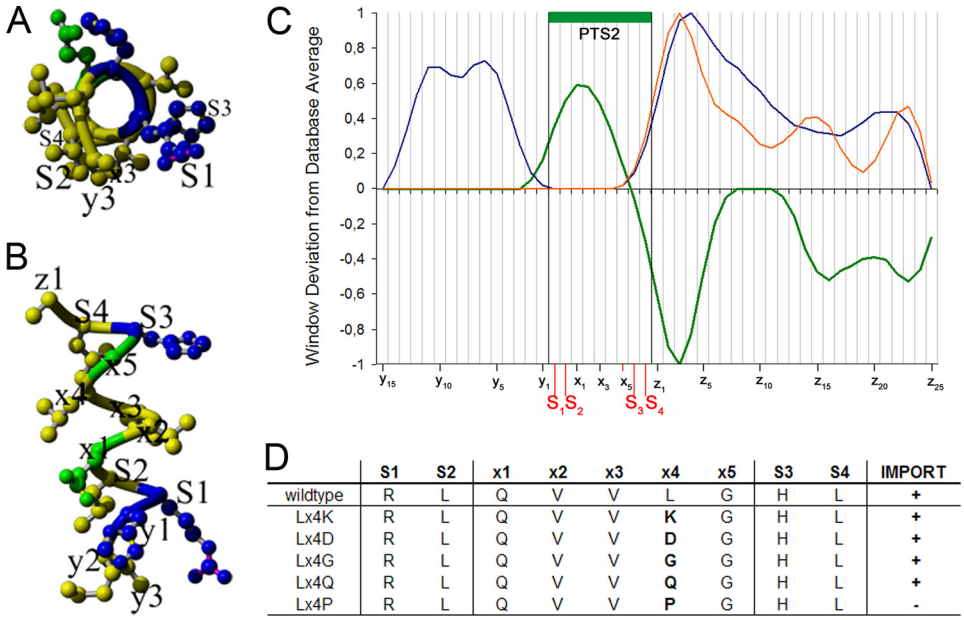
The relative enrichment of individual amino acids within the PTS2 motif alignment also reflects the canonical consensus sequence at positions S_1, S_2, S_3, and S_4 (Fig. 2*C*), except for a lack of lysine over-representation at position S_1. At position X_3, the bulky aliphatic residues leucine and isoleucine were found over-represented. Moreover, minor over-representations were found at many other positions inside and outside the PTS2, but only the preferences for alanine at position y_3, for arginine at position y_1, and for proline at position z_2 coincide with preferences at other levels of conservation. These results corroborate the experimental evidence for the importance of position X_3, but suggest further restrictions for PTS2 signals, which act within and outside the core PTS2.

Helical Structure of PTS2 Motifs—Next, we considered the contribution of position X_3 to functional signals, because there the preference for large and hydrophobic residues coincides with the inactivating effect of charged amino acids. As X_3 is separated from the two other hydrophobic amino acids (S_2 and S_4) by two and three amino acids, respectively, the hydrophobic residues S_2, X_3, and S_4 can be aligned on one side of an α-helix with seven amino acids per two turns. Moreover, the two basic amino acids (S_1 and S_3) would align alongside the helix leading to a positive flank. A model of such an α-helix with the charge distribution pattern of the human thiolase PTS2 is depicted in a top (Fig. 3*A*) or frontal (Fig. 3*B*) view.

When the sequences harboring PTS2 signals (positive data set) were analyzed for their probability to contain α-helical structures (Fig. 3*C*), we found amino acids supporting the formation of α-helices to be over-represented between positions y_2 and S_3 (*green line*), whereas the flanking regions are rich in amino acids mediating a high flexibility (*blue line*) or the formation of coiled structures (*orange line*) rather than regular structures. Thus, naturally occurring PTS2 signals are probably represented by an α-helix, which is flanked by unstructured regions. To corroborate the hypothesis that the ability to form an α-helical structure is necessary for a functional PTS2 signal, the helix-breaking amino acid proline was introduced at the least conserved position, X_4, of the human thiolase PTS2 within the reporter construct. When this plasmid was transfected into COS7 cells, the protein was found diffusely distributed across the cell, indicating that this mutation destroyed the PTS2 signal. In contrast, other mutations at the same position (basic, acidic, large, or small residues) did not interfere with peroxisomal import (Fig. 3*D*).

Interaction between PTS2 and Its Receptor PEX7—Provided that the PTS2 forms a well defined α-helical structure with a conserved charge distribution, the cognate receptor PEX7 should recognize this signal by a complementary domain on its surface. Thus, a three-dimensional homology-based model of human PEX7 was generated as described in detail under "Experimental Procedures." Because the structure of PTS2 appears conserved across evolution, the complementary PTS2-binding domain of PEX7 should behave similarly. Thus, the evolutionarily conserved amino acids on the surface of the PEX7 model were labeled by a color code (Fig. 4*A*), whereas the less conserved sequences are indicated in *gray*. The most conserved region of PEX7 is a cluster on top of the propeller-like structure of the WD-40 domain (Fig. 4*A*), whereas other areas appear less conserved as illustrated in side view (Fig. 4*B*) or a view from the bottom of this structure (Fig. 4*C*). This conserved area forms a groove-like structure, in which the PTS2 helix can be embedded (Fig. 4*B*). Furthermore, the surface charge pattern of this groove appears complementary to that of the hypothetical PTS2 helix (Fig. 4*A*). Interestingly, when mutations in PEX7 occurring in RCDP1 patients were entered into this model of PEX7, the majority of missense mutations, which still result in a stable protein, are located at the conserved part of the protein (indicated as spheres in the *side view* of Fig. 4, *D*, and *top view* of *E*).





D

	S1	S2	x1	x2	x3	x4	x5	S3	S4	IMPORT
wildtype	R	L	Q	V	V	L	G	H	L	+
Lx4K	R	L	Q	V	V	**K**	G	H	L	+
Lx4D	R	L	Q	V	V	**D**	G	H	L	+
Lx4G	R	L	Q	V	V	**G**	G	H	L	+
Lx4Q	R	L	Q	V	V	**Q**	G	H	L	+
Lx4P	R	L	Q	V	V	**P**	G	H	L	-

FIGURE 3. **PTS2 signal is a helical domain separated from the core protein by an unstructured amino acid stretch.** The sterical orientation of amino acid side chains in a model α-helix representing human thiolase PTS2 is depicted in a frontal (*A*) or top (*B*) view. *C,* relative abundance of amino acids supporting different secondary structures (helical *green*; flexible *blue*; and coiled *orange*) is compared between the positive set and the background set and is indicated as normalized fraction (see "Physical Property Window Deviation" under "Experimental Procedures"). *D,* subcellular localization of different EGFP reporter proteins harboring human thiolase PTS2 with single point mutations at position X_4. + indicates import, and − indicates no import.

To test this model, we investigated the interaction between the PTS2 and PEX7 in more detail. Two glutamate residues of PEX7 (Glu-113 and Glu-200) are predicted to lie in close proximity to R(S_1) and H(S_3) of the PTS2 (Fig. 4*A* and schematically depicted in Fig. 5*A*) and should contribute to the interaction between signal and receptor. A third glutamate residue (Glu-287), which resembles the other glutamates with respect to its position in the WD-40 domain and its conservation, but appears remote from the bound PTS2, served as a control. When the interaction between human PEX7 and $PTS2_{thiolase}$-EGFP was measured in a mammalian two-hybrid assay, we found that this interaction caused a strong and specific signal in the luciferase reporter activity (Fig. 5*B*). However, when the glutamate residues Glu-113 and Glu-200 of PEX7 were substituted by arginine, the interaction between PTS2 and these PEX7 variants was no longer detectable. In contrast, the substitution of Glu-287 retained most of the interaction between the PTS2 and PEX7, indicating that such mutations are compatible with a functional WD-40 structure. Next, the functional consequences of these mutations were tested by the restoration of PTS2-mediated import in cultured human fibroblasts of an RCDP1 patient lacking functional PEX7. Therefore, these fibroblasts were cotransfected with expression plasmids for $PTS2_{thiolase}$-EGFP together with either the empty vector (Fig. 5*C*) or with normal human PEX7 carrying an N-terminal Myc tag (Fig. 5*D*) or with mutated variants thereof (E113R in Fig. 5*E*; E200R in Fig. 5*F*). As expected $PTS2_{thiolase}$-EGFP was found evenly distributed across RCDP1 fibroblasts (Fig. 5*C*) but colocalized with PMP70 upon coexpression of myc-hPEX7 (Fig. 5*D*). myc-hPEX7 carrying the mutation E113R was not able to compensate for PEX7 deficiency (Fig. 5*E*), but cotransfection with myc-hPEX7 (E200R) caused a punctate staining against a cytosolic background (Fig. 5*F*) suggesting that the latter mutation can still partially complement PEX7 deficiency.

To further corroborate the close proximity between Glu-200 of PEX7 and histidine at position S_3 of the PTS2, a cross-complementation experiment was performed. When the histidine at position S3 of $PTS2_{thiolase}$-EGFP was substituted by glutamate (HS3E), the PTS2 signal was destroyed (Fig. 5*G*), and this effect was neither compensated by overexpression of myc-hPEX7 (Fig. 5*H*) nor by myc-hPEX7 (E113R) (Fig. 5*I*). However, when $PTS2_{thiolase}$-EGFP (HS3E) was coexpressed with myc-PEX7 (E200R), the reporter protein was found in peroxisomes (Fig. 5*J*), indicating that the mutation HS3E is specifically compensated by E200R. Western blot analysis of protein extracts from similarly transfected COS7 cells demonstrated comparable levels of $PTS2_{thiolase}$-EGFP-HS3E and of the myc-hPEX7 variants (Fig. 5*K*). Together, these results support the three-dimensional model of the interaction between PTS2 thiolase and PEX7 as illustrated in Fig. 4.

PTS2 in Silico Screening Algorithm—Provided that the helical structure followed by a flexible domain is an important characteristic of PTS2 signals, these parameters could serve to categorize peptides that fulfill the minimal consensus sequence of

FIGURE 2. **Computational sequence analysis of core PTS2 signals.** *A,* differences in Shannon entropy between the background and the positive set. *Red bars* indicate positions where the differences are significant ($p < 0.001$). The *green bar* indicates the PTS2 nonapeptide between S_1 and S_4. *B,* relative abundance of amino acids with specific physical properties (*red,* acidic; *blue,* basic; *green,* aliphatic, and *yellow,* bulkiness) is indicated as frequency compared with the background set; the *lower part* shows an enlargement of the core PTS2 sequence. *C,* relative abundance of individual amino acids in the positive set compared with the average of the background set. Basic amino acids are indicated in *blue*, acidic amino acids in *red*, and different colors have been attributed to the other amino acids. See detailed descriptions under the "Experimental Procedures."

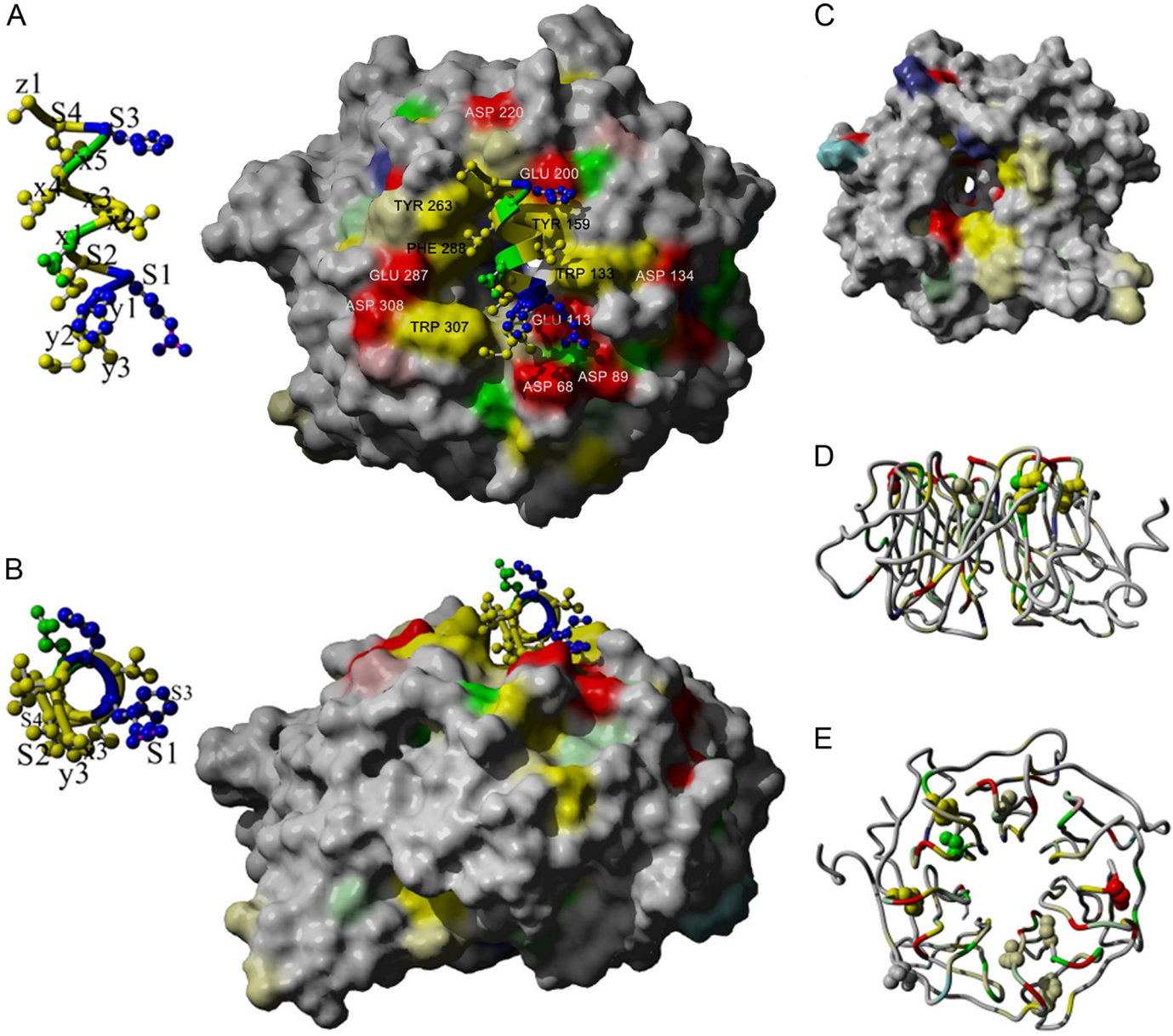


FIGURE 4. **Evolutionarily conserved groove on the surface of PEX7 interacts with the α-helical structure of a prototypical PTS2 motif.** Three-dimensional model of PEX7 based on homology modeling, wherein evolutionarily conserved residues are labeled in *red* (negative), *blue* (positive), *green* (polar), or *yellow* (hydrophobic), whereas nonconserved residues are colored *gray*. The helix representing the human thiolase PTS2 peptide was fitted into the groove and energy-minimized with short simulated annealing MD simulations (see "Experimental Procedures"). The depicted helix orientation represents the putative binding mode with charge complementarity between ligand (helix) and receptor (PEX7). Top (*A*) and side (*B*) projections of the PEX7 model together with the α-helical domain of human thiolase PTS2; *C*, bottom-up view of the PEX7 model; conserved residues are colored as in *A*. Overall, only the top side of the PEX7 propeller surface shows clusters of strongly conserved residues corresponding to the predicted PTS2-binding site. Side (*D*) and top view (*E*) of the PEX7 model with mutations occurring in RCDP1 patients are indicated as *balls*, which all localize to the top side of the propeller.

PTS2 signals. Thus, an *in silico* screening algorithm was developed, which evaluates the N-terminal 40 amino acids of putative PTS2-carrying proteins based on the following: (i) comparison of the core nonapeptide to amino acid frequencies found at each position of naturally occurring PTS2 signals; (ii) restrictions deduced from the mutational analysis of the central five positions (X_1–X_5) of human thiolase PTS2; (iii) evaluation of the helical propensity of the putative PTS2 signal; and (iv) the presence of an unstructured domain C terminus to the PTS2 signal (see supplemental material). This algorithm was used to evaluate the N termini of all human proteins (~39,000 human RefSeq sequences from the NCBI GenBank™ (30)), which were then ranked according to their prediction score. After the exemption of all transmembrane proteins, which should not be substrates for PTS2-mediated protein transport, a list of promising candidates was obtained (30 top candidates except proteins included in the learning set are listed in Table 2). Fourteen of these candidates were chosen for further investigation (boldface in Table 2) to evaluate the reliability of our algorithm. First, the minimal PTS2 signals encoded in these proteins were tested for their ability to mediate peroxisomal targeting in the context of the reporter protein. We found that three peptides encoded in the proteins KChIP4 (potassium channel interacting protein 4), GLOXD1 (glyoxalase domain containing protein 1), and

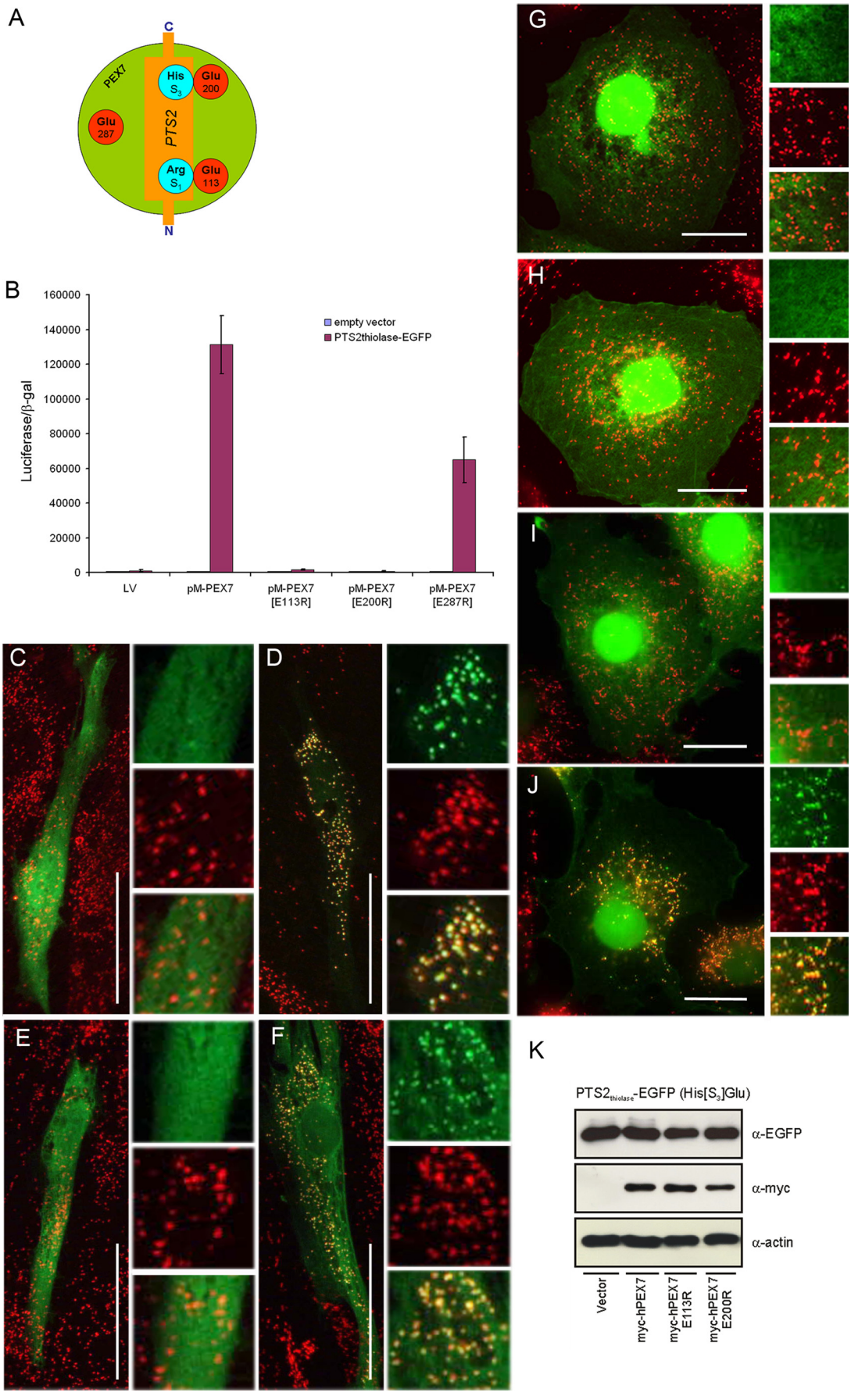


TABLE 2
Human proteins receiving the highest score upon evaluation of putatively encoded PTS2 signals

Score[a]	Position[b]	Accession-nr	PTS2	Description[c]
0.938	13	NP_065071.1	RLQCIKQHL	**Retinoic acid-induced 17**
0.841	31	NP_001001556.1	RVNIIGEHI	**Galactokinase 2 isoform 2**
0.755	4	NP_064587.1	RLALIQLQI	**Nitrilase family, member 2**
0.740	29	NP_077285.2	RLRRLQDQL	A20-binding inhibitor of NF-κB activation 2
0.730	4	NP_671710.1	RVESISAQL	**Kv channel interacting protein 4 isoform 2**
0.711	23	NP_997288.1	RVLTLQCQL	**Hypothetical protein LOC389197**
0.650	16	NP_076425.1	RVRALREQL	Mitochondrial ribosomal protein S34
0.648	35	NP_003229.1	RIEAIRGQI	**Transforming growth factor, β2**
0.614	5	NP_003672.1	RVLSIQSHV	**Pyridoxal kinase**
0.596	20	NP_001055.1	RLRISSIQA	**Transketolase**
0.593	28	NP_932076.1	KVNVFSRQL	**Nucleoside-diphosphate kinase 7 isoform b**
0.587	14	NP_079030.3	RVLLQALQI	Pentatricopeptide repeat domain 2
0.575	15	XP_950375.1	RLYKLHFQL	**Hypothetical protein LOC23045 isoform 6**
0.565	16	NP_056054.1	RIVGLLAQL	**ATP/GTP-binding protein 1**
0.551	23	NP_919339.1	RLDNLMSHL	Ring finger protein 41 isoform 2
0.538	13	NP_079185.1	RITVLDQHL	**Hypothetical protein LOC79969 isoform 2 (C6ORF)**
0.511	6	NP_116145.1	RLCHIAFHV	**Glyoxalase domain containing 1**
0.505	8	NP_116194.1	RLRELCGHW	Zinc finger protein 206
0.495	30	NP_001012979.1	RLAWFLSHL	RP11–506B15.1 protein isoform 2
0.449	26	NP_068741.1	RLLLQALQA	Fanconi anemia, complementation group E
0.428	24	NP_002797.2	RLKELREQL	Proteasome 26 S ATPase subunit 6
0.420	29	NP_689853.3	RIVDVASQV	Decapping enzyme Dcp1b
0.415	32	XP_936057.1	RVLVLATHF	Hypothetical protein LOC91351 isoform 2
0.413	7	NP_001019769.1	RINVSLEQL	Hairy and enhancer of split 3
0.402	2	NP_060238.3	RLKRIAGQD	Leucine-rich repeat containing 40
0.390	18	NP_085147.1	RVSPVHLQI	Apolipoprotein L3 isoform 2
0.387	25	NP_056083.2	RVFSVGTHA	**DnaJ (Hsp40) homologue, subfamily C, member 13**
0.376	34	XP_044178.5	KVKTLQQQL	Similar to neurofilament, heavy polypeptide isoform 1
0.374	34	NP_001014435.1	RLKQFHFHW	Carbonic anhydrase VII isoform 2
0.373	9	XP_939014.1	RIIAILLQV	G protein-coupled receptor 108 isoform 4

[a] Total score of a predicted motif (the higher the better) as evaluated by the basic PTS2 predictor.
[b] Distance of the position S_1 from the start methionine.
[c] Boldface type indicates PTS2 signals that were tested in the reporter protein context.

TGFβ2 (transforming growth factor β2), respectively, acted as functional PTS2 signals and transported EGFP selectively into peroxisomes (Fig. 6, *A–C*). The peptide encoded in the RAI17 (retinoic acid-induced protein/ZIMP10) acted as a PTS2 but also as mitochondrial targeting signal, because the reporter protein was found to colocalize with PMP70 (Fig. 6*D*) and with the mitochondrial marker ATPase (data not shown). The other 10 peptides investigated were not able to target the reporter protein to peroxisomes and thus do not represent PTS2 signals. Thus, roughly 28% of the chosen candidate proteins actually harbor a functional PTS2 signal.

To investigate whether the identified PTS2 signals are functionally active in their native protein context, the subcellular distribution of KChIP4, GLOXD1, TGFβ2, and RAI17 was investigated when expressed as EGFP-tagged full-length proteins. We found that KChIP4-EGFP (Fig. 6*E*) selectively colocalized with PMP70 indicating a peroxisomal localization of the fusion proteins. TGFβ2-EGFP was mainly found colocalized with the ER marker protein-disulfide isomerase, although additional peroxisomal targeting was observed in some cells (data not shown). GLOXD1-EGFP was found to colocalize with the mitochondrial marker MnSOD, and RAI17-EGFP was found in the cytosol and the nucleus of cells (data not shown). The overall summary of our investigation is depicted in Table 3.

DISCUSSION

Since its initial description, the PTS2 has attracted less attention than PTS1. Although major determinants for PTS2 motifs have been elucidated previously, the consensus sequence for this targeting signal was too loose to explain the low number of functional PTS2 signals. Here, important new properties of mammalian PTS2 signals were elucidated, and their binding site on the receptor PEX7 was identified.

The mutational analysis of the central five amino acids of a human thiolase PTS2 identified functional restrictions that exclude specific residues at individual positions of the motif. In contrast, the detailed sequence and physical property analysis of available PTS2 signals reveals the optimal shaping of PTS2 signals by evolutionary adaptation processes.

We can demonstrate experimentally that bulky aliphatic amino acids are not only preferred at position X_3 but are essential for a functional PTS2 as the conversion into a charged residue inactivates the signal. In contrast, at position X_2, lysine is well tolerated, whereas aspartate inactivates the signal although

FIGURE 5. **Interaction between PEX7 and PTS2.** *A,* schematic representation of the interaction face between PEX7 (*green*) and PTS2 (*orange*) as suggested by the three-dimensional model. *B,* mammalian two-hybrid assay. COS7 cells were cotransfected with plasmids encoding $GAL4_{DBD}$-PEX7 or variants thereof (E113R, E200R, and E287R) and VP16-AD-$PTS2_{thiolase}$-EGFP fusion proteins or the empty vector together with a plasmid encoding the UAS_{GAL4}-luciferase reporter and a plasmid expressing β-galactosidase for normalization. *C–F,* immunofluorescence microscopy of human RCDP1 fibroblasts lacking functional PEX7 after cotransfection of expression plasmids for $PTS2_{thiolase}$-EGFP and an empty vector (*EV*) (*C*), or for myc-hPEX7 (*D*), myc-hPEX7 (E113R) (*E*), or myc-hPEX7 (E200R) (*F*) using α-EGFP (*green*) and α-PMP70 (*red*) antibodies. *G–K,* immunofluorescence microscopy of COS7 cells that were cotransfected with the reporter plasmid $PTS2_{thiolase}$-EGFP encoding the mutation HS_3E together with either an empty vector (*G*), myc-hPEX7 (*H*), myc-hPEX7 (E113R) (*I*) or myc-hPEX7 (E200R) (*J*). *K,* Western blot analysis of protein extracts from COS7 cells cotransfected with $PTS2_{thiolase}$-EGFP (HS_3E) together with Myc-tagged versions of PEX7 (myc-hPEX7), myc-hPEX7 (E113R), or myc-hPEX7 E200R or an empty plasmid using α-EGFP and α-Myc antibodies. Labeling with α-β-actin served as loading control. *Scale bars,* 50 μm for *C–F* and 20 μm for *G–J*.

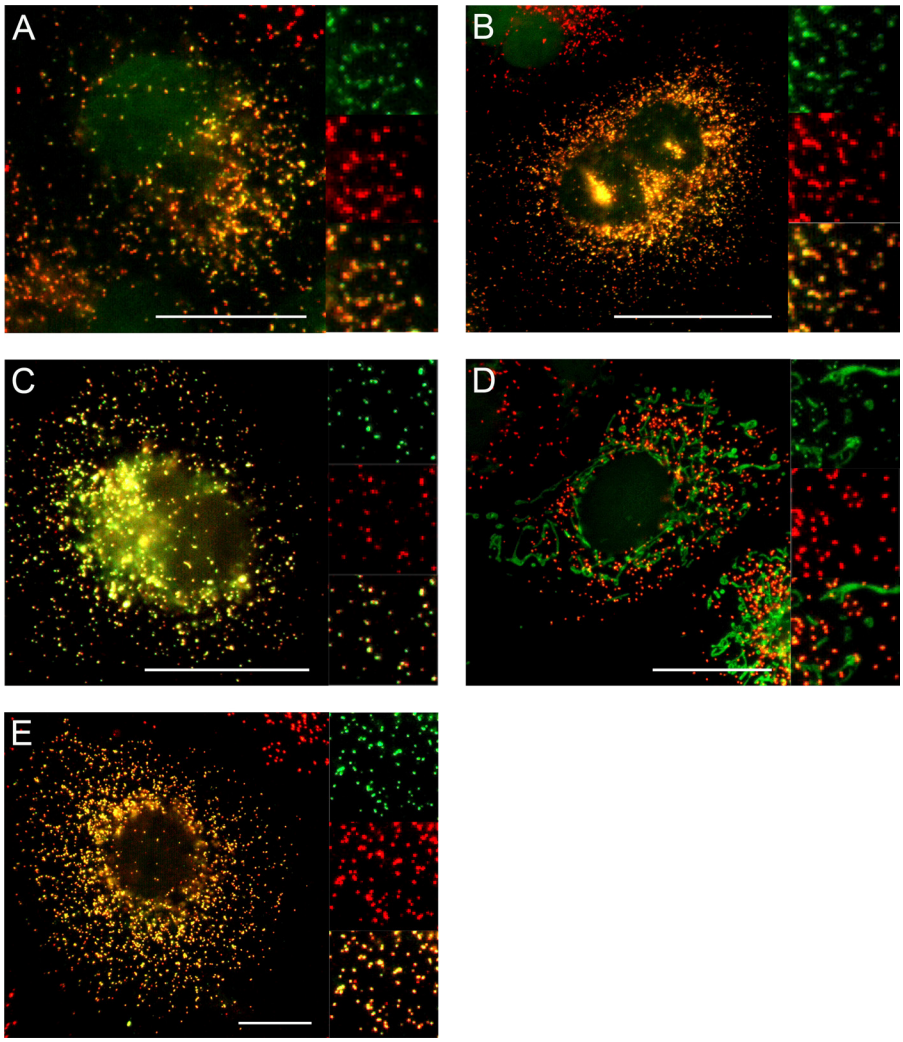


FIGURE 6. **Subcellular localization of reporter proteins with candidate nonapeptides and a full-length EGFP-tagged fusion protein.** Immunofluorescence analysis of COS7 cells transfected with plasmids encoding a reporter protein that harbors the putative PTS2 signal of either KChIP4 (*A*), GLOXD1 (*B*), TGFβ2 (*C*), or RAI17 (*D*) or with a plasmid encoding EGFP-tagged full-length protein of KChIP4 (*E*) using antibodies against EGFP (*green*) and PMP70 (*red*). *Scale bars,* 20 μm.

TABLE 3

Targeting properties of predicted PTS2 motifs of human proteins in the reporter context and in full-length EGFP fusion proteins

Protein	Species	Subcellular localization[a]		S_1	S_2	X_1	X_2	X_3	X_4	X_5	S_3	S_4
		Protein	Peptide									
KCHiP4	*H. sapiens*	PX	PX	R	V	E	S	I	S	A	Q	L
Transforming growth factor-β (*TGFβ*)	*H. sapiens*	ER/PX	PX	R	I	E	A	I	R	G	Q	I
Glyoxylase domain interacting protein (*GLOXD1*)	*H. sapiens*	MITO	PX	R	L	C	H	I	A	F	H	V
RIA17	*H. sapiens*	NUC/CYT	PX + MITO	R	L	Q	C	I	K	Q	H	L
Galactokinase 2 isoform2	*H. sapiens*	ND	CYT	R	V	N	I	I	G	E	H	I
Nitrilase	*H. sapiens*	ND	CYT	R	L	A	L	I	Q	L	Q	I
LOC389197	*H. sapiens*	ND	CYT	R	L	L	T	L	Q	C	Q	L
Pyridoxal kinase	*H. sapiens*	ND	CYT	R	V	L	S	I	Q	S	H	V
Transketolase	*H. sapiens*	ND	CYT	R	L	R	I	S	S	I	Q	A
Nucleoside-diphosphate kinase 7 isoform_b	*H. sapiens*	ND	CYT	K	V	N	V	F	S	R	Q	L
LOC23045	*H. sapiens*	ND	CYT	R	L	Y	K	L	H	F	Q	L
ATP/GTP-binding protein	*H. sapiens*	ND	MITO	R	I	V	G	L	L	A	Q	L
LOC79969 (C6ORF)	*H. sapiens*	ND	CYT	R	I	T	V	L	D	Q	H	L
DnaJ (Hsp40) homologue C13	*H. sapiens*	ND	CYT	R	V	F	S	V	G	T	H	A

[a] PX, peroxisomal; ER, endoplasmic reticulum; MITO, mitochondrial; NUC, nuclear; CYT, cytoplasmic; ND, not determined.

the amino acid preference resembles that of position X_3. However, basic residues at position X_2 generate an additional mitochondrial targeting signal, suggesting that the avoidance of a competing targeting signal is an additional reason for the under-representation of amino acid classes at individual positions. Similarly, basic residues at X_5 (arginine) can promote additional mitochondrial targeting, and large aliphatic residues (leucine) at X_1 can generate an additional targeting signal for the ER. In addition to the three type-changing substitutions that inactivate the PTS2, sequence alignment of known PTS2

motifs suggested further restrictions that are reflected by over- and under-representation of specific types of amino acids (*e.g.* bulky residues are significantly underrepresented at position X_5, but the large hydrophobic amino acid leucine is well tolerated).

It remains to be emphasized that our mutational analysis was performed in a reporter system that exposed the PTS2 at the extreme N terminus and thus might confer fewer restrictions than PTS2 signals located further away from the N terminus. Moreover, our investigations revealed a bipartite structural motif, which appears conserved across all PTS2 signals, namely a helical structure supposed to interact with PEX7 and an unstructured region connecting the core protein with the actual PEX7-interacting sequence segment.

The helical structure is indicated by the over-representation of amino acids supporting the formation of α-helices and the absence of the helix breaking amino acid proline in naturally occurring PTS2 signals. Furthermore, the PTS2 signal of human thiolase is sensitive to the insertion of proline at position X_4, whereas large, small, acidic, and basic amino acids are well tolerated. In a helical structure, the two basic residues (S_1 and S_3) as well as the two large hydrophobic residues (S_2 and S_4) are aligned on one side of the helix, separated by two turns (Fig. 3, *A* and *B*). Together with the important position X_3, they comprise one side of the helix with a hydrophobic area aligned by positive charges. The other side of the helix appears less conserved, and functional PTS2 signals are compatible with all amino acid classes at positions X_1, X_2, X_4, and X_5, except for a negative charge at position X_2. However, in an α-helix, X_2 is positioned in close proximity to the basic residues S_1 and S_3, and a negative charge could neutralize one of these charges and thereby inactivate the PTS2. The helical structure of PTS2 signals resembles mitochondrial targeting signals, which have been described as positively charged amphipathic helices (53). This similarity is supported by our observation that single point mutations in PTS2 thiolase such as V(X_2)K or G(X_5)R can generate a mitochondrial targeting signal without affecting peroxisomal targeting. These mutations alter the side of the helix that is not required for peroxisomal targeting, whereas in rat thiolase PTS2 the substitution of histidine (S_3) by basic amino acids also generated mitochondrial targeting but destroyed the PTS2 (5).

As a helical PTS2 motif exposes all highly conserved residues on one face of the helix, the interaction with the receptor PEX7 should involve this side of the helix. Accordingly, the newly generated three-dimensional model of PEX7 revealed a groove on the most conserved surface area, which shows a charge distribution complementary to the conserved side of the PTS2 helix. Moreover, many missense mutations in PEX7 identified in patients suffering from RCDP1 (54) affect this region.

Overall, our refined three-dimensional model of PEX7 resembles structures suggested previously (29, 55), but it identified conserved residues that appear concentrated on one side of the WD-40 structure and allowed the prediction of residues that contribute to the interaction. Two conserved glutamic acid residues of PEX7 (Glu-113 and Glu-200) are proposed to interact with arginine S_1 and histidine S_3, respectively. Substitution of each of these glutamates by arginine reduced the interaction of PEX7 with PTS2 thiolase below the detection level of the mammalian two-hybrid assay, whereas another glutamate to arginine mutation at a similar position in the WD-40 domain largely retained the interaction. Moreover, the mutation E113R in myc-hPEX7 destroyed its ability to complement PEX7 deficiency in RCDP1 fibroblasts, and the mutation E200R can restore the import of a PTS2 thiolase variant harboring the reciprocal charge exchange HS_3E, which normally inactivates the PTS2. Thus, we consider the predicted model of interaction highly probable despite the surprising finding that myc-hPEX7 (E200R) partially complements PEX7 deficiency. The latter could be due to residual binding between PTS2 thiolase and PEX7-E200R, which was below the detection limit of the two-hybrid assay. In the two-hybrid system, the PTS2 is fused to the VP16 activation domain, which dislocates the PTS2 from the extreme N terminus, and this might render the strength of the interaction with PEX7 more sensitive to mutations. Our prediction poses the PTS2 helix horizontally in the groove on top of the WD-40 domain of PEX7 and thus differs from the insertion of a linear unfolded peptide into the channel in the middle of the propeller as suggested previously (55). This orientation is supported by docking experiments and by the fact that the PTS2 signal of some proteins appears up to 37 amino acids away from the start (human alkylglycerone-phosphate synthase), which renders a linear insertion of the N terminus before the recognition of the PTS2 signal less probable.

The exposure of the PTS2 signal away from the bulk of the protein was concluded from predictions that indicate an unstructured linker region, as noticed between C-terminal PTS1 motifs and the core proteins. Our computational analysis is in agreement with three-dimensional structures of naturally occurring proteins, as in the available x-ray structures of human acyl-CoA thiolase (PDB code 2IIK) and human phytanoyl-CoA hydroxylase (PDB code 2A1X) the N-terminal sequences, including the putative linker regions, were not resolved suggesting that these sequences are not sufficiently structured. Such a linker has been shown to be of functional importance for the exposure of the PTS1 signal (56, 57), but in the case of PTS2 the linker domain should also contain the cleavage site for the processing peptidase.

We are confident that the identified criteria are relevant for typical PTS2 signals because their implementation into an *in silico* screening algorithm allowed the generation of a preliminary PTS2 prediction program, which led to a hit rate of 4 out of 14 when testing a list of PTS2 signals with a high PTS2 score derived from the whole human proteome. Moreover, KChIP4 (58) was imported into peroxisomes when expressed as EGFP-tagged full-length protein, and thus the algorithm led to the identification of a novel peroxisomal protein. EGFP-tagged full-length TGFβ2 (59) was found predominantly in the ER, GLOXD1 in mitochondria, and RAI17 in the nucleus and cytosol as described recently (60) suggesting that the targeting signals for these organelles can over-rule PTS2 as previously demonstrated for PTS1 signals (56). Alternatively, the lack of peroxisomal targeting despite a functional PTS2 might be due to a modulating influence of the amino acids directly surrounding the PTS2.

The newly identified peroxisomal protein KChIP4 (Kv-channel interacting protein 4) was first described to interact with a potassium channel and presenilin (60). However, the protein appears in various splice variants, some of which have been described in the cytosol or at the plasma membrane (61, 62). The subcellular localization of the variant analyzed in this study has not been investigated, although its expression is well described. KChIP4 belongs to a family of proteins harboring the structural motif of an EF-hand, which mediates structural changes upon Ca^{2+} binding (63). The expression profile of KChIP4 (64) shows brain selectivity, which explains the absence of the protein from peroxisomal fractions analyzed by proteomic approaches.

In summary, this investigation refines structural requirements for functional PTS2 signals and suggests a model for the interaction with the receptor PEX7. These criteria allowed the identification of four functional PTS2 signals encoded in human proteins and a novel peroxisomal protein.

Acknowledgments—We thank Manuela Haberl for technical assistance; Kalsoom Sughra for supporting the performance of luciferase assays; Sonja Forss-Petter, Christoph Wiesinger, and Fabian Dorninger for critically reading the manuscript, and Andreas Hartig and Michael Schuster for helpful discussions.

REFERENCES

1. Wanders, R. J., and Waterham, H. R. (2006) *Annu. Rev. Biochem.* **75,** 295–332
2. Gould, S. G., Keller, G. A., and Subramani, S. (1987) *J. Cell Biol.* **105,** 2923–2931
3. Miyazawa, S., Osumi, T., Hashimoto, T., Ohno, K., Miura, S., and Fujiki, Y. (1989) *Mol. Cell. Biol.* **9,** 83–91
4. Swinkels, B. W., Gould, S. J., Bodnar, A. G., Rachubinski, R. A., and Subramani, S. (1991) *EMBO J.* **10,** 3255–3262
5. Osumi, T., Tsukamoto, T., and Hata, S. (1992) *Biochem. Biophys. Res. Commun.* **186,** 811–818
6. Brocard, C., Kragler, F., Simon, M. M., Schuster, T., and Hartig, A. (1994) *Biochem. Biophys. Res. Commun.* **204,** 1016–1022
7. Van der Leij, I., Franse, M. M., Elgersma, Y., Distel, B., and Tabak, H. F. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90,** 11782–11786
8. Marzioch, M., Erdmann, R., Veenhuis, M., and Kunau, W. H. (1994) *EMBO J.* **13,** 4908–4918
9. Braverman, N., Steel, G., Obie, C., Moser, A., Moser, H., Gould, S. J., and Valle, D. (1997) *Nat. Genet.* **15,** 369–376
10. Holroyd, C., and Erdmann, R. (2001) *FEBS Lett.* **501,** 6–10
11. Glover, J. R., Andrews, D. W., and Rachubinski, R. A. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91,** 10541–10545
12. McNew, J. A., and Goodman, J. M. (1994) *J. Cell Biol.* **127,** 1245–1257
13. Reumann, S. (2004) *Plant Physiol.* **135,** 783–800
14. Glover, J. R., Andrews, D. W., Subramani, S., and Rachubinski, R. A. (1994) *J. Biol. Chem.* **269,** 7558–7563
15. Tsukamoto, T., Hata, S., Yokota, S., Miura, S., Fujiki, Y., Hijikata, M., Miyazawa, S., Hashimoto, T., and Osumi, T. (1994) *J. Biol. Chem.* **269,** 6001–6010
16. Flynn, C. R., Mullen, R. T., and Trelease, R. N. (1998) *Plant J.* **16,** 709–720
17. Petriv, O. I., Tang, L., Titorenko, V. I., and Rachubinski, R. A. (2004) *J. Mol. Biol.* **341,** 119–134
18. Lazarow, P. B. (2006) *Biochim. Biophys. Acta* **1763,** 1599–1604
19. Legakis, J. E., and Terlecky, S. R. (2001) *Traffic* **2,** 252–260
20. Hijikata, M., Ishii, N., Kagamiyama, H., Osumi, T., and Hashimoto, T. (1987) *J. Biol. Chem.* **262,** 8151–8158
21. Kurochkin, I. V., Mizuno, Y., Konagaya, A., Sakaki, Y., Schönbach, C., and Okazaki, Y. (2007) *EMBO J.* **26,** 835–845
22. Motley, A. M., Hettema, E. H., Hogenhout, E. M., Brites, P., ten Asbroek, A. L., Wijburg, F. A., Baas, F., Heijmans, H. S., Tabak, H. F., Wanders, R. J., and Distel, B. (1997) *Nat. Genet.* **15,** 377–380
23. Purdue, P. E., Skoneczny, M., Yang, X., Zhang, J. W., and Lazarow, P. B. (1999) *Neurochem. Res.* **24,** 581–586
24. de Vet, E. C., and van den Bosch, H. (2000) *Cell Biochem. Biophys.* **32,** 117–121
25. Jansen, G. A., Mihalik, S. J., Watkins, P. A., Moser, H. W., Jakobs, C., Denis, S., and Wanders, R. J. (1996) *Biochem. Biophys. Res. Commun.* **229,** 205–210
26. Biardi, L., Sreedhar, A., Zokaei, A., Vartak, N. B., Bozeat, R. L., Shackelford, J. E., Keller, G. A., and Krisans, S. K. (1994) *J. Biol. Chem.* **269,** 1197–1205
27. Olivier, L. M., and Krisans, S. K. (2000) *Biochim. Biophys. Acta* **1529,** 89–102
28. Ghys, K., Fransen, M., Mannaerts, G. P., and Van Veldhoven, P. P. (2002) *Biochem. J.* **365,** 41–50
29. Braverman, N., Chen, L., Lin, P., Obie, C., Steel, G., Douglas, P., Chakraborty, P. K., Clarke, J. T., Boneh, A., Moser, A., Moser, H., and Valle, D. (2002) *Hum. Mutat.* **20,** 284–297
30. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. (2009) *Nucleic Acids Res.* **37,** D26–D31
31. Hubbard, T. J., Aken, B. L., Ayling, S., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., Clarke, L., Coates, G., Fairley, S., Fitzgerald, S., Fernandez-Banet, J., Gordon, L., Graf, S., Haider, S., Hammond, M., Holland, R., Howe, K., Jenkinson, A., Johnson, N., Kahari, A., Keefe, D., Keenan, S., Kinsella, R., Kokocinski, F., Kulesha, E., Lawson, D., Longden, I., Megy, K., Meidl, P., Overduin, B., Parker, A., Pritchard, B., Rios, D., Schuster, M., Slater, G., Smedley, D., Spooner, W., Spudich, G., Trevanion, S., Vilella, A., Vogel, J., White, S., Wilder, S., Zadissa, A., Birney, E., Cunningham, F., Curwen, V., Durbin, R., Fernandez-Suarez, X. M., Herrero, J., Kasprzyk, A., Proctor, G., Smith, J., Searle, S., and Flicek, P. (2009) *Nucleic Acids Res.* **37,** D690–D697
32. Passreiter, M., Anton, M., Lay, D., Frank, R., Harter, C., Wieland, F. T., Gorgas, K., and Just, W. W. (1998) *J. Cell Biol.* **141,** 373–383
33. Su, X., Han, X., Yang, J., Mancuso, D. J., Chen, J., Bickel, P. E., and Gross, R. W. (2004) *Biochemistry* **43,** 5033–5044
34. Kersey, P. J., Duarte, J., Williams, A., Karavidopoulou, Y., Birney, E., and Apweiler, R. (2004) *Proteomics* **4,** 1985–1988
35. Li, W., and Godzik, A. (2006) *Bioinformatics* **22,** 1658–1659
36. Vacic, V., Iakoucheva, L. M., and Radivojac, P. (2006) *Bioinformatics* **22,** 1536–1537
37. Eisenhaber, F., and Bork, P. (1998) *Trends Cell Biol.* **8,** 169–170
38. Kawashima, S., Ogata, H., and Kanehisa, M. (1999) *Nucleic Acids Res.* **27,** 368–369
39. Maurer-Stroh, S., and Eisenhaber, F. (2005) *Genome Biol.* **6,** R55
40. Fauchère, J. L., Charton, M., Kier, L. B., Verloop, A., and Pliska, V. (1988) *Int. J. Pept. Protein Res.* **32,** 269–278
41. Zimmerman, J. M., Eliezer, N., and Simha, R. (1968) *J. Theor. Biol.* **21,** 170–201
42. Zvelebil, M. J., Barton, G. J., Taylor, W. R., and Sternberg, M. J. (1987) *J. Mol. Biol.* **195,** 957–961
43. Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R., and Wu, C. H. (2007) *Bioinformatics* **23,** 1282–1288
44. Chou, P. Y., and Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol.* **47,** 45–148
45. Karplus, P. A., and Schulz, G. E. (1985) *Naturwissenschaften* **72,** 212–213
46. Robson, B., and Suzuki, E. (1976) *J. Mol. Biol.* **107,** 327–356
47. Ginalski, K., Elofsson, A., Fischer, D., and Rychlewski, L. (2003) *Bioinformatics* **19,** 1015–1018
48. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U., and Sali, A. (2007) *Curr. Protoc. Protein Sci.* Chapter 2, Unit 2.9
49. Schneider, A., Dessimoz, C., and Gonnet, G. H. (2007) *Bioinformatics* **23,** 2180–2182
50. Katoh, K., Kuma, K., Toh, H., and Miyata, T. (2005) *Nucleic Acids Res.* **33,** 511–518

51. Mihalek, I., Res, I., and Lichtarge, O. (2004) *J. Mol. Biol.* **336,** 1265–1282
52. Krieger, E., Koraimann, G., and Vriend, G. (2002) *Proteins* **47,** 393–402
53. Roise, D., and Schatz, G. (1988) *J. Biol. Chem.* **263,** 4509–4511
54. Motley, A. M., Brites, P., Gerez, L., Hogenhout, E., Haasjes, J., Benne, R., Tabak, H. F., Wanders, R. J., and Waterham, H. R. (2002) *Am. J. Hum. Genet.* **70,** 612–624
55. Stanley, W. A., Fodor, K., Marti-Renom, M. A., Schliebs, W., and Wilmanns, M. (2007) *FEBS Lett.* **581,** 4795–4802
56. Neuberger, G., Kunze, M., Eisenhaber, F., Berger, J., Hartig, A., and Brocard, C. (2004) *Genome Biol.* **5,** R97
57. Neuberger, G., Maurer-Stroh, S., Eisenhaber, B., Hartig, A., and Eisenhaber, F. (2003) *J. Mol. Biol.* **328,** 567–579
58. Morohashi, Y., Hatano, N., Ohya, S., Takikawa, R., Watabiki, T., Takasugi, N., Imaizumi, Y., Tomita, T., and Iwatsubo, T. (2002) *J. Biol. Chem.* **277,** 14965–14975
59. de Martin, R., Haendler, B., Hofer-Warbinek, R., Gaugitsch, H., Wrann, M., Schlüsener, H., Seifert, J. M., Bodmer, S., Fontana, A., and Hofer, E. (1987) *EMBO J.* **6,** 3673–3677
60. Sharma, M., Li, X., Wang, Y., Zarnegar, M., Huang, C. Y., Palvimo, J. J., Lim, B., and Sun, Z. (2003) *EMBO J.* **22,** 6101–6114
61. Jerng, H. H., and Pfaffinger, P. J. (2008) *J. Biol. Chem.* **283,** 36046–36059
62. Liang, P., Wang, H., Chen, H., Cui, Y., Gu, L., Chai, J., and Wang, K. (2009) *J. Biol. Chem.* **284,** 4960–4967
63. Grabarek, Z. (2006) *J. Mol. Biol.* **359,** 509–525
64. Xiong, H., Kovacs, I., and Zhang, Z. (2004) *Brain Res. Mol. Brain Res.* **128,** 103–111
65. Kuiken, C., Yusim, K., Boykin, L., and Richardson, R. (2005) *Bioinformatics* **21,** 379–384

Suppl.TAB.I List of selected PTS2 carrying proteins and their orthologues

Protein	Organism	Acc.number[a]	Distance	Import	PTS2[b]	Literature
3-ketoacyl-CoA thiolase	*Danio rerio*	NP_001002207.1	3	nd.	sim+	
3-ketoacyl-CoA thiolase	*Xenopus laevis*	AAH54299	3	nd.	sim+	
3-ketoacyl-CoA thiolase	*Dictyostelium discoideum*	AAO51864	6	nd.	sim+	
3-ketoacyl-CoA thiolase A	*Candida tropicalis*	P33290	3	+	sim+	Kurihara (1988) FEBS Lett 229(1): 215-218
3-ketoacyl-CoA thiolase	*Yarrowia lipolytica*	Q05493	3	nd.	sim+	
KAT5/3-ketoacyl-CoA thiolase	*Arabidopsis thaliana*	NP_568704.2	7	+	sim+	Carrie (2007), Plant Mol. Biol, 63:97-108
3-ketoacyl-CoA thiolase A	*Rattus norvegicus*	P21775	13	+	+	Swinkels (1991) EMBO J. 10: 3255-3262; Osumi (1991) BBRC 181(3) 947-954
3-ketoacyl-CoA thiolase	*Cucurbita sp.*	D70895	7	+	+	Kato (1996) Plant Mol.Biol. 31(4):843-52.
3-ketoacyl-CoA thiolase	*Saccharomyces cerevisiae*	NP_012106	4	+	+	Erdmann (1994) Yeast 10: 935-944
Phytanoyl-CoA dioxygenase	*Rattus norvegicus*	P57093	9	+	sim+	Jansen (1999) J Lipid Res 40: 2244-2254
Phytanoyl-CoA dioxygenase	*Xenopus laevis*	ENSXETP00000019993	15	nd.	sim+	
Phytanoyl-CoA dioxygenase	*Tetraodon nigroviridis*	CAF96289	7	nd.	sim+	
Alkylglycerone phosphate synthase	*Danio rerio*	AAH45516.1	15	nd.	sim+	
Alkylglycerone phosphate synthase	*Xenopus laevis*	ENSXETP00000018518	17	nd.	sim+	
Alkylglycerone phosphate synthase	*Homo sapiens*	NP_003650	37	+	sim+	De Vet (1998), JBC 273/17, pp. 10296–10301
Citrate synthase	*Dictyostelium discoideum*	XP_647596.1	6	nd.	sim+	
Citrate synthase 2 (CSY2)	*Arabidopsis thaliana*	NP_191434.1	6	+	+	Pracharowenwattana (2005), Plant Cell, 17:2037-48
Citrate synthase	*Cucurbita maxima*	P49299	16	+	+	Kato (1995) Plant Mol Biol 27(2): 377-390, Kato, 1996
Acyl CoA oxidase homolog	*Cucurbita sp.*	T07901	21	+	sim+	Hayashi (1998) JBC 273(14): 8301-8307
AFR302Wp (ACOX)	*Eremothecium gossypii*	NP_985849	4	nd.	sim+	
Acyl-CoA oxidase	*Arabidopsis thaliana*	T52120	23	nd.	sim+	
Fructose-bisphosphate aldolase	*Trypanosoma brucei brucei*	P07752	4	+	+	Blattner (1995) FEBS Lett 360: 310-314
Fructose-bisphosphate aldolase	*Leishmania mexicana*	CAB55315	3	nd.	sim+	
Malate dehydrogenase	*Cucurbita sp.*	D90257	10	+	+	Kato (1998), Plant Cell Physiology, 39/2 186-195
Malate dehydrogenase	*Oryza sativa*	Q42972	10	nd.	sim+	
NAD-dependent malate dehydrogenase	*Arabidopsis thaliana*	CAA10321	10	nd.	sim+	
Aspartate aminotransferase/Asp3	*Glycine max*	AAC50015	25	+	+	Gebhardt (1998), Plant Mol.Biology 37: 99-108
Aspartate aminotransferase 3	*Arabidopsis thaliana*	NP_196713.1	15	nd.	sim+	
Hsp26 (AtAcd3.1-2)	*Arabidopsis thalliana*	At1g06460	8	+	+	Ma (2006), Plant Physiol 141: 47-60
Alpha-crystalling domain 31.2	Medicago truncatula	ACJ84938.1	9	nd.	sim+	
Alpha-crystalling domain 31.2	Vitis vinifera	CAO69939.1	9	nd.	sim+	
Transthyretin-like protein (TTLP)	Arabidopsis thalliana	NP_200630.1	182	+	+	Reumann (2007), Plant Cell 19: 3170-93
Transthyretin-like protein (TTLP)	Zea mays	ACF86780.1	202	nd.	sim+	
Probable long-chain-fatty-acid-CoA ligase	*Brassica napus*	T07944	10	nd.	sim+	
Long-chain acyl-CoA synthetase 6	*Arabidopsis thaliana*	NP_566265	15	+	+	Fulda (2002) The Plant J 32: 93-103
Long chain acyl-CoA synthetase 7	*Arabidopsis thaliana*	AAM28874	10	+	+	Fulda (2002) The Plant J 32: 93-103

[a] accession number derived from NCBI protein database or Ensembl protein database

[b] + import proved, sim+ homologous protein to protein with proved import

SUPPL.TAB. 2: Oligonucleotides

Cloning for PTS2-tester

Oligonucleotide number	*Direction*	*Protein*	*Species*	*Oligo sequences*
Oli_600	fw	EGFP_N1		CCACCGGTCGCCACGATCGTGAGCAAGGG
Oli_601	rev	EGFP_N1		CCCTTGCTCACGATCGTGGCGACCGGTGG
Oli_1093	fw	Thiolase	*Rattus norvegicus*	TCGAGatggcaagtctgcagagaagcttggG
Oli_1094	rev	Thiolase	*Rattus norvegicus*	AATTCccaagcttctctgcagacttgccatC
Oli_1095	fw	Thiolase	*Rattus norvegicus*	aattctggccggccggcccgagtcgagctccgcgctgc aagctgcgccctgctccgctaggttcggG
Oli_1096	rev	Thiolase	*Rattus norvegicus*	GATCCcgaacctagcggagcagggcgcagcttgcagc gcggagctcgactcgggccggccggccaG

Thiolase mutations

Oligonucleotide number	Direction	Protein	Species	Mutation	Oligo sequences	Plasmid-number
Oli_1133	fw	thiolase	*Homo sapiens*		GaggctgttctcagctctgctgcaggttgG	P629
Oli_1134	rev	thiolase	*Homo sapiens*		AATTCcaacctgcagcagagctgagaacagcctCTGCA	
Oli_1415	fw	thiolase	*Homo sapiens*	Q1L	GcggctgcTAgtagtgctgggccacctggG	P914
Oli_1416	rev	thiolase	*Homo sapiens*	Q1L	aattcccaggtggcccagcactacTAgcagccgctgca	
Oli_1190	fw	thiolase	*Homo sapiens*	Q1D	GcggctgGACgtagtgctgggccacctggG	P668
Oli_1191	rev	thiolase	*Homo sapiens*	Q1D	aattcccaggtggcccagcactacGTCcagccgctgca	
Oli_1192	fw	thiolase	*Homo sapiens*	V2D	GcggctgcaggATgtgctgggccacctggG	P669
Oli_1193	rev	thiolase	*Homo sapiens*	V2D	aattcccaggtggcccagcacTAcctgcagccgctgca	
Oli_1417	fw	thiolase	*Homo sapiens*	V2K	GcggctgcaAAAGgtgctgggccacctggG	P927
Oli_1418	rev	thiolase	*Homo sapiens*	V2K	aattcccaggtggcccagcacCTTTtgcagccgctgca	
Oli_1154	fw	thiolase	*Homo sapiens*	V3D	GcggctgcaggtagATctgggccacctggG	P654
Oli_1155	rev	thiolase	*Homo sapiens*	V3D	aattcccaggtggcccagATctacctgcagccgctgca	
Oli_1156	fw	thiolase	*Homo sapiens*	V3K	GcggctgcaggtaAAgctgggccacctggG	P655
Oli_1157	rev	thiolase	*Homo sapiens*	V3K	aattcccaggtggcccagcTTtacctgcagccgctgca	
Oli_1194	fw	thiolase	*Homo sapiens*	V4D	GcggctgcaggtagtgGACggccacctggG	P670
Oli_1195	rev	thiolase	*Homo sapiens*	V4D	aattcccaggtggccGTCcactacctgcagccgctgca	
Oli_1198	fw	thiolase	*Homo sapiens*	L4P	GcggctgcaggtagtgcCgggccacctggG	P672
Oli_1199	rev	thiolase	*Homo sapiens*	L4P	aattcccaggtggcccGgcactacctgcagccgctgca	
Oli_1419	fw	thiolase	*Homo sapiens*	V4G	GcggctgcaAgtagtgGGGggccacctggG	P916
Oli_1420	rev	thiolase	*Homo sapiens*	V4G	aattcccaggtggccCCCcactacTtgcagccgctgca	
Oli_1421	fw	thiolase	*Homo sapiens*	V4Q	GcggctgcaAgtagtgcAAggccacctggG	P915
Oli_1422	rev	thiolase	*Homo sapiens*	V4Q	aattcccaggtggccTTGcactacTtgcagccgctgca	
Oli_1247	fw	thiolase	*Homo sapiens*	G5L	GcggctgcaggtagtgctgctgcacctggG	P729
Oli_1248	rev	thiolase	*Homo sapiens*	G5L	Aattcccaggtgcagcagcactacctgcagccgctgca	
Oli_1196	fw	thiolase	*Homo sapiens*	G5D	GcggctgcaggtagtgctgGATcacctggG	P671
Oli_1197	rev	thiolase	*Homo sapiens*	G5D	aattcccaggtgATCcagcactacctgcagccgctgca	
Oli_1682	fw	thiolase	*Homo sapiens*	RS1E	GGAActgcaagtagtgctgggccacctggG	P1280
Oli_1683	rev	thiolase	*Homo sapiens*	RS1E	AATTCccaggtggcccagcactacttgcagTTCCTGCA	

Putative PTS2 nonapeptides

Oligonucleotide number	Direction	Protein	Species	Acc.number	Oligo sequences	Plasmid-number
Oli_1135	fw	Pyridoxalkinase	*Homo sapiens*	NP_003672.1	GcgggtgctctccatacagagccacgtcgG	P631
Oli_1136	rev	Pyridoxalkinase	*Homo sapiens*		AATTCcgacgtggctctgtatggagagcacccgCTGCA	
Oli_1164	fw	Hypertension related protein	*Homo sapiens*	NP_001074472.1	GcgagtgagtttcttgggggcccagcttgG	P684
Oli_1165	rev	Hypertension related protein	*Homo sapiens*		aattccaagctgggcccccaagaaactcactcgctgca	
Oli_1235	fw	RAI17	*Homo sapiens*	NP_065071.1	GcgactgcagtgcatcaagcagcacttaaG	P717
Oli_1236	rev	RAI17	*Homo sapiens*		aattcttaagtgctgcttgatgcactgcagtcgctgca	
Oli_1237	fw	Nitrilase	*Homo sapiens*	NP_064587.1	GcgcttggccctcatccagcttcagattaG	P727
Oli_1238	rev	Nitrilase	*Homo sapiens*		aattctaatctgaagctggatgagggccaagcgctgca	
Oli_1239	fw	NP_997288.1	*Homo sapiens*	NP_997288.1	GcgggtgctgaccctgcagtgccagctcaG	P728
Oli_1240	rev	NP_997288.1	*Homo sapiens*		aattctgagctggcactgcagggtcagcacccgctgca	
Oli_1241	fw	XP_950375.1	*Homo sapiens*	XP_950375.1	GagattatacaaattgcattttcaattgaG	P731
Oli_1242	rev	XP_950375.1	*Homo sapiens*		aattctcaattgaaaatgcaatttgtataatctctgca	
Oli_1243	fw	-GTP binding dor	*Homo sapiens*	NP_056054.1	GaggatcgtaggactcctggctcaactgaG	P758
Oli_1244	rev	-GTP binding dor	*Homo sapiens*		Aattctcagttgagccaggagtcctacgatcctctgca	
Oli_1245	fw	domain containir	*Homo sapiens*	NP_116145.1	GcgtttgtgccacatcgccttccacgtgaG	P718
Oli_1246	rev	domain containir	*Homo sapiens*		Aattctcacgtggaaggcgatgtggcacaaacgctgca	
Oli_1255	fw	KvCHiP4	*Homo sapiens*	NP_671710.1	GagggtggaaagcatttcggctcagctgAG	P716
Oli_1256	rev	KvCHiP4	*Homo sapiens*		aattCtcagctgagccgaaatgctttccaccctCtgca	
Oli_1308	fw	C6ORF (LOC79969)	*Homo sapiens*	NP_079185.1	GcggatcacggtgctggaccagcacctggG	P742
Oli_1309	rev	C6ORF (LOC79969)	*Homo sapiens*		AATTCccaggtgctggtccagcaccgtgatccgCTGCA	
Oli_1310	fw	galactokinase 2 isoform 2	*Homo sapiens*	NP_001001556.1	GagagtcaacataataggagagcatatagG	P743
Oli_1311	rev	galactokinase 2 isoform 2	*Homo sapiens*		AATTCctatatgctctcctattatgttgactctCTGCA	
Oli_1312	fw	nucleoside-diphosphate kinase 7	*Homo sapiens*	NP_932076.1	GaaagtgaatgtcttttctcgacaactggG	P744
Oli_1313	rev	nucleoside-diphosphate kinase 7	*Homo sapiens*		AATTCccagttgtcgagaaaagacattcactttCTGCA	
Oli_1314	fw	DnaJ (Hsp40) homolog C13 isoform 2	*Homo sapiens*	NP_056083.2	GcgagtcttttcagttggaacccatgcagG	P745
Oli_1315	rev	DnaJ (Hsp40) homolog C13 isoform 2	*Homo sapiens*		AATTCctgcatgggttccaactgaaaagactcgCTGCA	
Oli_1352	fw	Transketolase	*Homo sapiens*	NP_001055.1	GcgcctgcgcatcagctccatccaggccGG	P853
Oli_1353	rev	Transketolase	*Homo sapiens*		AATTCCggcctggatggagctgatgcgcaggcgCTGCA	
Oli_1561	fw	TGF beta	*Homo sapiens*	NP_003229.1	gCGCATCGAGGCCATCCGCGGCCAGATgg	P1041
Oli_1562	rev	TGF beta	*Homo sapiens*		aattccATCTGGCCGCGGATGGCCTCGATGCGctgca	

Full length EGFP fusion proteins

Oligonucleotide number	Direction	Protein	Organism	Restriction enzyme	Oligonucleotide sequence	Plasmid-number
Oli_1297	fw	KvCHiP4	*Homo sapiens*	HindIII	gatctatttgacaagcttaatcacattttc	P795
Oli_1298	rev	KvCHiP4	*Homo sapiens*	PstI	ccctgcgtccctGCagacatgaatgtgag	
Oli_1299	fw	Glyoxalase domain containing protein 1	*Homo sapiens*	XhoI	ggtcgccgtcTCgagcgccatggc	P829
Oli_1300	rev	Glyoxalase domain containing protein 1	*Homo sapiens*	SmaI	cctgggcatcccgggcttcctggct	
Oli_1462	fw	RAI17	*Homo sapiens*	SalI	GTCGACatgaattctatggacaggcacatccag	P992
Oli_1330	rev	RAI17	*Homo sapiens*	HindIII	cgggtggccctcaagcttgttctcaaatag	
Oli_1605	rev	TGF_beta2	*Homo sapiens*	XhoI	CTCGAGatgcactactgtgtgctgag	P1121
Oli_1606	fw	TGF_beta2	*Homo sapiens*	BamHI	GGATCCgctgcatttgcaagactttac	

Mammalian Two Hybrid System

Oligonucleotide number	Direction	Protein	Organism	Mutation	Oligonucleotide sequence	Restriction enzyme	Plasmid-number
Oli_939	rev	PEX7	*Homo sapiens*		cggAATTcgggatgagtg	EcoRI	P1398
Oli_940	fw	PEX7	*Homo sapiens*		ctgtCgAcccagctactcag	SalI	
Oli_1670	rev	PEX7	*Homo sapiens*	E113R	gctcagAGGgtgtaCagtgttgattggagccaaacc	BsrGI	P1399
Oli_1671	fw	PEX7	*Homo sapiens*	E113R	actGtacacCCTctgagcgtgttctttatagacttgc	BsrGI	
Oli_1672	rev	PEX7	*Homo sapiens*	E200R	caggcaAGGatcCtgagttgtgactggtgtaaatac	BamHI	P1400
Oli_1673	fw	PEX7	*Homo sapiens*	E200R	actcaGgatCCTtgcctgatgtgcaggaatcacg	BamHI	
Oli_1703	rev	PEX7	*Homo sapiens*	E287R	catcatacaAGgtttacAtgtggtttagacttc		P1401
Oli_1704	fw	PEX7	*Homo sapiens*	E287R	ccacaTgtaaacCTtgtatgatgctccactg		
Oli_1719	rev	PTS2-tester			GTCGACcatcgATggcaagtctgcag	SalI	P1402
Oli_1720	fw	PTS2-tester			GTCGACgcggccgctttacttgtacagc	SalI	