J|A|C|S ARTICLES

Published on Web 02/23/2010

A Proteome-Wide Perspective on Peroxisome Targeting Signal 1(PTS1)-Pex5p Affinities

Debdip Ghosh and Jeremy M. Berg*

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, Maryland 20892

Received December 27, 2009; E-mail: bergj@mail.nih.gov

Abstract: Most proteins are targeted to the peroxisomal matrix by virtue of a peroxisomal targeting signal-1 (PTS1), a short carboxy-terminal sequence specifically recognized by the PTS1 receptor Pex5p. We had previously developed a model that allowed the estimation of the affinities of many PTS1 sequences within the human proteome for Pex5p that revealed a wide range of predicted affinities. We have now experimentally determined the affinities of the PTS1-containing peptides from 42 proteins from the human proteome for Pex5p and show that these range over 4 orders of magnitude. These affinities correlate reasonably well with the predicted values and are substantially more precise. In an attempt to provide a possible explanation for the wide range of PTS1−Pex5p affinities, we compared these affinities with mRNA levels (as a proxy for rates of protein production) of the genes encoding these proteins in 79 human tissues and cell types. We note that high affinity PTS1−Pex5p interactions tend to correspond to proteins encoded by genes expressed at relatively low levels, whereas lower affinity PTS1−Pex5p interactions tend to correspond to proteins encoded by genes exhibiting higher levels and wider ranges of expression. Further analysis revealed that these relationships are consistent with the notion that a relatively uniform pool of protein−Pex5p complexes is maintained for appropriate peroxisome assembly.

Introduction

Protein targeting depends critically on specific molecular interactions. Short stretches of polypeptide sequence within proteins interact with specific receptors that carry their cargo to particular intracellular compartments.[1] The thermodynamic strength of the interactions between these targeting signals and their receptors may be a primary determinant of protein targeting. One of the most well characterized protein targeting systems is that utilized by peroxisomes.[2−8] In humans, peroxisomes function primarily to catalyze fatty acid beta-oxidation, particularly of long and very long chain fatty acids, and to degrade the metabolic byproduct hydrogen peroxide.[9,10] Peroxisomes also house a number of other activities including those involved in purine and amino acid metabolism and the biosynthesis of cholesterol and plasmalogens. Proper peroxisome assembly is essential for human health as demonstrated by the occurrence of the peroxisome biogenesis disorders including the Zellweger spectrum disorders.[11−13] Most proteins destined for the peroxisomal matrix include a carboxy-terminal peroxisomal targeting signal-1 (PTS1) that is recognized by a specific receptor, Pex5p.[6,7,14,15] The PTS1 sequence was initially regarded as a tripeptide of the form Ser-Lys-Leu-COO$^-$ or conservative variations thereof.[3] Subsequent direct thermodynamic measurements, functional selection experiments, and structural studies have refined the definition of the PTS1 to include residues prior to the terminal tripeptide.[7,16−18] For example, the minimal targeting signal for the archetypal peroxisomal enzyme catalase is **Lys**-Ala-Asn-Leu-COO$^-$.[19] However, even the full catalase targeting signal has a substantially lower affinity for Pex5p than do many other PTS1s.[16]

We recently developed a model that allows the prediction of the Pex5p binding affinities of most of the PTS1s in the human

(1) Blobel, G. *ChemBioChem* **2000**, *1*, 86–102.
(2) Gould, S. G.; Keller, G. A.; Subramani, S. *J. Cell Biol.* **1987**, *105*, 2923–31.
(3) Gould, S. J.; Keller, G. A.; Hosken, N.; Wilkinson, J.; Subramani, S. *J. Cell Biol.* **1989**, *108*, 1657–64.
(4) Subramani, S. *J. Membr. Biol.* **1992**, *125*, 99–106.
(5) Teter, S. A.; Klionsky, D. J. *Trends Cell Biol.* **1999**, *9*, 428–31.
(6) Gatto, G. J., Jr.; Geisbrecht, B. V.; Gould, S. J.; Berg, J. M. *Nat. Struct. Biol.* **2000**, *7*, 1091–5.
(7) Stanley, W. A.; Filipp, F. V.; Kursula, P.; Schuller, N.; Erdmann, R.; Schliebs, W.; Sattler, M.; Wilmanns, M. *Mol. Cell* **2006**, *24*, 653–63.
(8) Stanley, W. A.; Fodor, K.; Marti-Renom, M. A.; Schliebs, W.; Wilmanns, M. *FEBS Lett.* **2007**, *581*, 4795–802.
(9) Lazarow, P. B. *J. Neuropathol. Exp. Neurol.* **1995**, *54*, 720–5.
(10) Schrader, M.; Fahimi, H. D. *Histochem. Cell Biol.* **2008**, *129*, 421–40.
(11) Gould, S. J.; Valle, D. *Trends Genet.* **2000**, *16*, 340–5.
(12) Fujiki, Y. *FEBS Lett.* **2000**, *476*, 42–6.
(13) Faust, P. L.; Banka, D.; Siriratsivawong, R.; Ng, V. G.; Wikander, T. M. *J. Inherit. Metab. Dis.* **2005**, *28*, 369–83.
(14) Fransen, M.; Brees, C.; Baumgart, E.; Vanhooren, J. C.; Baes, M.; Mannaerts, G. P.; Van Veldhoven, P. P. *J. Biol. Chem.* **1995**, *270*, 7731–6.
(15) Distel, B.; Erdmann, R.; Gould, S. J.; Blobel, G.; Crane, D. I.; Cregg, J. M.; Dodt, G.; Fujiki, Y.; Goodman, J. M.; Just, W. W.; Kiel, J. A.; Kunau, W. H.; Lazarow, P. B.; Mannaerts, G. P.; Moser, H. W.; Osumi, T.; Rachubinski, R. A.; Roscher, A.; Subramani, S.; Tabak, H. F.; Tsukamoto, T.; Valle, D.; van der Klei, I.; van Veldhoven, P. P.; Veenhuis, M. *J. Cell Biol.* **1996**, *135*, 1–3.
(16) Maynard, E. L.; Gatto, G. J., Jr.; Berg, J. M. *Proteins* **2004**, *55*, 856–61.
(17) Lametschwandtner, G.; Brocard, C.; Fransen, M.; Van Veldhoven, P.; Berger, J.; Hartig, A. *J. Biol. Chem.* **1998**, *273*, 33635–43.
(18) Brocard, C.; Hartig, A. *Biochim. Biophys. Acta* **2006**, *1763*, 1565–73.
(19) Purdue, P. E.; Lazarow, P. B. *J. Cell Biol.* **1996**, *134*, 849–62.

proteome.[20] This model predicted a range exceeding 3 orders of magnitude in Pex5p dissociation constants for apparently functional PTS1 signals within the human proteome. The model was based on an analysis of the results of Lametschwandtner et al. who reported the sequences of 63 peptides selected for binding to human Pex5p from a randomized library of carboxyl-terminal peptides through the use of a yeast two-hybrid screen.[17] These sequences were analyzed to yield frequencies for each amino acid as a function of position from the carboxyl terminus. With the assumptions that each position is independent and that these frequencies (corrected for codon frequency within the library) correspond to a Boltzmann distributed set, these frequencies can be converted into relative free energies through the equation $\Delta\Delta G_{ij} = -RT \ln(c_i f_{ij})$ where f_{ij} is the frequency of recovery of amino acid i in position j and c_i is a correction factor for the number of codons for amino acid i in the NN(G,T) randomized pool. While this model provided some useful insights into Pex5p including the biological impact of mutations,[20] the underlying assumptions are untested and the predicted affinities are relatively imprecise.

Here, we report the determination of a wide range of Pex5−PTS1 peptide affinities and compare the experimentally determined affinities with those previously predicted. The wide range of Pex5p affinities was confirmed, and the experimental results correlated reasonably well with those predicted. In addition, in an initial attempt to understand a possible basis for the observed range of affinities, we compared the experimental affinities with the levels of peroxisomal protein mRNA expression as a proxy for relative rates of protein synthesis in a wide range of human tissues and cell lines. Analysis of the combination of the Pex5p−PTS1 affinity data and these mRNA levels revealed that the calculated concentrations of Pex5p−protein complexes are relatively uniform in most tissues, suggesting how the PTS1 signals and protein expression levels may have coevolved to facilitate appropriate peroxisome assembly.

Experimental Section

Materials. LB medium, tryptone, and yeast extract were obtained from Becton Dickinson. Isopropylthiogalactoside (IPTG) was obtained from M.P. Biomedical LLC. Competent *Escherichia coli* BL21(DE3) cells were obtained from Bioline. Complete protease inhibitor cocktail tablets were obtained from Roche, and DNase and RNase from New England Biolabs. Kanamycin, Tris-HCl, ethylenediaminetetraacetic acid (EDTA), benzamidine-HCl, sodium fluoride, dithiothreitol, ammonium sulfate, and sodium chloride were obtained from Sigma-Aldrich. For the peptide synthesis, Fmoc amino acids, Rink-amide resins, and *O*-(benzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-hydroxymethyluronium tetrafluoroborate (HBTU) were obtained from Anaspec Inc. Piperidine, dimethyl formamide, and diisopropylethylamine were obtained from Sigma-Aldrich. Lissamine sulfonyl chloride was obtained from Molecular Probes.

Protein Expression and Purification. The PTS1 binding domain of human Pex5p was expressed and purified as described previously.[6] A cDNA encoding amino acids 235−602 of human Pex5p was constructed by the polymerase chain reaction, cloned into the pT7 plasmid, and transformed into *Escherichia coli* BL21(DE3) cells. Cells were grown in rich media, 2YT (2YT is 16 g tryptone, 10 g yeast extract and 5 g NaCl per liter), to an OD_{600} of 1.0 at 37 °C, induced with 1 mM IPTG, and grown for an additional 4 h at 25 °C. Cell lysis was performed using 0.1x complete protease inhibitor cocktail and 0.02 mg/mL DNase and RNase followed by two cycles of French press. Pex5p was then purified by sequential column chromatography steps over Q-sepharose (Buffer 50 mM Tris, pH 7.5, 2 mM EDTA, 5 mM benzamidine-HCl, 5 mM NaF, 1 mM DTT), C-4 butyl-sepharose (buffer 50 mM Tris, pH 7.5, 2 mM EDTA, 5 mM benzamidine-HCl, 5 mM NaF, 1 mM DTT, 0.25 M ammonium sulfate, and Superdex 75 (buffer 50 mM Tris, 200 mM NaCl).

Peptide Synthesis. Peptides were synthesized by manual solid-phase methods using 9-fluorenylmethoxycarbonyl (Fmoc) chemistry with *O*-(benzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-hydroxymethyluronium tetrafluoroborate (HBTU) activation. The Fmoc-protecting groups were removed using 20% piperidine. To introduce the lissamine fluorescent label, 100 mg of the Fmoc cleaved, side chain protected peptide resin were suspended in 2 mL of dimethylformamide (DMF) with a 4-fold molar excess of lissamine sulfonyl chloride (Molecular Probes) and a 10-fold molar excess of diisopropylethylamine (DIEA). The reaction was kept in the dark on ice for 2 h with occasional stirring, after which the resin was washed with DMF and dichloromethane and cleaved. The peptides were purified by reversed phase high performance liquid chromatography, and their identity and purity were assessed by mass spectrometry.

Anisotropy Binding Studies. Anisotropy assays were performed using the peptide YKSKL labeled on the amino terminus with lissamine (Liss-YKSKL) as described previously.[6] The fraction of the fluorophore bound (f_B) for a given anisotropy value (r) and quantum yield ratio (Q) were computed using the expression:

$$f_B = \frac{r - r_{free}}{(r_{bound} - r)Q + (r - r_{free})}$$

Liss-YKSKL was used at a concentration of 100 nM with parameters $r_{bound} = 0.3$, $r_{free} = 0.075$, K_d (labeled peptide) = 2.7 nM, and $Q = 0.8 \pm 0.1$.

The value of 0.3 for r_{bound} and 0.075 for r_{free} were extracted from anisotropy experiments demonstrating binding of Pex5p to the fluorophore labeled peptide Liss-YKSKL. The corresponding trace is provided in the Supporting Information. This demonstrates that the free unbound fluorophore, Liss-YKSKL, has an anisotropy of 0.075 (r_{free}). Titration of Pex5p leads to an increase in the anisotropy up to a final value of 0.3 (r_{bound}). The corresponding dissociation constant extracted was 2.7 nM (K_d).

The expression for fraction bound (f_B) needs to be modified reflecting changes in the fluorescence intensity during the binding experiments. In this case, a correction factor, Q, representing the quantum yield ratio of the bound to the free form, is introduced. Q can be estimated by the ratio of the intensities of the bound to the free species. Prior to all curve fitting in these experiments, this correction was applied since a decrease in the fluorescence intensity was consistently observed upon ligand binding ($Q = 0.8$).

Calculation of Relative Pex5p-protein Complex Concentrations. Pex5p is expressed at relatively constant and low levels in all tissues (based on mRNA levels[22]) and, hence, is limiting in determining the distribution of species within the Pex5p−protein complex pool. The ratio of the concentrations of any two complexes Pex5p−protein1 and Pex5p−protein2 can be calculated from [Pex5p−protein1]/[Pex5p−protein2] = ([protein1][K_{d2}])/([protein2][K_{d1}]) where K_{d1} and K_{d2} are the dissociation constants for Pex5p−protein1 and Pex5p−protein2, respectively.

Calculation of Nonuniformity Scores and Estimation of Statistical Significance. The nonuniformity score is defined by

$$Q = \sum_{i=1}^{n} (f_i - f_u)^2$$

where f_i is the fraction of the i^{th} complex and f_u is the fraction of each complex within a uniform population ($f_u = 1/42 = 0.024$ for $n = 42$ proteins). The parameter Q is defined such that Q is zero if all complexes are estimated to be present at the same concentration and increases as the estimated distribution of complexes become less uniform. Q was calculated for each tissue and cell type. To

(20) Maynard, E. L.; Berg, J. M. *J. Mol. Biol.* **2007**, *368*, 1259–66.

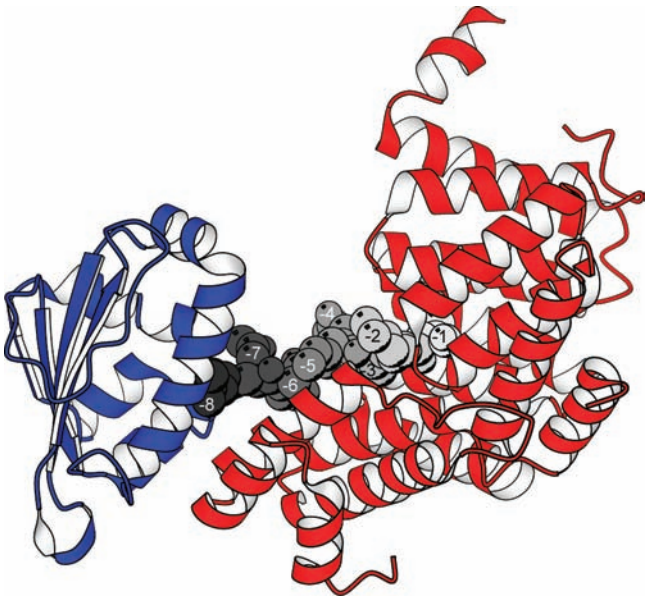

Figure 1. Structure of the Pex5p-Sterol carrier protein-2 complex (16). The PTS1-binding domain of human Pex5p is shown in red, and the sterol carrier protein-2 domain is shown in blue. The eight amino acids corresponding to the PTS1 at the carboxyl terminus of the sterol carrier protein-2 domain are shown as space-filling models in shades of gray and are labeled −1 through −8. Note that there are only minor contacts between Pex5p and the sterol carrier protein domain not involving these residues.

estimate the likelihood that a given nonuniformity score occurred by chance, the dissociation constants were randomly reassigned to mRNA expression levels and the resulting nonuniformity score was calculated. This was repeated 100 times for each tissue and cell type to build up a score distribution. The position of the nonuniformity score for the actual set of mRNA expression levels and dissociation within each distribution was determined to estimate the likelihood that this score occurred by chance.

Results and Discussion

A total of 42 peptides corresponding to the carboxyl-terminal 8 amino acids (with an amino-terminal Tyr added to assist peptide quantitation) for each of the proteins within the human proteome known to be targeted to peroxisomes by virtue of a PTS1 sequence were prepared by solid-phase peptide synthesis. Peptides of this length were chosen for study since these include the carboxyl terminal residues that contact Pex5p in the crystal structure of human Pex5p-Sterol carrier protein-2 complex[7] (Figure 1). Furthermore, a six amino acid peptide has been shown to have an affinity within a factor of 6 of that for the corresponding intact protein domain.[7] The use of the nine amino acid peptides in the present study reduced this difference to a factor of 4. No additional thermodynamic or structural studies of Pex5p complexes with intact proteins have been reported, but crystal structures of other proteins that include PTS1 signals reveal that these occur in poorly structured extensions from the protein core, supporting the notion that thermodynamic properties determined for the peptides used in the present study should provide reasonable approximations for values expected for the intact proteins.

The affinities of these peptides for the PTS1-binding domain of human Pex5p were determined by a fluorescence anisotropy-based assay[6,16] and are shown in Table 1 (see also Supporting Information). The free energies associated with these affinities correlate reasonably well with those previously predicted ($r = 0.48$) (Figure 2). Furthermore, the values encompass a range (with dissociation constants from approximately 1 nM to >1 μM) quite similar to that anticipated. Consistent with earlier studies,[16] the PTS1 peptide from acyl-CoA oxidase 1 is found to have a high affinity (5.6 nM) whereas the PTS1 peptide from catalase is found to have a relatively low affinity (1.2 μM).

A total of 42 peptides were studied experimentally whereas affinities had been estimated previously[20] for only 30 PTS1 sequences, due to the limitation of the sequences represented within the yeast two-hybrid data set on which the affinity estimation model was based.[17] Furthermore, the experimentally determined affinities and corresponding standard free energies of binding are substantially more precise. Since the affinity estimation model was based on only 63 sequences, uncertainties due to limited sampling are substantial, leading to uncertainties greater than 0.5 kcal/mol compared with the experimental uncertainties that are typically 0.1−0.2 kcal/mol.

The experimentally determined affinities tend to be lower than those previously estimated with a mean difference of −1.2 kcal/mol. Given the assumptions that were used in the affinity estimation model such as the interpretation of populations deduced from the yeast two-hybrid experiments in terms of Boltzmann distribution and that preferences in different positions are strictly additive, there are many possible explanations for these differences. Further studies with peptides specifically selected to probe these different possible explanations will be required to provide additional clarity about the relative importance of these effects. Despite these issues, the reasonable agreement in overall trends supports the general approach used to estimate affinities while the experimentally determined values provide a solid foundation for additional consideration of the significance of these affinities.

How can one account for these Pex5p−PTS1 peptide affinities? Why do the PTS1 signals from some proteins have substantially higher affinities for Pex5p than do others? One simple hypothesis would suggest that these affinities reflect the efficiency of peroxisomal protein targeting. However, examination of the limited data available about relative protein abundances for a small number of peroxisomal proteins in liver,[21] the tissue in which peroxisomes are most abundant, reveals that proteins with low affinity PTS1 signals such as catalase are relatively abundant (4.0 μg/mg of total protein) compared with acyl-CoA oxidase 1 with a high affinity (0.3 μg/mg of total protein).

An alternative hypothesis is that Pex5p−protein affinities and rates of protein production in the cytoplasm prior to targeting are related so that the pool of Pex5p−protein complexes is relatively balanced with similar concentrations of each protein within the pool. Of course, this balance need not be perfect as the requirements for different enzymes and proteins within the peroxisome are almost certainly different. To test this hypothesis, information regarding the rates of protein production for many proteins destined for the peroxisome matrix is required. Extensive data regarding mRNA levels from a wide range of human tissues are available.[22] Such mRNA levels should serve as a reasonable proxy for rates of protein production. Note that, in several systems, such mRNA levels correlate only moderately with protein abundances,[23–25] but protein abundances reflect the

(21) Hashimoto, T. *Neurochem. Res.* **1999**, *24*, 551–63.
(22) Su, A. I.; Wiltshire, T.; Batalov, S.; Lapp, H.; Ching, K. A.; Block, D.; Zhang, J.; Soden, R.; Hayakawa, M.; Kreiman, G.; Cooke, M. P.; Walker, J. R.; Hogenesch, J. B. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6062–7.

Table 1. Dissociation Constants of PTS1 Peptides from the Human Proteome for Human Pex5p[a]

Protein (Gene Name)	Sequence	K_d, nM (SD)	$\Delta G°$, kcal/mol
Acyl-CoA oxidase 1 (ACOX1)	YHKSLQSKL	5.6(1.0)	−11.2
Acyl-CoA oxidase 2 (ACOX2)	YLQSWRSKL	257(5)	−9.0
Acyl-CoA oxidase 3 (ACOX3)	YVGSLKSKL	1.6((1.0)	−12.0
L-Bifunctional enzyme (EHHADH)	YAGSPSSKL	1096(156)	−8.1
D-Bifunctional enzyme (HSD17B4)	YILKDYAKL	195(62)	−9.1
Sterol carrier protein 2 (SCP2)	YLQPGNAKL	380(26)	−8.8
3,2-*trans*-enoyl-CoA isomerase (PECI)	YFLSRKSKL	14(3)	−10.7
2,4-dienoyl-CoA reductase 2 (DECR2)	YFASFSAKL	60(2)	−9.8
Δ-3,5-Δ-2,4-Dienoyl-CoA isomerase (ECH1)	YKTVTFSKL	120(32)	−9.4
trans-2-enoyl-CoA reductase (PECR)	YTFKEKAKL	21(3)	−10.5
2-Hydroxyphytanoyl-CoA lyase (HPCL2)	YHWLTRSNM	4467(1289)	−7.3
Methylacyl-CoA racemase (AMACR)	YSNKVKASL	1023(295)	−8.2
Carnitine acyltransferase 1 (CRAT)	YQSHPRAKL	59(7)	−9.9
Carnitine octanoyl transferase (CROT)	YQLMNSTHL	590(46)	−8.5
Acyl-CoA thioesterase 2 (PTE1)	YPQVSESKL	316(129)	−8.9
Acyl-CoA thioesterase 1B (ZAP128)	YEGTIPSKV	646(51)	−8.4
Catalase (CAT)	YAAREKANL	1230(314)	−8.1
Peroxiredoxin (PRDX5)	YAPNIISQL	128(60)	−9.4
D-amino acid oxidase (DAO)	YSRMPPSHL	372(182)	−8.8
D-aspartate oxidase (DDO)	YTPIPKSNL	148(46)	−9.3
Hydroxyacid oxidase 1 (HAO1)	YNPLAVSKI	26 303(1348)	−6.2
Hydroxyacid oxidase 2 (HAO2)	YNLVQFSRL	93(23)	−9.6
Very-long-chain acyl-CoA synthetase (SLC27A2)	YISAKTLKL	463(58)	−8.6
Dihydroxyacetone phosphate acyltransferase (GNPAT)	YGKPATAKL	288(85)	−8.9
Isocitrate dehydrogenase (IDH1)	YIKLAQAKL	186(56)	−9.2
Alanine-glycolate aminotransferase (AGXT)	YQHCPKKKL	13 490(2340)	−6.6
Pipecolic acid oxidase (PIPOX)	YPSLGKAHL	240(52)	−9.0
Epoxide hydrolase 2 (EPHX2)	YNPPVVSKM	33(7)	−10.2
Glutathione S-transferase class κ (GSTK1)	YPPAVNARL	5012(515)	−7.2
Polyamine oxidase (PAOX)	YVQQPRPRL	2.6(0.4)	−11.7
Malonyl-CoA decarboxylase (MLYCD)	YQFQKNSKL	562(59)	−8.5
3-Hydroxy-3-methylglutaryl-CoA lyase (HMGCL)	YVAQATCKL	1995(292)	−7.8
Nudix hydrolase specific for CoA (NUDT7)	YHKKATSRL	79(11)	−9.7
Insulysin (IDE)	YINFMAAKL	234(62)	−9.0
Lon protease (LONP)	YPGLLNSKL	22(3)	−10.4
Nudix hydrolase specific for CoA (NUDT19)	YSVVSKSHL	162(23)	−9.3
Bile acid Coenzyme A (BAAT)	YIPDVTSQL	113(12)	−9.5
Dehydrogenase/reductase SDR family member 4 (DHRS4)	YGGGTPSRL	98(9)	−9.6
Fission 1 (mitochondrial outer membrane) (FIS1)	YLAVSKSKS	211(17)	−9.1
Nitric oxide synthase 2A (INOS)	YSSLEMSAL	189(16)	−9.2
Peroxisomal membrane Mpv17 protein (MPV17)	YLSWKAHRL	343(22)	−8.8
Peroxisomal acyl-CoA thioesterase 2B (PTE2b)	YQKTAVPKL	623(48)	−8.5

[a] Standard deviations for each dissociation constant, K_d, are given in parentheses.

impact of the relative rates of protein degradation in addition to the rates of protein production.

Messenger RNA levels (determined by RNA microarray analysis) are available for most human proteins levels from a wide range of tissues.[22] The expression results for the 42 peroxisomal proteins are shown in Figure 3 (also see Supporting Information). Overall, mRNA levels for peroxisomal proteins are highest in the liver (column 21) followed by the kidney (column 20), consistent with the known tissue distribution of peroxisomes.

The PTS1 peptide−Pex5p affinities can be compared with these mRNA levels. The genes for some proteins such as acyl-CoA oxidases 1, 2, and 3, the enzymes that catalyze the rate-determining steps in peroxisomal fatty acid beta oxidation[26] (top three rows in Figure 1), are expressed at relatively low levels in all tissues. In contrast, genes encoding proteins such as catalase (CAT, row 17) and peroxiredoxin 5 (PRDX5, row 18), key peroxide metabolizing enzymes, are expressed at much higher levels. The levels of expression for all 42 proteins in 79 tissues and cell types are compared with the values of the Pex5p−protein dissociation constants in Figure 4. These data reveal that proteins with PTS1s with high affinities for Pex5p tend to be expressed at relatively low levels, whereas proteins with PTS1 signals with more modest affinities are expressed with higher maximum levels and with broader ranges of expression. Overall, the correlation is only modest with a correlation coefficient of $r = 0.12$ between the logarithm of the dissociation constant and the logarithm of the median expression level. The correlation is somewhat better for expression levels in the liver with $r = 0.29$ where peroxisomal protein expression is higher. Furthermore, the correlation is better for proteins with PTS1s with relatively high affinities of Pex5p (corresponding to data points toward the left side of Figure 4). For example, for the 15 proteins with PTS1s with the highest affinity for Pex5p, $r = 0.44$ for median expression levels and $r = 0.49$ for liver expression levels. These correlation coefficients should be considered in the context that one would not anticipate a very strong correlation due to (among other factors) (i)

(23) Greenbaum, D.; Colangelo, C.; Williams, K.; Gerstein, M. *Genome Biol.* **2003**, *4*, 117.
(24) Greenbaum, D.; Jansen, R.; Gerstein, M. *Bioinformatics* **2002**, *18*, 585–96.
(25) Ideker, T.; Thorsson, V.; Ranish, J. A.; Christmas, R.; Buhler, J.; Eng, J. K.; Bumgarner, R.; Goodlett, D. R.; Aebersold, R.; Hood, L. *Science* **2001**, *292*, 929–34.
(26) Aoyama, T.; Souri, M.; Kamijo, T.; Ushikubo, S.; Hashimoto, T. *Biochem. Biophys. Res. Commun.* **1994**, *201*, 1541–7.

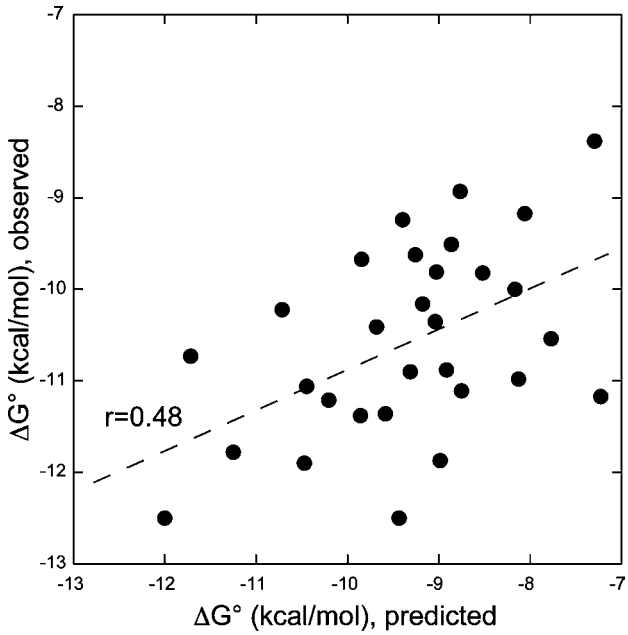

Figure 2. A comparison of the experimentally determined standard free energies of binding to human Pex5p for 30 PTS1 peptides with those previously predicted.[20] The line of best fit and correlation coefficient are shown.

limitations in the expression data which come generally from entire tissues rather than pure cell types, (ii) errors inherent in microarray data, and (iii) the relatively simplistic notion that all peroxisomal matrix proteins should be targeted with equal efficiency.

As an alternative approach to examining the relationship between the Pex5p−PTS1 affinities and expression levels, the expression data can be used to test more directly the hypothesis that the population of Pex5p−protein complexes is relatively uniform. An estimate of the relative population of each Pex5p−protein complex can be calculated from the relative mRNA expression level (as an estimate of the initial protein concentration) and the corresponding Pex5p−protein complex dissociation constant with the assumption (supported by known levels of Pex5p mRNA) that Pex5p is present at a concentration much lower than the total concentration of potential binding partners. Calculated percentages of each protein within the pool of Pex5p−protein complexes for two tissues are shown in Figure 5A.

If the dissociation constants and expression levels were perfectly correlated, then all Pex5p−protein complexes would be present at the same concentration. Again, there is no reason for a perfect balance within the Pex5p−protein complex pool as differences in protein activity and metabolic requirements will lead to a different optimal level for each protein. However,

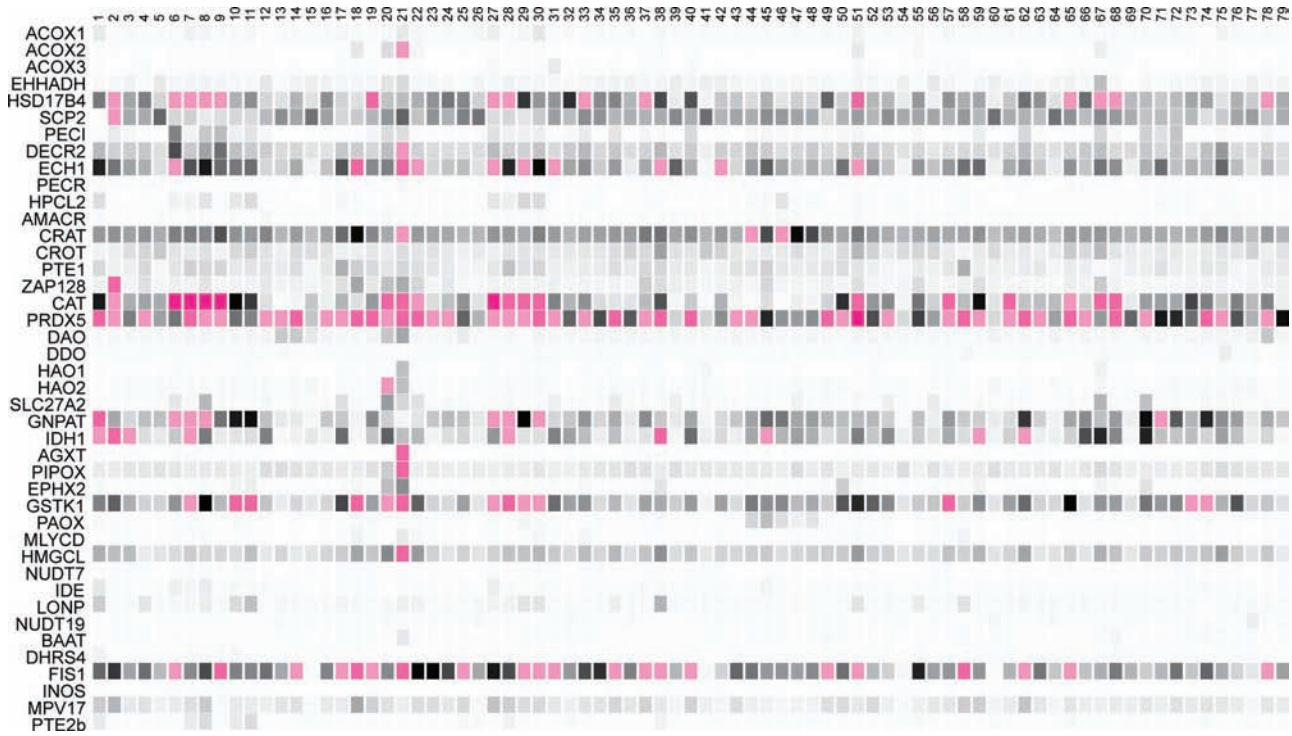

Figure 3. A plot of the relative mRNA expression levels of 42 genes for peroxisomal proteins in 79 tissues and cell types from ref 22. The expression levels are shown in gray scale from 0−100 (in arbitrary units) with higher levels shown in color: 100−200 (light pink), 200−500 (medium pink), >500 (dark pink). Each row corresponds to one gene indicated while the columns are numbered from 1−79 as follows: 1, 721 B-lymphoblasts; 2, adipocytes; 3, adrenal cortex; 4, amygdala; 5, appendix; 6, bone marrow $CD105^+$ endothelial cells; 7, bone marrow $CD33^+$ myeloid cells; 8, bone marrow $CD34^+$ cells; 9, bone marrow $CD71^+$ early erythroid cells; 10, $CD4^+$ T cells; 11, $CD8^+$ T cells; 12, cardiac myocytes; 13, cerebellum; 14, cerebellum peduncle; 15, ciliary ganglion; 16, cingulate cortex; 17, colorectal adenocarcinoma; 18, heart; 19, hypothalamus; 20, kidney; 21, liver; 22, lung; 23, medulla oblongata; 24, occipital lobe; 25, olfactory bulb; 26, ovary; 27, peripheral blood $BDCA4^+$ dendritic cells; 28, peripheral blood $CD14^+$ monocytes; 29, peripheral blood $CD19^+$ B cells, 30, peripheral blood $CD56^+$ NK cells; 31, placenta; 32, pancreas; 33, pancreatic islets; 34, parietal lobe; 35, pituitary; 36, pons; 37, prefrontal cortex; 38, prostate; 39, skeletal muscle; 40, smooth muscle; 41, superior cervical ganglion; 42, tongue; 43, temporal lobe; 44, testis; 45, testis germ cells; 46, testis interstitium; 47, testis Leydig cell; 48, testis seminiferous tubule; 49, thalamus; 50, thymus; 51, thyroid; 52, tonsil; 53, trachea; 54, trigeminal ganglion; 55, uterus; 56, uterus corpus; 57, whole blood; 58, whole brain; 59, adrenal gland; 60, atrioventricular node; 61, bone marrow; 62, bronchial epithelial cells; 63, caudate nucleus; 64, dorsal root ganglia; 65, fetal thyroid; 66, fetal brain; 67, fetal liver; 68, fetal lung; 69, globus pallidus; 70, leukemia, chronic myelogenous (k562); 71, leukemia, lymphoblastic (molt4); 72, leukemia, promyelocytic (hl60); 73, lymph node; 74, lymphoma, Burkitt's Daudi; 75, lymphoma, Burkitt's Raji; 76, salivary gland; 77, skin; 78, spinal cord; 79, subthalamic nucleus.

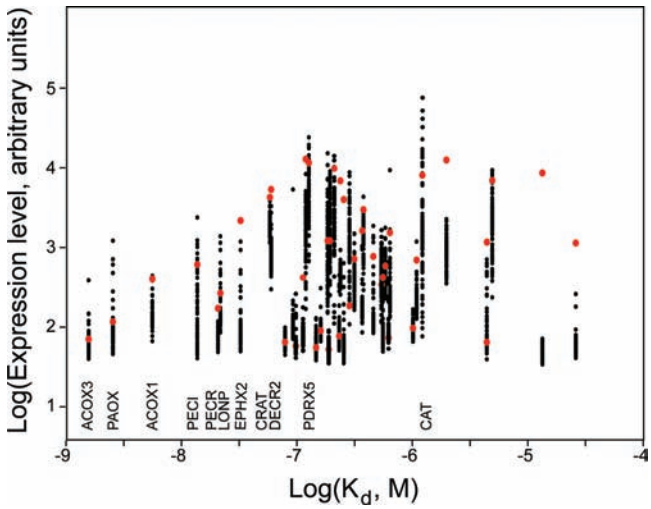

Figure 4. A plot of the log of mRNA expression level from 79 tissues and cell types versus the log of the dissociation constant for the Pex5p−PTS1 complex for 42 peroxisomal proteins with selected genes indicated. The expression levels for liver are shown in red.

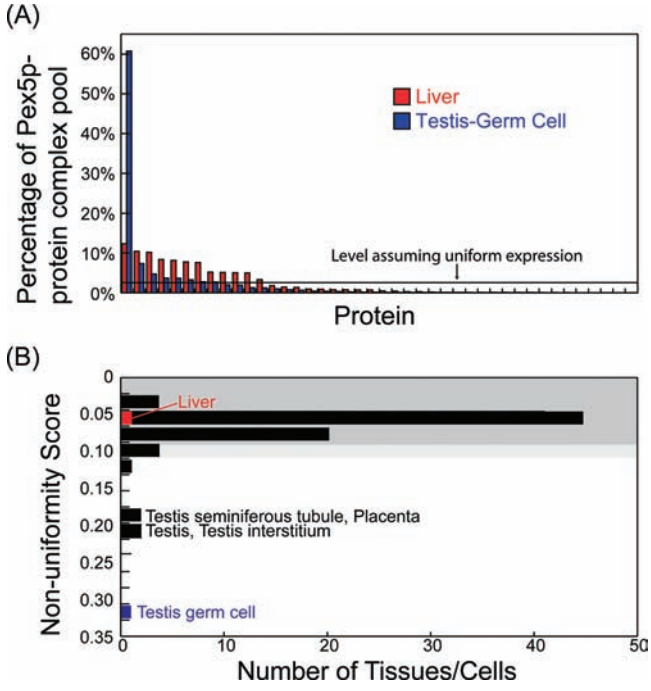

Figure 5. Calculated distribution of proteins within the pool of Pex5p−protein complexes. (A) The distributions of Pex5p−protein complexes for liver (red) and testis germ cells (blue) based on the expression levels and Pex5p−PTS1 dissociation constants. The line shows the level (2.4%) if all Pex5p−protein complexes were equally represented. (B) The distribution of nonuniformity scores for the 79 tissues and cell types. The darker gray region corresponds to nonuniformity scores estimated to have probabilities of occurring by chance of less than 5%, while the lighter gray region corresponds to nonuniformity scores estimated to have probabilities of occurring by chance of less than 10% based on jumbling analyses. The nonuniformity scores for liver (red) and testis germ cells (blue) are indicated.

the presence of a large imbalance with one or a few proteins dominating the complex pool could significantly affect peroxisome assembly and function. For the 42 proteins under consideration, a perfectly balanced pool would have each complex represented at 2.4% of the pool. Based on the liver expression levels, the most prevalent complex is calculated to represent 13% of the pool and more than half of the complexes are predicted to be within a factor of 5 of the uniform level. In contrast, for testis germ cells, the most prevalent complex dominates the distribution, representing 61% of the Pex5p−protein complex pool despite representing only 3.5% of the pool of mRNA.

The deviation of the overall population from a hypothetical uniform population can be quantified by the nonuniformity score Q defined by

$$Q = \sum_{i=1}^{n} (f_i - f_u)^2$$

where f_i is the fraction of the i^{th} complex and f_u is the fraction of each complex within a uniform population ($f_u = 1/42 = 0.024$ for $n = 42$ proteins). For a uniform population, $Q = 0$. For the liver expression pattern, $Q = 0.050$ while, for the testis germ cell expression pattern, $Q = 0.359$. The nonuniformity scores, calculated for each of the 79 tissue and cell expression patterns, are shown in Figure 5B. The nonuniformity scores for all but five of the expression patterns are clustered between 0.047 and 0.135. An advantage of the use of the nonuniformity score is that the statistical significance of these results can be estimated by repeatedly jumbling the linkages between gene expression levels and Pex5p−protein dissociation constants to build up a population of nonuniformity scores. For 61 of the 79 tissues and cell types, the observed nonuniformity scores are predicted to occur by chance less than 5% of the time, while for an additional 9 the nonuniformity scores are predicted to occur by chance less that 10% of the time.

Five nonuniformity scores are separated from the remainder. The expression pattern with the highest nonuniformity score is from testis germ cells. The high nonuniformity score is due to the high calculated percentage (61%) of the Pex5p−polyamine oxidase (PAOX) complex. The high expression of this enzyme also accounts for the relatively high nonuniformity scores for testis interstitium, testis, and testis seminiferous tubules. While the activity levels of polyamine oxidase in peroxisomes from these testicular tissues do not appear to have been reported, the occurrence of specialized polyamine metabolic pathways in these tissues would not be surprising given the key roles that polyamines play in sperm.[27] Note that histological staining of peroxisomes for specific activities has demonstrated that peroxisomes from different tissues can have significantly different activities.[28] The remaining high nonuniformity score is for placenta due to the relatively high calculated percentage (45%) for the Pex5p-acyl-CoA oxidase 3 (ACOX3). Human acyl-CoA oxidase 3 gene has been hypothesized to be preferentially expressed at specific developmental stages although these stages do not appear to have been identified.[29]

These calculations relate to the composition of the Pex5p−protein complex population. Given the role of these complexes in protein targeting, the fraction of each complex may reflect the relative likelihood of targeting a specific protein to a newly forming or dynamic peroxisome. Thus, these fractions, determined by the relationships between the levels of protein expression and the Pex5p−PTS1 dissociation constants, represent the priorities for targeting particular proteins and may contribute to the assembly of peroxisomes with appropriate basal

(27) Jänne, J.; Alhonen, L.; Pietilä, M.; Keinänen, T. A. *Eur. J. Biochem.* **2004**, *271*, 877–94.
(28) Van den Munckhof, R. J. *Histochem. J.* **1996**, *28*, 401–29.
(29) Vanhooren, J. C.; Marynen, P.; Mannaerts, G. P.; Van Veldhoven, P. P. *Biochem. J.* **1997**, *325* (Pt 3), 593–9.

activities as well as peroxisomes specialized for particular functions in specific tissues.

Summary and Conclusions

We have determined the affinities for human Pex5p of 42 peptides that correspond to the PTS1 peroxisome targeting signals from all of the proteins in the human proteome known or suspected to be targeted by a PTS1 signal. These peptides show a wide range of Pex5p affinities with dissociation constants ranging from 1.6 nM through greater than 25 mM. These experimentally determined standard free energies of binding correlate reasonably well with those predicted previously based on quantitative analysis of the results of a yeast two-hybrid screen for Pex5p-binding peptide sequences.

PTS1 sequences occur on carboxyl-terminal extensions that do not participate in many interactions with the remainder of the protein.[30] Thus, PTS1s are relatively unconstrained from an evolutionary perspective as each PTS1 sequence can evolve to optimize peroxisomal targeting in the context of all other PTS1 sequences and the expression levels of other peroxisomal proteins. In general, it appears that these PTS1 sequences and gene expression levels may have evolved to provide a relatively uniform population of Pex5p−protein complexes and, hence, similar probabilities of initial targeting for most peroxisomal proteins while still allowing a substantial range of expression for some peroxisomal proteins. This may be particularly crucial since the expression of some, but not all, peroxisomal protein genes is significantly induced by compounds, peroxisome proliferators, that trigger an increase in the number of peroxisomes.[31] The thermodynamic logic that we propose makes testable predictions for this and other aspects of peroxisomal function and evolution.

Acknowledgment. This work was supported by the intramural research program of the National Institute of Diabetes and Digestive and Kidney Diseases and the National Institute of General Medical Sciences. We thank Drs. Barbara Amann, Douglas Barrick, and Gregory J. Gatto, Jr. for useful discussions.

Supporting Information Available: The results of titrating Pex5p with lissamine-YQSKL, the results of competition titrations of the 42 peptides with Pex5p-lissamine-YQSKL, correlations between the observed PST1 peptides affinities and those previously predicted in ref 19, and the gene expression levels for the 42 genes in 79 tissues and cell types are available. This material is available free of charge via the Internet at http://pubs.acs.org.

JA9109049

(30) Eisenhaber, B.; Eisenhaber, F. *Curr. Protein Pept. Sci.* **2007**, *8*, 197–203.

(31) Issemann, I.; Green, S. *Nature* **1990**, *347*, 645–50.