Abstract

Genomes provide us with a blue print for the potential of a cell. However, the activity of a cell is expressed in its proteome. Full understanding of the complexity of cells demands a comprehensive view of the proteome; its interactions, activity states and organization. Comprehensive proteomic approaches applied to peroxisomes have yielded new insights into the organelle and its dynamic interplay with other cellular structures. As technologies and methodologies improve, proteomics hold the promise for new discoveries of peroxisome function and a full description of this dynamic organelle.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Peroxisome; Proteome; Mass spectrometry; Protein quantification; Protein localization; In silico prediction

1. Introduction

Ideally, proteomics studies should provide a comprehensive list of the relevant protein players (in the system of focus), their abundance (in absolute terms), their activity states (reflected in post translational states and interactions) and a quantitative measure of their dynamic changes. Current technologies provide some foothold into reaching these goals, but much remains to be accomplished. New advances in mass spectrometry, in silico methodologies, and biochemical and reporter assays have led to large data sets of the proteins expressed in cells, their post translational states and their membership in organelles and complexes. Here we focus on the impact these methods have had on the fields of peroxisome biogenesis and function. Arguably, focusing on a dynamic, cellular subsystem like the peroxisome not only allows us to have a realistic goal of obtaining a comprehensive list of players, but it also allows us to develop methods to understand their dynamic interplay.

2. In silico proteomics

Experimental proteomics has become intertwined with computational approaches to both de-convolute data generated from mass spectrometry experiments, and to combine and analyze large scale datasets to make predictions regarding the potential nature of the proteome. For the peroxisome field, many of these studies make use of comparative genomics, the end result being a prediction regarding the potential status of a given protein as a peroxisomal protein. Cells utilize two known peptide signals, PTS1 (peroxisomal targeting signal 1) and PTS2, to localize proteins to the peroxisomal matrix. PTS1 is a three amino acid, variable motif found at the extreme C-terminus of peroxisomal matrix proteins, that is generally described by the consensus sequence serine–lysine–leucine (SKL) [1]. Emmanuelsson and colleagues combined machine-learning techniques to do PTS1 domain-based cross-species comparisons across eight eukaryotic genomes [2]. The proteins that were selected were predicted to be neither secreted nor transmembrane proteins, and contained a variant of PTS1. An interesting aspect of this study was an analysis of the nine amino acids that immediately precede PTS1. This 9-mer was analyzed by machine learning pattern recognition techniques to establish another criterion by which the authors could exclude false positives. Looking at genomes from yeast to humans under stringent criteria, the analysis generated a core of approximately 60 proteins that are predicted to be peroxisomal in most of the genomes. Using more relaxed criteria, their analysis yielded a data set of 277 putative peroxisomal proteins for Saccharomyces cerevisiae and 1427 for humans. Clustering was then used to organize the data sets, and a clustering cutoff imposed to...
improve the accuracy of their predictions. In the end this groups’ analysis identified 21 of the known yeast peroxisomal matrix proteins and predicted 430 peroxisomal proteins over 8 different genomes, many of which lack localization annotations. The localizations of these unknown proteins remain to be demonstrated experimentally.

Schonbach’s group also utilized the presence of a PTS1 motif to predict peroxisomal proteins [3]. A Genbank dataset of 177,765 mRNA transcripts was pared down to 241 candidates by a Fuzzpro search for C-terminally located (S/A) KL motifs. After removing redundant sequences and matching to Locuslink accession numbers, 85 candidates remained plus an additional 30 candidates that could not be matched by accession number. The 85 candidates were curated by Medline searches and both sets of candidates were curated by experts, resulting in the identification of 117 transcripts known to encode 21 different peroxisomal enzymes across the human, mouse, rat, and guinea pig genomes. They also identified 15 novel peroxisomal proteins. These genes included an alcohol dehydrogenase superfamily member that may function as an antioxidant, and a putative peptidase with protease-related domains. Interestingly, these studies identified Scarb1 as a putative peroxisomal protein. Scarb1 is thought to function in the uptake of high density lipoprotein lipid and is known to interact with Pex5 [4].

Groups studying peroxisomes in plants have also used the presence of peroxisomal targeting signals to predict peroxisomal components of the proteome. Nishimura’s group used the PTS1 or PTS2 to identify putative peroxisomal proteins in Arabidopsis thaliana. Using variants of the canonical PTSs they were able to identify 182 proteins with PTS1 motifs and 74 with PTS2 motifs [5]. These predictions were validated by expression arrays, revealing organ specific differentiation of peroxisomes. Similarly, Reumann used a scheme whereby all known PTS1 and PTS2 targeted peroxisomal matrix proteins were retrieved from protein databases of different plant species [6]. Within these data A. thaliana orthologues were identified and those orthologues with high sequence similarity to proteins known to have non-peroxisomal localizations were excluded. This set of proteins was used to query the nonredundant database and the database of expressed sequence tags (ESTs) at the NCBI. This analysis defined 9 major PTS1 tripeptides, all variants of the canonical SKL motif, and two major variants of the PTS2 motif that can be used to screen proteomes for peroxisomal matrix proteins. Again the sequences adjacent to the PTS peptides were found to contain conserved residues, in particular proline and basic amino acid residues that confer an overall net charge to the region, that are predictive for a protein with a PTS localizing to peroxisomes. This group went on to establish a database AraPerox (http://www.araperox.uni-goettingen.de/) with information on the subcellular targeting predictions, homology, and in silico expression analyses of putative peroxisomal proteins in A. thaliana [7].

Relying strictly on the presence of a consensus peroxisomal targeting signal as a predictor of peroxisomal localization has limitations. These signals can be masked by protein conformation, or multimer formation, or proteins could be subjected to post translational modification to inactivate, or remove the PTS. In addition, proteins that do not have an evident PTS can be imported into peroxisomes by other means, for example, by yet undiscovered PTSs, or by piggy-backing on other proteins recognized by the peroxisomal targeting and import apparatus. Moreover, in silico predictions, stemming from sequenced genomes can extend beyond inventories of the organelle. In the case of Pujol’s group they used the proteomes and sequenced genomes of different organisms to investigate the origin of the organelle[8]. By comparison of the known peroxisomal proteomes from S. cerevisiae and humans, to the predicted proteomes based on the genomes of 19 different organisms, they established a high degree of similarity among membrane proteins and between components of the peroxisomal matrix import machinery and the secretory and proteasome systems. But these proteins had little similarity to bacteria or archa proteins. Based on the dictum that membranes arise from preexisting membranes, they concluded that these results do not support an endosymbiotic origin for the peroxisome. Rather, peroxisomes likely arise from internal membrane systems. Moreover, the authors present a phylogenetic tree in which a set of peroxins (Pex2p, Pex10p, Pex12p, Pex3p, Pex19p, Pex16p and Pex11p) appear to be sufficient to predict the presence of peroxisomes throughout the eukaryotic kingdom, with Pex3p, Pex19p, Pex10p and Pex12p representing a core of required proteins. By contrast, peroxisomal proteins that have overall homology to bacteria or archaea or to proteins from organisms lacking peroxisomes such as Encephalitozoon cuniculi or Giardia lamblia, are likely usurped from other roles and are not good indicators of the presence of the organelle. Thus, by these criteria, while classically defined as an organelle containing enzymes such as oxidases and catalases that catalyze the production and breakdown of hydrogen peroxide, perhaps it is time for a new definition of what constitutes a peroxisome.

3. High throughput localization studies

While no high throughput localization studies have been done that focus specifically on the peroxisomal proteome, the entire S. cerevisiae proteome has been interrogated via two genome-wide subcellular localization studies. The first study, by Kumar et al., was done by indirect immunofluorescence [9]. These studies made use of not only traditional plasmid-based tagging of 2085 S. cerevisiae PCR amplified open reading frames (ORFs), but also employed random tagging of ORFs by transposon mutagenesis. In the latter approach, bacterial transposons were modified to carry a reporter, allowing for the selection of in-frame insertions. A search of the associated TRIPLES database, (http://ygac.med.yale.edu/triples/triples.htm) yields information on four peroxins (Pex1p, Pex5p, Pex8p, and Pex11p).

The second genome wide study was done by O’Shea’s group in which the genome was systematically tagged, one gene at a time, with green fluorescent protein (GFP) [10]. One obvious advantage of this approach is the assessment of the yeast proteome under live cell conditions. A search of the database
under conditions that induce peroxisomes holds promise of not revealed by genome-wide screens. A large scale analysis actin cytoskeleton (for a review see[12]). Careful examination the establishment of cell polarity and reorganization of the For example Rho1p is a well characterized protein involved in peroxisome induction, to increase chances of identification. and the cells grown in a condition that will stimulate the genes are tagged N-terminally, to expose PTS1 signals, addressed by the generation of a localization dataset in which peroxisomes. The paucity of peroxisomal proteins could be 25) and a few of the peroxisomal matrix proteins (7 of 29) to Kohlwein groups localized many peroxisomal proteins (17 of GFP blocked putative PTS1 sequences, these localize to the peroxisome in a PTS1/PTS2 independent manner. In total, 17 known yeast peroxisomal proteins were localized to peroxisomes.

The localization data produced by both the O’Shea and Kohlwein groups localized many peroxisomal proteins (17 of 25) and a few of the peroxisomal matrix proteins (7 of 29) to peroxisomes. The paucity of peroxisomal proteins could be addressed by the generation of a localization dataset in which the genes are tagged N-terminally, to expose PTS1 signals, and the cells grown in a condition that will stimulate peroxisome induction, to increase chances of identification. For example Rho1p is a well characterized protein involved in the establishment of cell polarity and reorganization of the actin cytoskeleton (for a review see [12]). Careful examination of Rho1p under peroxisome inducing conditions showed localization of Rho1p to peroxisomes (see Fig. 1)—a feature not revealed by genome-wide screens. A large scale analysis under conditions that induce peroxisomes holds promise of identifying many more peroxisomal proteins and yielding a more comprehensive picture of the yeast peroxisomal proteome.

4. Interactome studies

High throughput interaction studies have the potential to identify new interactions between peroxisomal proteins and by association, new potential functions and functional relationships. The two most comprehensive experimental approaches have been global systematic mass spectrometry (MS) analyses of immunoprecipitated protein complexes, and global two hybrid studies. Comprehensive MS studies by Gavin et al. [13,14] and Ho et al. [16] revealed numerous components of the yeast interactome. The S. cerevisiae strains used by Gavin et al. in their MS studies were generated by genomically integrating a cassette that contained two epitope tags, the tandem affinity purification (TAP) tags. This allowed for an initial isolation of tagged proteins using one tag, wash and release, then a second isolation using the other tag. The TAP protein assemblies were then identified by MS. Through this method, Gavin et al. identified 1440 distinct proteins in 232 protein complexes[13], while in a later study Gavin et al. were able to repeatedly identify 491 complexes, of which 275 were novel [14]. In the Gavin et al. papers, there is an under representation of peroxisomal proteins. This is understandable since peroxisomes are dynamic organelles, and these experiments were done under conditions in which peroxisomal genes are known to be repressed [15].

By using a different approach Ho et al. [16], were able to bypass the expression problem associated with peroxisomal proteins. This group chose a set of 725 bait proteins, representing numerous functional classes such as signaling molecules and DNA repair proteins. These proteins were overexpressed by means of a GALLI or tet promoter, and immunopurified via the plasmid encoded FLAG tag. Rather than tandem affinity purifications, these purifications were done in a single step. The proteins were run on gels, the bands cut out, and the co-precipitating proteins identified by MS. Once nonspecific proteins were removed from the data sets, 3617 interactions involving 1578 different proteins were identified from a core of 493 bait proteins. The 1578 interacting proteins represent 25% of the yeast proteome. At least six different peroxisomal proteins are present in these data, though the functional relevance of many of these interactions is unclear.

Large potential interaction maps have also been generated by two hybrid studies. The two hybrid method also bypasses the problem of a lack of expression of peroxisomal proteins in non-peroxisome proliferating conditions. Two hybrid studies done by Uetz et al. yielded 957 potential interactions between 1004 proteins [17] while Ito et al. also used the same approach to identify 4549 putative interactions between 3278 proteins in S. cerevisiae [18]. Both of these studies detected numerous interactions between peroxisomal proteins and other proteins. Ito et al. detected interactions with at least 19 different peroxisomal proteins, though many of the interactions are difficult to understand in light of our current knowledge. The interactions detected by Uetz et al. correspond with many known peroxisomal interactions. A search of the word “pex” in the

Fig. 1. GFP fluorescent localization of Rho1p is altered in three different carbon sources. In glucose grown S. cerevisiae cells, Rho1p appears to be cytoplasmic, in glycerol there is GFP signal in what appears to be the endoplasmic reticulum. In oleic acid, a condition which induces peroxisomes, Rho1p localizes to subcellular foci, which co-localize with peroxisomes. Figure taken from [45].
‘Pathcalling Yeast Interaction’ database (http://portal.curagen.com), yields 9 peroxisomal proteins, with numerous interactions. For example Pex14p, which has two Pex5p and one Pex7p interacting domains [19], is involved in the import of proteins into the peroxisomal matrix and has six interactions, which include both Pex5p and Pex7p. Pex5p in turn has four interactions and Pex7p has five interactions, including Pex14p. Conversely a matrix enzyme such as Pot1p appears to be a network “dead end”, connected only to Pex7p. While interesting, there are a number of known interactions that are not detected.

Of course, a constant problem with high throughput studies is validation of the data. While the studies by both Gavin et al. and Ho et al. were able to validate numerous interactions, there were nonspecific interactions and failure to detect some known interactions [20]. Still one should not trivialize the value of these high throughput interaction studies. Granted in both the affinity purification and mass spectrometry and two hybrid studies a false negative is a missed interaction, but with respect to the two hybrid studies, once the auto-activating constructs have been removed from the data, the question of what exactly a false positive is becomes more philosophical. If two proteins are annotated as belonging to different cellular compartments does it mean there is no possibility for interaction? Perhaps these data can be viewed as a framework for interactions that have yet to be uncovered by other means.

A focused interaction matrix of 12 human peroxins (Pex2, 3, 5, 7, 10, 11α, 11β, 12, 13, 14, 16, and 19) was done by Fransen et al. in both yeast two hybrid and bacterial two hybrid systems [21]. While complementary, each assay yielded results unique to that system. For example in the yeast two hybrid system, Pex5p and Pex19p, when fused to the binding domain, auto-activated the reporter, while known interactions between Pex7p–Pex5p, Pex12p–Pex5p,10p, and Pex14p–Pex14,19p were not identified [22]. In contrast, in the bacterial system, neither Pex5p nor Pex19p displayed this self activation while interactions could be detected between Pex14p–Pex14p, and Pex14p–Pex19p. Consistent between the two systems were interactions between Pex5p and Pex14p, and between Pex19p and Pex3p and Pex11p, results that are in good agreement with previously published studies [23–32]. Interestingly this group also found that the use of fragments of Pex12p, Pex13p, and Pex14p yielded an additional 6 interactions in yeast and two interactions in the bacterial system. While these interactions may be artifacts, they may also result from the uncovering of an interaction motif. The interaction matrix done by Fransen et al. is a nice approach to the problem of mapping interactions in the peroxisomal proteome. A focused two hybrid study of this nature of all the known peroxisomal proteins would be of great interest to the peroxisome community. An interaction map of yeast peroxisomal proteins would be a matrix of substantial size and would likely yield novel interactions.

5. Mass spectrometry

Mass spectrometry (MS) based proteomic approaches have been combined with a repertoire of techniques that can be used to improve the coverage of the peroxisomal proteome. These include different strategies to separate the proteins in a fraction (one and two dimensional gel electrophoresis, liquid chromatography), different ways to operate the mass spectrometer (single range, gas phase fractionation), and different methods of quantification (spot intensity, chemical labeling). Kikuchi et al. used one dimensional gel electrophoresis to separate the proteins in a fraction of peroxisomal membranes from rat liver [33]. The membrane proteins were resolved by electrophoresis in a single dimension (size separation) and the lanes of each gel were cut into 64 evenly sized pieces. The proteins were trypsin digested, and the digested peptides were identified by liquid chromatography (LC) coupled tandem mass spectrometry (MS/MS) using data dependant peptide selection. Data dependant ion selection is a method which allows for computer controlled selection of ions with specific characteristics for collision induced dissociation (CID). For example, ions can be selected based on their intensity or m/z ratio. CID is a process in which peptide ions are accelerated by a high energy potential and fragmented by collision with an inert gas. CID is an integral part of MS/MS, a technique where in the first MS, a peptide ion with a specific m/z ratio is selected for CID, and fragmented. This MS/MS spectrum can be used by sequence database search algorithms to identify the sequence of the parent ion. Through the use of these MS techniques, Kikuchi et al. identified a Lon protease, along with other proteins, in a peroxisomal fraction and showed by immunofluorescence microscopy that this protein is peroxisomal. Though comparatively low in coverage of the peroxisomal proteome, these experiments are excellent examples of the power of MS analysis to reveal new components of peroxisomal membrane.

Kikuchi et al. also found that immunopurification of the membranes eliminated the majority of contamination from both the mitochondria and the endoplasmic reticulum (ER). The authors argue that, based on the decreased detection of several mitochondrial and cytoplasmic proteins after immunopurification, the peroxisomal membrane fraction is free of contaminants. The remaining proteins that have been previously annotated as belonging to the ER and mitochondrial may represent bona fide constituents of the peroxisome. For example the P450 cytochromes, which are major constituents of the ER, were not identified in the immunoisolated membranes while aldehyde dehydrogenase, another ER protein, was identified and may be a peroxisomal protein. Previous work done by immunoblotting, though inconclusive, also lends support to the idea that aldehyde dehydrogenase is a peroxisomal protein [34].

The earliest high throughput characterization of the yeast peroxisomal membrane proteome was also done by one-dimensional (1D) electrophoresis, cutting out bands on a gel, and identifying peptides by MS [35]. Cultures of S. cerevisiae were induced in olate containing media and peroxisomes were isolated by differential centrifugation. An osmotic lysis was done to extract the peroxisomal membranes, reducing the quantity of peroxisomal matrix proteins that could mask the identification of the lower abundance proteins in the membrane. The extracted membranes were then resolved by 1D gel electrophoresis, and bands excised. The gel slices were processed and the peptides either applied to a matrix assisted
laser desorption/ionization (MALDI) target or separated by liquid chromatography (LC) and ionized by electrospray ionization (ESI), then identified by MS. Using these techniques Schafer et al. identified 30 peroxisomal proteins including all known peroxisomal integral membrane proteins, and most known peripheral membrane proteins. A group of 15 mitochondrial and cytoplasmic proteins were also identified. While mitochondrial proteins are common contaminants in peroxisomal fractions, both the cytoplasmic and mitochondrial proteins may have uncharacterized and perhaps dynamic peroxisomal association.

An enhancement to 1D electrophoresis is the use of a second dimension: two-dimensional (2D) electrophoresis. In the first dimension proteins are resolved by isoelectric focusing (IEF) and then the proteins are size separated in the second dimension. There are potential problems with this approach, in particular when analyzing membrane proteins. Hydrophobic proteins, such as peroxisomal membrane proteins are difficult to focus based on PI and are frequently insoluble in the buffers used for IEF and may precipitate. Nevertheless, 2D electrophoresis has been used successfully for both mammalian and plant studies of peroxisomes [36,37]. When Islinger et al. [37] used 2D electrophoresis to study peroxisomal proteins from rat liver, they used a modified approach to enhance their ability to resolve and identify peroxisomal membrane proteins. Rather than using the traditional isoelectric focusing in the first dimension, Islinger et al. used benzyltrimethyl-\( n \)-hexadecylammoniumchloride (16-BAC) to separate proteins in the first dimension. Membrane proteins have increased solubility in 16-BAC, preventing precipitation and enhancing identification. Fifty of the most prominent spots on the gel were selected for identification by MALDI-TOF (time of flight) MS and from these, 17 proteins were identified. Again this group found that their enrichment was done by Yi et al. to study the peroxisome proteome in \( S. \) cervicesia [43]. In these experiments, the isolated peroxisomal proteins were not first resolved by gel electrophoresis but rather, the proteins were digested, separated by liquid chromatography, and ionized by electrospray ionization before mass spectrometry (\( \mu \)LC-ESI-MS/MS). Two additional features of these experiments were the use of data dependent ion selection, similar to Schafer et al. [35], and the use of variable range gas phase fractionation (GPF). GPF is the use of iterative mass spectrometric analysis of a given sample over multiple, narrow mass to charge ratios rather than analyzing a sample over a single large range. The rationale behind this approach is that by focusing on narrow mass ranges, the MS has more time to sample peptides within this mass range during and ESI-MS/MS analysis. During ESI-MS/MS peptides are continuously eluting off of a reverse phase column and being sampled in real time by the mass spectrometer. Because the MS is typically programmed to select the most abundant ions for CID, GPF should permit the selection of less abundant ions within the defined mass range for CID. GPF has been shown to increase the number of protein identifications by increasing the number of distinct peptides identified. However, the use of GPF may result in a decrease in the number of times a specific peptide is identified. When these techniques were applied to the analysis of peroxisomal membranes, they proved to have a high identification yield. Yi et al. identified 181 proteins, which included 46 known peroxisomal proteins. While sensitive enough to identify proteins with bipartite localizations, low abundance proteins, and proteins that only transiently associate with the peroxisome (for example Pex5p, which fits all three of these criteria), there was no quantification of relative protein amounts. With over 50% of the proteins identified by this technique previously annotated in the \( S. \) cervicesia Genome Database (SGD) as components of non-peroxisomal compartments, the issue of contamination from other organelles remained a significant shortcoming.

6. Quantification

When DeDuve initially began work on subcellular fractionation by centrifugation, it became readily apparent that absolute purification was a significant challenge. Instead one could enrich for an organelle and monitor this enrichment using marker enzymes. Co-purification of new proteins based on these markers then became a test for organelar localization of the proteins. As we continue subfractionating and analyzing fraction constituents, problems arise when annotating proteins with multiple localizations. In particular when samples are analyzed by MS it is problematic to differentiate those proteins

is interesting in light of ubiquitin activity at the peroxisomes of the yeast \( S. \) cervicesia where Pex4p and Pex10p mediate import to the peroxisome via a ubiquitin mediated mechanism [39]. Pex5p, which is the peroxisomal import receptor for PTS1 signals, has also been shown to be monoubiquitinated [40,41]. A similar approach was then used by the same group to characterize the proteome of glyoxysomes in \( A. \) thaliana [42].

Quite a different type of mass spectrometric analyses was done by Yi et al. to study the peroxisome proteome in \( S. \) cervicesia [43]. In these experiments, the isolated peroxisomal proteins were not first resolved by gel electrophoresis but rather, the proteins were digested, separated by liquid chromatography, and ionized by electrospray ionization before mass spectrometry (\( \mu \)LC-ESI-MS/MS). Two additional features of these experiments were the use of data dependent ion selection, similar to Schafer et al. [35], and the use of variable range gas phase fractionation (GPF). GPF is the use of iterative mass spectrometric analysis of a given sample over multiple, narrow mass to charge ratios rather than analyzing a sample over a single large range. The rationale behind this approach is that by focusing on narrow mass ranges, the MS has more time to sample peptides within this mass range during and ESI-MS/MS analysis. During ESI-MS/MS peptides are continuously eluting off of a reverse phase column and being sampled in real time by the mass spectrometer. Because the MS is typically programmed to select the most abundant ions for CID, GPF should permit the selection of less abundant ions within the defined mass range for CID. GPF has been shown to increase the number of protein identifications by increasing the number of distinct peptides identified. However, the use of GPF may result in a decrease in the number of times a specific peptide is identified. When these techniques were applied to the analysis of peroxisomal membranes, they proved to have a high identification yield. Yi et al. identified 181 proteins, which included 46 known peroxisomal proteins. While sensitive enough to identify proteins with bipartite localizations, low abundance proteins, and proteins that only transiently associate with the peroxisome (for example Pex5p, which fits all three of these criteria), there was no quantification of relative protein amounts. With over 50% of the proteins identified by this technique previously annotated in the \( S. \) cervicesia Genome Database (SGD) as components of non-peroxisomal compartments, the issue of contamination from other organelles remained a significant shortcoming.

6. Quantification

When DeDuve initially began work on subcellular fractionation by centrifugation, it became readily apparent that absolute purification was a significant challenge. Instead one could enrich for an organelle and monitor this enrichment using marker enzymes. Co-purification of new proteins based on these markers then became a test for organelar localization of the proteins. As we continue subfractionating and analyzing fraction constituents, problems arise when annotating proteins with multiple localizations. In particular when samples are analyzed by MS it is problematic to differentiate those proteins
that are low abundance contaminants from those proteins which have enriched specifically with a given fraction. The chemical characteristics of a peptide dictate, to a large extent, the detection of the peptide. MS is inherently poorly quantitative so methods have evolved to aid in the quantification of peptides in MS experiments.

Relative protein quantities can be determined before MS analysis. In both 1D and 2D gel electrophoresis, the gels can be stained and the intensity of the stain quantified by imaging and quantification software. Kikuchi et al. [33] used both gel analysis and sequence coverage by MS to calculate relative protein amounts in their analysis of the proteome of rat liver peroxisomes. Sequence coverage, where gels are bypassed, counts both the number of different peptides from a protein and the number of times those peptides were identified, which can be used to generate an estimate of relative protein amounts. A similar quantification method exploits the use of spectral counts, where the intensity of the spectra is used to estimate quantities. As the mass spectra are acquired, they can be plotted over time and a curve drawn at the spectral peaks. The area under this curve is linearly related to the peptide amount [44].

The relative quantities of proteins in the organelle fraction are compared to a control fraction, such as a total cell lysate or a fraction enriched for a different cell compartment.

A third method to study relative protein quantities is the use of chemical labeling of peptides and comparison of different fractions. This was the method used by Marelli et al. in their analysis of the peroxisomal proteome in S. cerevisiae [45]. The chemical labeling in these experiments was done using isotope-coded affinity tagging (ICAT). ICAT is a technique wherein two fractions can be differentiated by labeling the proteins in each, with chemically identical but isotopically different tags, allowing for the direct comparison of peptide pairs from differentially labeled fractions. The ICAT labeled fractions can then be purified by chromatography [46]. One of the advantages of ICAT is that the reagents are cysteine reactive, thus sample complexity is reduced by only having cysteine containing peptides in the MS analysis. Two different strategies were employed to optimize the correct identification of peroxisomal components. In the first approach (ICAT I), peroxisomes were compared to mitochondria. Both organelles were isolated from olate induced cells by differential centrifugation and lysed. The membrane enriched fractions were then collected, differentially labeled using ICAT reagents, and analyzed by MS. In the second approach (ICAT II), only peroxisomes were isolated, but this time from a strain in which the peroxisomal membrane protein Pex11p had been tagged with protein A (Pex11pA). Again the organelles were lysed and membrane enriched fractions collected by centrifugation. The Pex11pA containing membranes were immunoisolated by chromatography on an IgG resin. The immunoisolated membranes were then compared to a peroxisomal membrane fraction using ICAT labeling and MS.

When the data from both ICAT experiments were analyzed, the results of the two different analyses were found to be complementary. ICAT I was of course better at identifying those proteins that are likely to be mitochondrial contaminants, but was unable to differentiate contaminants from proteins with bipartite localizations to both peroxisomes and mitochondria. This could be determined by comparison to the ICAT II data set. Those true bipartite proteins would be enriched in the affinity purified peroxisomal fraction. The increase in complexity of the protein mix that arises from the mixing of mitochondrial and peroxisomal membrane fractions also concomitantly decreased the number of proteins identified in the mass spectrometer in ICAT I. ICAT II was superior in this respect, as more peroxisomal proteins were identified in this less complex sample. Weaknesses of ICAT II are the potential for over representation of abundant proteins or nonspecific binding of proteins to the chromatography resin. However both these issues can be resolved by comparison to ICAT I.

A critical component to this study was the unbiased way in which the data were analyzed. For each protein the probability of enrichment in the peroxisomal membrane fraction was calculated. These calculations were based on the distributions of proteins that were previously known to be peroxisomal but were not modified to account for proteins that would normally be regarded as contaminants due to annotation to another cellular compartment. This allowed for the inclusion of proteins that may function in or at the peroxisome as well as at another location. It was through this type of mathematical modeling that several proteins with functions seemingly unrelated to peroxisomes, were predicted to also be components of peroxisomal function. These proteins were selected for further study and shown to be peroxisomal by both subcellular fractionation and Western blotting, and also by subcellular localization under peroxisome inducing conditions. Rho1p, a component of the secretory pathway, was shown to be recruited to peroxisomes during peroxisome induction and interact with Pex25p, regulating the state of actin assembly at the peroxisome. This example demonstrates the power of proteomic analysis to provide an opportunity to learn new biology, in particular of proteins with multiple roles and dynamic localizations.

The use of chemical tagging approaches such as ICAT is a significant advancement in the ability to differentiate bona fide peroxisomal components from ever present mitochondrial and ER contaminants [47]. Since fractionation does not purify but rather enriches for a given organelle, it is critical to be able to resolve which peaks are enriching peptides and which are low abundance contaminants. The decision as to whether any given protein may be a peroxisomal component is confounded by the realization that many of these proteins are indeed part of the proteomes of more than one organelle. The combination of different labeling schemes that were employed in these studies (mitochondria versus peroxisomes, then peroxisomes versus affinity-purified peroxisomal membranes) is also an important consideration in these types of experiments. A variant of ICAT, the isobaric Tag for Relative and Absolute Quantitation (iTRAQ), has great potential for this type of analysis. Where ICAT is cysteine reactive, iTRAQ is amine reactive. iTRAQ has not been applied to peroxisomal MS analyses but has been used for numerous other analyses from E. coli [48] to human fibroblasts [49]. Four different tags, of mass 114, 115, 116, 117 Da, are used to generate specific reporter ions when
groups who used 15N labeling to do quantitative proteomics. The use of complementary data sets provides more parameters, aiding in the mathematical modeling of the data. This allows for greater confidence when annotating a peptide as peroxisomal or otherwise.

7. Dynamics

Peroxisomes are dynamic organelles whose components are differentially regulated in response to environmental stimuli [15,50]. One way in which the dynamic peroxisomal proteome can be interrogated is by the use of metabolic labeling. This technique was first developed in the late 1990s by several groups who used 15N labeling to do quantitative proteomics [44,51]. As in the ICAT method, the intensity of the spectra resulting from chemically identical, but isotopically different pairs of peptides can be used to quantify the relative abundance of the proteins isolated from differentially labeled samples. This technique has since evolved into the SILAC (stable isotope labeling by amino acids in cell culture) method. As the name implies, heavy (13C, 15N) amino acids such as l-lysine are provided to growing cultures in the media. The amino acid then becomes incorporated throughout the bulk of the proteome, and the relative abundance is determined by comparing (lysine-containing) differentially labeled peptide pairs.

The main advantage of metabolic labeling over chemical labeling techniques such as ICAT or iTRAQ is the incorporation of the tag into living cells. Because the cells are differentially labeled before harvesting, the two different cell populations can be mixed. By mixing the cells populations before lysis, subcellular fractionation, or purification of the peroxisomal membranes, the quantification will not be affected by these steps. While this technique has yet to be applied to peroxisomes, it has been successfully applied to the study of lipid rafts. Foster et al. [52] compared HeLa cells treated with cholesterol-disrupting drugs to untreated cultures. One of these cultures was grown in the presence of deuterium-labeled leucine while the other was not. The extracts from each culture were quantified by Coomassie blue stain and equal amounts of lysate, with respect to protein content, were mixed together. The lipid rafts were isolated by differential centrifugation and the isolated lipid raft proteins identified by LC-MS/MS. From these analyses, a quantifiable depletion of 241 proteins was detected in response to the cholesterol-disrupting drugs. Since its development this type of metabolic labeling has been used to study dynamic changes in protein phosphorylation [53], RNA binding proteins [54], prostate cancer [55], and other processes. The application of SILAC promises to be a general method for identifying dynamic changes in complexes or organelles, in response to a perturbation or through a developmental process.

As we are yet to fully enumerate bona fide versus contaminating proteins of the peroxisome, an added twist on SILAC is I-DIRT, which is described as a general method for discriminating between contaminants and true components of a given fraction [56]. This protocol also utilizes the stable incorporation of isotopically heavy amino acids into two strains grown under identical conditions. The heavy, wild type strain is mixed with a strain encoding a tagged protein of interest at a 1:1 ratio and the cells are cryogenically lysed. The tagged protein can then be immunoisolated, along with co-isolating proteins. Proteins that were specifically associated with the protein of interest should only be isotopically light since they were associated before thawing while non-specific proteins that associate after cell lysis, during the procedure, will be both heavy and light. This approach was used to study the yeast DNA polymerase ε complex. Though numerous proteins were identified, only three proteins (Dpb2p, Dpb3p, and Dpb4p) were determined, by their isotopic ratios, to have been specifically interacting with Pol2p prior to mixing in a stable complex. This approach should be readily applicable to large complexes or organelles that can be isolated by affinity chromatography.

8. Bottlenecks and solutions

While critical improvements have been made in sample purification, mass spectrometers, use of quantitative procedures, and the analysis of the material, problems remain. Differential purification is a major bottleneck in the analysis of peroxisomal proteomes. In S. cerevisiae, 40 g of cells will yield less than 100 μg of purified peroxisomes and the isolation process is labor intensive. From the time the cells are harvested to where the peroxisome fraction can be recovered takes the better part of a day. One has to wonder how many dynamic interactions are missed due the length of time subcellular fractionation takes. The capture of these dynamic interactions is one of the areas where significant improvement can be made.

One of the challenges in this approach will be to detect, isolate, and quantify peroxisomal subpopulations. Rapid isolation procedures and/or crosslinking can be envisaged as methods to maximize our ability to maintain and isolate separate populations prior to quantitative MS.

Immunopurification of the peroxisome using different peroxisomal targets, in combination with differential centrifugation and free flow electrophoresis, also has the potential to reveal subpopulations of the organelle [57–59]. Free flow electrophoresis is not generally the method of choice to resolve most organelles, due to the pI similarities between the various cell components. However, Völkl and co-workers used antibodies directed against the cytoplasmic domain of the peroxisomal membrane protein PMP70; this had the net effect of altering the pI so that the antibody decorated organelles could be separated from the main peak of organelles. Using this technique, they were able to sort not only peroxisomes, but were later able to extend the technique to isolate subpopulations of peroxisomes.

The challenge at the MS level will be the relative quantification of all the proteins in the complex fractions. In a technique such as electrospray ionization, a complex fraction is ionized into the MS, which now has to identify individual peptides in what is a very complex mix of peptides. Techniques that reduce sample complexity will aid in the identification of proteins in the sample. Both Yi and Marelli used reverse phase
liquid chromatography to simplify complex mixtures of peptides. Reverse phase matrices are resins that are commonly used to separate peptides on the basis of hydrophobicity. Link et al. [60] modified this procedure by integrating a second liquid chromatography step. Thus two dimensional liquid chromatography or multidimensional protein identification technology (MudPIT) was born. This protocol makes use to two different columns to provide even greater separation of peptides. The complex fraction of peptides is first bound on a strong cation exchange (SCX) column then serially eluted using increasing salt concentrations. The proteins eluted from the SCX column are then immobilized on a reverse phase column, the salts washed away and the peptides eluted using increasing concentrations of organic solvents. The now greatly simplified peptides fractions can be analyzed directly by MS/MS.

Post translational modifications are an important component to the understanding of proteomes. Studies have been done looking at many different types of modifications including glycosylation [61–63], sumoylation [64–67], and phosphorylation [68–70], both on small and global scales, in numerous organisms. When characterizing the proteome of S. cerevisiae peroxisome membranes, Shafer et al. were able to identify, by MS, the phosphorylation site in the peroxisomal matrix protein Faa2p (long chain fatty acid CoA ligase 2), raising the possibility that this protein is regulated via phosphorylation events [35]. Perhaps the single biggest bottleneck in the whole of peroxisomal proteomics, and for that matter any global analysis is data validation. Most of the mass spectrometry papers that we have discussed attempt to validate at least some of those proteins predicted to be peroxisomal. It would be nice to see validation of these data sets in terms of functionality. While not exactly high throughput, Kunau’s group did an elegant characterization of peroxisomal protein import complexes and their major components, finding that Pex8p is a central component of both a Pex14p–Pex17p–Pex13p and a Pex2p–Pex10p–Pex12p protein import complexes [71]. These types of experiments will be increasingly important as we acquire more data that require both validation and contextualization.

Fig. 2. A physical interaction map of the Saccharomyces cerevisiae peroxisomal proteome. Physical interactions amongst proteins with “peroxisomal” gene ontology are displayed using Cytoscape 2.2 network visualization software (www.cytoscape.org). Proteins involved in peroxisome biogenesis and peroxisomal metabolism are displayed as hexagons and circles, respectively, whereas interacting metabolites are shown as yellow diamonds. Interactions amongst proteins (downloaded from SGD in Cytoscape format using Batch Download tool) and between proteins and metabolites (metabolites and their interactions are described in [72]) are shown as black edges. Edges are thicker for protein–protein interactions that have been identified by more than one approach. For proteins whose corresponding genes change in expression more than two fold in response to oleate (26 h oleate versus glycerol; [15]), the transcriptional response is reflected by color gradients. Gradations of red are used for genes induced by oleate (no genes were repressed, white indicates no change in expression). Definitions of metabolites are the following: 4HPP: 3-(4-Hydroxyphenyl) pyruvate; ACAR: O-acetylcarnitine; ACCOA: acetyl-CoA; ACOA: acyl-CoA; ASP: l-aspartate; C140: myristate; C160: palmitate; C180: stearate; CAR: carnitine; FADH2M: reduced flavin adenine dinucleotide mitochondrial; FADM: oxidized flavin adenine dinucleotide mitochondrial; GL3P: glycerol 3-phosphate; GLX: glyoxylate; H2O2: hydrogen peroxide; HACOA: (3S)-3-hydroxyacyl-CoA; ICT: isocitrate; LCCA: long chain carboxylic acid; MAL: malate; NAC: Nicotinate; NACM: Nicotinate mitochondrial; NAM: nicotinamide; NAMM: nicotinamide mitochondrial; NH3M: ammonia mitochondrial; OA: oxaloacetate; OACOA: 3-Oxoacyl-CoA; OSUC: oxalosuccinate; T3P2: glycerone phosphate; TYR: l-tyrosine.
Part of validation and contextualization will be the integration of both multiple datasets and other data types. By incorporating microarray expression analyses, metabolite interactions, physical interaction data from multiple sources and localization data, we hope to develop a more complete picture of the network that forms the peroxisomal proteome (Fig. 2).

9. Conclusions

Much progress has been made in recent years in this field. We have gone from a proteome map that was constructed by single, focused experiments, to a much more global picture of the peroxisome, in which the organelle is not viewed as an isolated entity but rather a fluid structure that interacts with, or at the very least shares, proteins with numerous other cell compartments, in particular the mitochondria and endoplasmic reticulum. Many of global analyses discussed herein appear to have an inherent under representation of peroxisomal proteins, indicative of a need on the part of the peroxisome community to work to not only adapt existing protocols to capture more of the peroxisomal proteome, but also develop new tools to allow for the analysis of the dynamic nature of the proteome.

Acknowledgements

The authors acknowledge support from the Canadian Institutes for Health Research for Post-Doctoral Fellowships to R.A.S. and J.J.S., and operating support from the National Institute for Health Research to J.D.A. (RO1 GM067228; U54 GM076547).

References


R.A. Saleem et al. / Biochimica et Biophysica Acta 1763 (2006) 1541–1551 1549


