

Chapter 16

Proteomic Analysis of the Plant Nucleolus

Olga Koroleva, Peter McKeown, Alison Pendle, and Peter Shaw

Abstract The nucleolus is a prominent sub-nuclear structure found in all eukaryotes. It is where the ribosomal RNA genes are transcribed and ribosomes are synthesised. However, much evidence has now accumulated that the nucleolus is involved in many other nuclear processes. Nucleoli are of moderate protein complexity, comprising a few hundred proteins, and can be isolated for proteomic analysis. In this chapter we describe the purification and analysis of plant nucleoli by proteomic methods and summarise the current results. We also discuss more specific tagging methods that have been used to analyse individual protein complexes, as well as methods for analysing post-translational modifications of nucleolar proteins. Finally we discuss the assessment of the reliability of such proteomic data, and the presentation and curation of this type of data.

16.1 Introduction

The nucleolus is a prominent structure found in the nuclei of all eukaryotes. It is where ribosomal RNA genes are transcribed by RNA polymerase I and where these transcripts are processed to form pre-ribosomes. Like all such nuclear sub-structures, the nucleolus lacks any bounding membrane, and must therefore be held together by intermolecular interactions as well as by the continuity of the rDNA

Olga Koroleva, Department of Cell and Developmental Biology, John Innes Centre, Colney, Norwich NR4 7LX, UK

Peter McKeown, Department of Cell and Developmental Biology, John Innes Centre, Colney, Norwich NR4 7LX, UK

Alison Pendle, Department of Cell and Developmental Biology, John Innes Centre, Colney, Norwich NR4 7LX, UK

Peter Shaw, Department of Cell and Developmental Biology, John Innes Centre, Colney, Norwich NR4 7LX, UK, E-mail: peter.shaw@bbsrc.ac.uk

within it with the rest of the nuclear DNA. It has been suggested that the biochemistry of ribosome biogenesis itself is responsible for the existence of the nucleolus as a distinct compartment (Melese and Xue 1995; Hernandez-Verdun et al. 2002) and nucleolar volume correlates with transcription levels. However rRNA transcription occurs without nucleolus formation in Archaea (Omer et al. 2000), whereas nucleoli are formed in organisms lacking RNA polymerase I (Conrad-Webb and Butow 1995) or multiple rRNA repeats (Matsuzaki et al. 2004), suggesting that the formation of a nucleolus is not essential to rRNA transcription, but is an adaptation to create a compartment in which pre-ribosomes may be formed with high efficiency (Raska et al. 2006).

However, much evidence now suggests that there must be more to nucleolar structure and function than making ribosomes. In the first place, nucleolar structure varies considerably between species, between cell types and between individual cells. Second, a number of unexpected, unconventional activities have been localised, at least in part, to the nucleolus (Raska et al. 2006). In most cases these unconventional activities have been investigated in only one or two species, so that it is not yet clear whether they are general nucleolar functions or specific adaptations. For example the nucleolus has been implicated in many other aspects of RNA biology, including biogenesis of snRNAs and snoRNAs, of the signal recognition particle, of tRNAs and RNase P, and of telomerase, as well as in mRNA surveillance. The nucleolus has also been linked to viral infections, to cell cycle regulation, to cancer and to stress responses (Hiscox 2002; Rubbi and Milner 2003; Maggi and Weber 2005; Mayer and Grummt 2005; Yuan et al. 2005). In many cases the involvement of the nucleolus was discovered by localising factors involved in the various processes to the nucleolus. However, recent approaches using high throughput localisation studies and proteomic analysis of purified nucleoli have put these unconventional activities on a more systematic basis (Scherl et al. 2002; Andersen et al. 2005; Pendle et al. 2005).

For the most part, proteins are recruited to the nucleolus as a consequence of their intermolecular interactions rather than by recognisable nucleolar targeting sequences. This has frustrated attempts to predict nucleolar localisation *in silico*. Nucleolar localisation sequences (NoLS) exist in viral proteins, such as the plant umbravirus ORF3 protein (Kim et al. 2004) and both potato leaf-roll virus (PLRV) capsid proteins (Haupt et al. 2005). Indeed, the nucleolar location of one PLRV capsid protein was predicted through NoLS identification (Haupt et al. 2005), and comparison of coronavirus N-protein sequences allowed similar predictions for components of the SARS virus (Reed et al. 2006). However, these reflect binding to existing nucleolar proteins (nucleolin, in the latter case) and the interacting sequences are not well conserved. Similarly, a native nucleolar protein may include sequences important for localisation, as in the basic C-terminal domain of yeast Nop25 (Fujiwara et al. 2006) but even this is not widely conserved. Hence, nucleolar localisation remains difficult to predict.

Without reliable prediction, direct determination of the composition of nucleoli is the best option, and it has recently become possible to use high-throughput proteomic techniques to identify large numbers of nucleolar proteins from

complex mixtures derived from purified nucleoli. Direct visualisation of sub-cellular localisation has also been used in mice to identify nucleolar proteins amongst other nuclear proteins via an enhancer-trap system, although this study took as a starting point a broad range of predicted nuclear proteins of which only 10% were exclusively nucleolar (Sutherland et al. 2001). In pilot screens aimed at systematically localising proteins in mammalian (Simpson et al. 2000) and plant cells (Cutler et al. 2000; Escobar et al. 2003; Tian et al. 2004; Koroleva et al. 2005) by systematic green fluorescent protein (GFP) open reading frame fusion expression, 2–3% of GFP fusions were highly enriched within the nucleolus. As expected, these included many RNA processing functions as well as protein kinases and phosphatases that may regulate various aspects of rRNA metabolism (Koroleva et al. 2005). Perhaps more surprising was the finding that certain transcription factor-like proteins were preferentially located in the nucleolus, whereas related family members were found mainly in the nucleoplasm (Tian et al. 2004; Koroleva et al. 2005).

16.2 Proteomics Techniques used for Plants

The two techniques commonly used for proteomic purposes are 2D polyacrylamide gel electrophoresis (2-DE) and liquid chromatography / mass spectrometry (LC/MS) methods. Although 2D gel approaches do have some advantages (see Chap. 2 by Hurkman and Tanaka, this volume), such as quick visualisation of the distribution of major protein groups, accessibility and relatively low basic costs, and ease of sample and data storage, they are applicable only to the major protein components of the proteome, and not suitable for proteins with extreme values of M_r (relative molecular mass) or pI (isoelectric point), hydrophobic membrane components or low-abundance proteins. More advanced uses of this methodology, such as difference in gel electrophoresis (DiGE) (Unlu et al. 1997; Lilley and Dupree 2006) significantly increase the costs – fluorescent dyes and image analysis software are expensive – but the dynamic resolution is still not very high. Furthermore, since many plant proteins are extensively modified, plant protein mixtures do not generally resolve very well on 2-DE gels. It has been estimated that single gel-based analyses allow the identification of approximately 5% of expressed cellular proteins (Heazlewood and Millar 2003). Only an estimated 120 nucleolar proteins had been identified by such techniques prior to the use of high throughput MS methods (Coute et al. 2006).

Therefore, there seems to be more future in non-gel-based approaches for analysing the plant proteome, such as recently developed techniques like LC-MALDI MS; LC ESI MS/MS and MuDPIT (Aebersold and Mann 2003). Nevertheless, there are several examples where a combination of 2D gel separation with subsequent identification of proteins from excised spots have allowed analysis of the plant nuclear proteome. In *Arabidopsis*, 500–700 spots were detected on 2D gels, and analysis using MALDI-TOF (matrix assisted

laser desorption / ionisation–time of flight) MS led to the identification of 184 spots corresponding to 158 different proteins (Bae et al. 2003). In rice nuclei isolated from suspension culture cells, from a total of 549 proteins resolved on 2-DE, 190 proteins were identified by MALDI-TOF MS from 257 major protein spots (Khan and Komatsu 2004).

MALDI-TOF is the most widespread type of MS analysis. This method simply measures the mass of each ionised peptide, providing a “fingerprint” of protein composition, with protein identification based on matches between measured masses of peptides and predicted masses of proteolytic cleavage products of the proteins present in databases. The caveat is that with the increasing length of the peptide chain, many peptide fragments with different sequences will have very similar masses. The same problem sometimes results from protein modifications. Besides MALDI, another common ionisation method is electro-spray ionisation (ESI). The multiply charged ions generated by ESI often produce MS/MS spectra that are cleaner and potentially easier to interpret (Bodnar et al. 2003). In general, thorough protein separation/purification is required for ESI, since this technique is very intolerant of any contaminants.

A more complex analytical approach is tandem mass spectrometry (MS/MS), which has two MS stages, often a combination of quadrupole and time-of-flight mass detectors (Q-TOF). In this technique, peptide fragments of a given mass are resolved by the first MS stage, then fragmented in a collision chamber and the mass fragmentation patterns are recorded by the second MS stage. The advantage of tandem MS over single MS fingerprinting is that the precise sequence of amino acids in each peptide can be determined, which allows much more reliable identification.

An important principle for the preparation of samples for proteomic analysis is to reduce sample complexity by protein fractionation, therefore increasing the possibility of detecting proteins with lower abundance in the complex protein mixture.

Very commonly, LC separation precedes the ionisation and MS stages. MALDI tandem mass spectrometers allow a preliminary LC separation “off-line”, in which the eluted fractions are spotted onto a MALDI sample plate for later MS analysis. Therefore, LC-MALDI techniques do not suffer from the time constraints imposed by the transient presence of peptides eluting from a column; if necessary, each sample can be analysed more than once. On the other hand, with LC-ESI techniques the sample eluted from the LC is sprayed directly into the ESI source and data is acquired immediately. Because in LC-MALDI the LC is decoupled from the MS analysis, the data is not available immediately, as it is with ESI analyses, and chromatography problems can be present but undetected until too late. MuDPIT (multi-dimensional protein identification technology) is a development of an online LC approach, in which two sequential microcapillary columns (an ion exchange column followed by a reverse phase column) are used to separate very complex mixtures of peptides for MS/MS analysis. This type of analysis allows the identification of very large numbers of proteins/peptides from complex mixtures with the minimum of pre-treatment.

16.3 Purification of Nucleoli

Nucleoli are not bounded by membranes and, since they contain rDNA, they are connected by these DNA strands to the rest of the nuclear chromatin. For these reasons, all methods for preparing nucleoli have relied on mechanical fragmentation procedures. In the case of plant cells, the cell wall must be digested by degrading enzymes, such as cellulases and pectinases, to produce protoplasts. This is necessary to avoid the nuclei and nucleoli being trapped inside the cell wall residues after cellular fragmentation. Pendle et al (2005) used *Arabidopsis* suspension cultures as a convenient, reproducible and abundant source of cells. The cells were protoplasted, and then the protoplasts were carefully disrupted by a small number of strokes of a stainless steel homogeniser, with a clearance between the piston and chamber of 25 μm , observing the state of the preparation after every few strokes. In general the first few strokes released mostly intact nuclei, while further homogenisation then disrupted the nuclei to release nucleoli.

Nucleoli constitute the densest cellular component with the exception of starch granules, which may be an unavoidable contaminant, and can be quickly and efficiently purified from the other cellular debris by differential centrifugation. One of the most critical factors is the Mg^{2+} concentration. Magnesium ions causes chromatin to cross-link into an unworkable network, and centrifugation steps then bring down the chromatin with the nucleoli enmeshed in it. To solve this problem, Pendle et al (2005) simply left Mg^{2+} out of the homogenisation buffer. The nucleoli could then be separated from the chromatin. However, a small concentration of Mg^{2+} was added immediately after centrifugation, since lack of Mg^{2+} caused a gradual disintegration of the nucleoli.

Similar methods have been used for the purification of nucleoli from human cell culture (Andersen et al. 2002; Scherl et al. 2002), although in this case nuclei were purified first, with sonication being necessary to release the nucleoli. This may indicate that human nucleoli are more tightly associated with the chromatin and the rest of the nucleoplasm. We have extracted nucleoli from *Arabidopsis* seedlings on a pilot scale, although not so far on a preparative scale, by chopping tissue with a razor blade to release nuclei, which can then be homogenised in the same way as culture cells and nuclei.

Studies during the past few years using live cell imaging of fluorescently tagged proteins have caused a reappraisal of the dynamic nature of the nucleolus, and indeed all other nuclear compartments. It has become clear that virtually all nuclear and nucleolar proteins are rapidly exchanged with the nucleoplasm, and that what distinguishes 'nucleolar' components is a mean nucleolar residence time of the order of seconds rather than of fractions of a second (Phair and Misteli 2000; Raska et al. 2006). This raises the interesting questions of how nucleoli can be isolated as distinct structures over the timescale of minutes or hours necessary for the purification, and how the structures isolated relate to nucleoli seen in living cells. It would be expected that most of the associated protein and other mobile factors would be lost to the medium during extraction and purification, leaving only the most

strongly attached factors. The fact that most of the expected proteins are present in purified nucleolar fractions suggests that the situation is more complicated. One possibility is that, along with the rapidly exchanging population of proteins, there is a more stable population. Another possibility is that the changes in medium, ionic strength, metabolite concentrations, etc., that occur during nuclear fragmentation and nucleolar purification cause the dynamic exchange of many of the proteins to stop, either because active processes necessary are halted or because changes in the medium cause an effective precipitation of nucleolar contents into a less soluble state. It is important to bear in mind that these processes may affect different nucleolar constituents to different extents, and that 'purified' nucleoli are unlikely to have exactly the same composition as their *in vivo* counterparts. However, extracted *Arabidopsis* nucleoli remain transcriptionally competent (P. McKeown and P. Shaw, unpublished data), which gives some confidence in the validity of the purified fractions.

16.4 Isolation and Analysis of *Arabidopsis* Nucleolar Proteins

These approaches were successfully applied in the two first published proteomic analyses of human nucleoli purified from HeLa cells, in which 257 proteins (Andersen et al. 2002) and 210 proteins (Scherl et al. 2002), respectively, were identified. Recently, a new study identified 667 proteins within nucleoli prepared from HeLa cells (Andersen et al. 2005). So in total, the results obtained from the three independent analyses of HeLa nucleoli provided a list of 713 individual proteins (Coute et al. 2006). The latest study by Andersen et al (2005) employed both an LC MS/MS Q-ToF instrument and a linear ion trap Fourier-transform ion-cyclotron resonance mass spectrometer (FT-ICR-MS), which provided very high resolution and mass accuracy. The development of hybrid mass spectrometers employing FT-ICR potentially presents new opportunities for proteomics analysis of the nucleolus (Coute et al. 2006).

The *Arabidopsis* nucleolar preparation obtained as described above from suspension cell cultures was subjected to high throughput proteomic analysis as described by Pendle et al (2005), which identified 217 proteins. This has been subsequently increased to over 500 proteins by a MuDPIT approach using two-stage microcapillary LC linked to a Q-TOF instrument (P. McKeown, P. Shaw, A. Bottrill, unpublished data). Many of the proteins that were identified were expected: known nucleolar proteins, ribosomal proteins, proteins involved in rDNA transcription, and other RNA-interacting proteins involved in ribosome biogenesis. However, many unexpected proteins were also found in the nucleolus, including for example, spliceosomal proteins, small nuclear RNP (snRNP) proteins and translation factors (see Fig. 16.1). These results reinforce the results of several previous studies, implicating the nucleolus in a variety of functions in addition to ribosome biogenesis, including the biogenesis or transport of a range of RNAs and RNPs, and roles in mRNA maturation, cell cycle control and stress

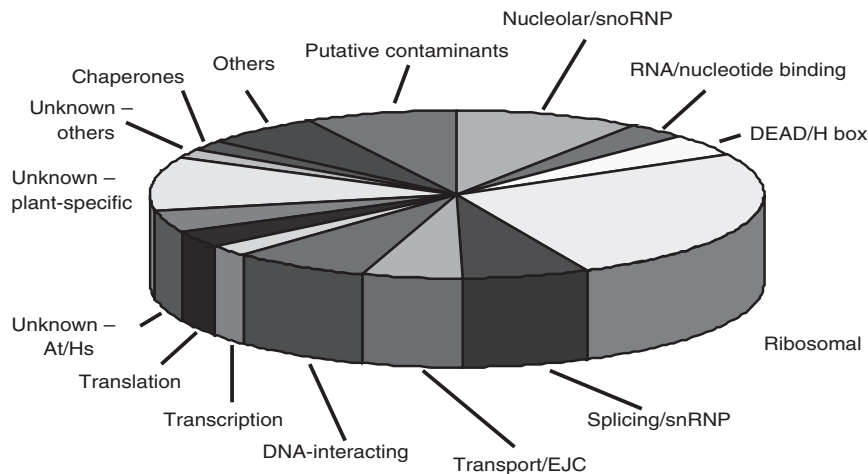


Fig. 16.1 Relative proportions of different protein classes found in Arabidopsis nucleolar proteome

responses (Rubbi and Milner 2003; Andersen et al. 2005; Olson and Dundr 2005; Pendle et al. 2005; Pontes et al. 2006).

It could be argued that many of the proteins identified were contaminants – a large and poorly defined structure such as the nucleolus is likely to be impossible to fully separate from other nuclear and cellular components. Nevertheless, the proteomic analysis has been confirmed by systematic protein expression studies using GFP-tagged proteins. For example, Pendle et al (2005) showed that the vast majority (87%) of the proteins identified by proteomic analysis were also found to be nucleolar-located by the structural criterion of expressing GFP fusions of the identified proteins. Many of the unexpected human proteins have also subsequently been confirmed as being in the nucleolus by structural methods such as GFP tagging and immunofluorescence (Fox et al. 2002). It is also notable that a large proportion of the nucleolar proteins identified in these proteomic analyses are currently completely uncharacterised, further supporting the view that we have as yet a very incomplete picture of the biochemistry that is taking place in the nucleolus. The following websites host current databases of identified nucleolar proteins: <http://lamondlab.com/NOPdb/>; [http://bioinf.scri.sari.ac.uk/cgi-bin/atnopdb/home](http://bioinf.scri.sari.ac.uk/cgi-bin/atnopdb/home;); <http://www.expasy.org/ch2d/>.

16.5 Analysis of Post-Translational Protein Modifications

Post-translational modifications (PTMs) are very common in nucleolar proteins, and several hundred specific modifications, many of which are likely to be functionally important in nucleoli, are known. For example, arginine dimethylation has

defined roles in nucleoplasmic shuttling, and unknown roles on the abundant nucleolar proteins fibrillarin, nucleolin and GAR1 (Lapeyre et al. 1986; Najbauer et al. 1993; Xu et al. 2003). Histone modifications are implicated in the control of rDNA transcription (Earley et al. 2006) and both ubiquitination and SUMOylation pathways are active in the nucleolus (Mo et al. 2002; Song and Wu 2005; Panse et al. 2006). Modifications of histones include phosphorylations, methylations, acetylations and deaminations, and the complex patterns of histone modifications have been proposed to constitute a 'histone code' (Nightingale et al. 2006). A detailed analysis of nucleolar histone modifications could in principal explore the potential histone code as it applies to the rDNA. It is also likely that PTMs of other non-histone proteins have been underestimated, and will be revealed by detailed MS analyses.

The use of tandem MS/MS processing to sequence covalently modified peptides was first reported by Olsen and Mann (2004), and the method is amenable to high-throughput techniques by using a quadrupole as the first MS stage. As with identification of unmodified peptides, it is possible to study either a pre-selected group of proteins of interest, or to sample the entire proteome. More focussed approaches have been used in several organisms for different modifications, and may allow a greater level of detail. Nevertheless, full determination of all modification sites is still technically difficult. For example, analysis of phosphorylation sites of the nucleolar RENT complex in yeast failed to detect all known sites, even when several different techniques were focussed on just these proteins (Chen et al. 2002).

When extracted *Arabidopsis* nucleoli were studied with Q-ToF MS/MS, a number of modifications were detected, including a ribosomal protein found in an acetylated form (P. McKeown and P. Shaw, unpublished work). In the same study, eEF1a (1g07930) was found acetylated at two sites, and the mammalian equivalent has also been detected in an acetylated form (Kim et al. 2006). The total number of acetylations found was fewer than in mouse, partly because the majority of acetylated proteins in mouse are mitochondrial or cytoplasmic, and many acetylated nuclear proteins are probably not enriched in the nucleolus (Kim et al. 2006). Kim et al. made use of affinity purification with an anti-acetyl lysine antibody, but were still unable to detect many known acetylations, especially those associated with rare transcription factors. Combining fractionation with affinity tagging may increase the number of nucleolar covalent modifications detected, but restricts the number of modifications that can be searched. For example, our survey detected seven sites of methionine acetylation that would not have been detected had an acetyl lysine-specific antibody been used.

Care is needed, however, in the interpretation of this type of result. Some small modifications have very similar masses: methylation and formylation are close enough in Mr for methylated peptides to also be detected in a screen for formylations, although with lower confidence ratings, which may allow them to be rejected. Dimethylation and citrullination are isobaric and hence cannot be distinguished by this method. Additionally, it may not be possible to distinguish between the modifications of different residues within peptides, a particular problem with highly modified proteins such as histones. Modifications have to be actively sought in

search engines such as MASCOT, precluding the detection of novel modifications, although known modifications can be found on novel substrates or at positions not previously identified.

In studies of large modifications such as ubiquitination or SUMOylation, which themselves fragment, it has been found to be necessary to isolate the modified proteins before carrying out MS analysis. Plasma membrane-bound proteins modified with glycosylphosphatidylinositol (GPI)-anchor proteins were identified by a 'shave and conquer' technique: isolated membranes were treated with phospholipase D to liberate any GPI-anchored protein, and this protein fraction was analysed (Elortza et al. 2003). Covalent modifications have also been determined by SILAC (stable isotope labelling by amino acids in cell culture), with one population of HeLa cells fed with labelled tyrosine and the other fed with labelled arginine and lysine. Again, this allowed identification of unknown gamma-phosphorylation sites, and could be applicable to nucleoli (Amanchy et al. 2005).

16.6 Quantification Techniques

A typical proteomic analysis of a protein mixture produces a list of identified components but without any quantification of the relative amounts of the different proteins. The detection and successful identification of a particular peptide will depend on a several other factors apart from the relative abundance of the original protein in the mixture: first, on effective digestion by the proteolytic enzyme used to produce the peptide; second, on how easily the peptide is volatilised and ionised; third, on the way the peptide fragments in the collision cell; fourth, on the presence of modifications, which may not have been predicted. Finally, it will also depend on a presence of the peptide sequence in the database or databases to be searched subsequently. The last factor is a particular problem for species without completely sequenced genomes, since MASCOT and similar search engines search for a sequence match that is based on accurate mass data and therefore cannot rely on homology.

For many proteomic applications, for example, cataloguing a particular tissue/organellar proteome or getting a list of binding partners in a protein complex, simple identification of proteins present in the sample is all that is required. On the other hand, quantitative analysis is essential for many other proteomic applications, where the research involves comparison between different tissues and/or developmental stages of an organism, specific conditions or treatments, mutants or transgenic organisms with over-expressed or silenced genes. Therefore, methods have been developed for both relative and absolute quantification of protein amounts in proteomic analysis. Most biological applications require relative analysis, to compare two or more datasets, and in this review we will consider current proteomics techniques for relative quantification.

One possibility for quantification is 2-DE comparison. Although significant progress has been achieved recently with the development of fluorescent stains

and the DiGE technique (Unlu et al. 1997; Lilley and Dupree 2006), 2-DE still allows only a subset of the proteome to be analysed, as certain groups of proteins, such as membrane proteins, co-migrating, low abundance proteins and those with extreme values of molecular mass and pI, cannot be clearly separated and/or visualised. In our experience this means that the majority of plant nuclear and nucleolar proteins either do not run cleanly in 2D gels or do not enter the gel at all. The low dynamic range of DiGE is also a major limitation of this technique. Protein identification from excised gel spots also has limitations because the spots are often contaminated with other proteins. Alternative approaches, non-gel-based quantitative proteomic methods, are based on comparison of peptide abundance, determining the isotopic composition of one or more elements in a compound after stable isotope labelling of a sample, either *in vivo* or *in vitro*. We will describe here a selection of isotopic labelling techniques that have been used on plant material. The labels can either be incorporated *in vivo* as substituted metabolites such as amino acids, or after purification and digestion by reaction with the resulting peptides.

16.6.1 In Vivo Isotope Labelling: SILAC

SILAC uses the *in vivo* incorporation of isotopically substituted amino acids to shift the molecular masses of the resulting peptides. Two or more cell cultures are grown in parallel using different substituted amino acid mixtures, and then the cultures are subjected to different conditions or treatments, such as drug treatments or stress. After treatment, the cultures are pooled and used for biochemical purification. At the MS stage the relative amounts of each peptide can be quantified, since the origin of the peptide can be determined by its isotopic composition. In general, isotopically substituted lysine and arginine are used together, since all tryptic peptides should then be labelled. It is usually necessary to grow the cell culture in the substituted amino acids for a few days to get a good level of incorporation. This approach was used by Andersen et al (2005) to determine the effects of different drug treatments on human nucleoli. Other typical applications are analysing dynamics events by adding isotopically substituted amino acids to a cell culture at a given time, and analysing samples during a time course (Ong et al. 2002; Blagoev et al. 2004). A large scale proteomic study tested the feasibility and technical challenges associated with SILAC to uncover quantitative changes during apoptosis in the nuclear proteome, resulting in the identification and quantification of 1,174 putative nuclear proteins (Hwang et al. 2006). Another recent example of the application of this technique is selective isotope labelling of proteins from *Arabidopsis* cell cultures by growing cells in the presence of a single stable isotopically labelled amino acid (Gruhler et al. 2005). A potential problem in applying this method to plants is that plant cells are better at inter-converting amino acids than animal cells, and so there is a risk that other amino acids may eventually be labelled.

16.6.2 In Vitro Isotope Labelling

16.6.2.1 Isotope Coded Affinity Tagging

In isotope coded affinity tagging (ICAT), free cysteines in a protein are reacted with a special affinity tag. Labeled proteins are enzymatically digested and labelled peptides are separated from the bulk mixture, first using affinity chromatography and then, in a second round, using ion-exchange chromatography prior to MS. The tag has three functional elements: a biotin tag, used during affinity capture (avidin chromatography); the isotopically encoded linker chain (with either eight hydrogens or eight deuteriums); and the reactive group, which will bind to and modify cysteine residues of the protein (Gygi et al. 1999; Li et al. 2003). The tag is marketed by Applied Biosystems (Foster City, CA). However, ICAT has limitations, because it only labels cysteine-containing peptides; 10–15% of every genome codes for proteins with no cysteine, and many proteins contain only a single cysteine, providing only a single peptide that can be potentially quantified.

16.6.2.2 Localisation of Organelle Proteins by Isotope Tagging

Localisation of organelle proteins by isotope tagging (LOPIT), a variation of the ICAT technique, was developed for discovering novel proteins in endomembrane organelles (Dunkley et al. 2004b; Lilley and Dupree 2006), and uses analytical centrifugation in combination with differential isotope labelling. The method involves partial separation of organelles by density gradient centrifugation, thus producing over-lapping fractions, followed by the analysis of protein distributions in the gradient by ICAT labelling (Dunkley et al. 2004a, 2004b) and MS. Multivariate data analysis techniques are then used to group proteins. A good correlation was observed between identification lists of proteins clusters in LOPIT and previous experimental evidence of protein locations within sub-cellular structures, and it has been shown that the LOPIT technique can be used to discriminate Golgi, endoplasmic reticulum (ER), plasma membrane, and mitochondrial/plastid proteins (Lilley and Dupree 2006). In the latter paper the more versatile ITRAQ method (see below) was used instead of ICAT.

16.6.2.3 Isobaric Tag for Relative and Absolute Quantification

Isobaric tag for relative and absolute quantification (ITRAQ) involves chemical tagging of the N-terminus of peptides generated from protein digests; the reagent used reacts with primary amines (Zieske 2006). Fragmentation of the tag attached to the peptides generates a low mass unique reporter ion. There are four tags available, all with an identical mass of 145 Da. The advantages of this approach are that all tryptic peptides are labelled, resulting in higher quality data; up to four labels

can be used for multiple experiments; improved MS/MS fragmentation results in better confidence identification; and post translational modifications can be detected. The subsequent data analysis requires specialised software, and apart from the ProQuant software supplied by Applied Biosciences, non-commercial i-Tracker software is also available for quantitative proteomics using iTRAQ (Shadforth et al. 2005). Disadvantages of iTRAQ include the increased MS time required because of the increased number of peptides, and the fact that samples must be prepared according to strict guidelines.

16.6.3 Comparison of Quantification Methods

What is the best MS quantification technique to use? Recently a comparative study was carried out on three proteomic quantitative methods, DiGE, ICAT, and iTRAQ, using either 2-DE or LC-MALDI TOF/TOF (Wu et al. 2006). All three approaches yielded quantitative results with reasonable accuracy when the same protein mixture was used. In DiGE, accurate quantification was sometimes compromised due to the full or partial co-migration of proteins. The iTRAQ method was more susceptible to errors in precursor ion isolation, especially with increasing sample complexity. The global-tagging iTRAQ technique was more sensitive than the cysteine-specific ICAT method, which in turn was as sensitive as, if not more sensitive than, the DiGE technique.

16.7 Proteomic Analysis of Nuclear Protein Complexes

16.7.1 TAP Tag Strategy and Variations

The tandem affinity purification (TAP) method (Rigaut et al. 1999) was developed to improve the purification of protein complexes, particularly for subsequent proteomic analysis. The most common previous purification methods for protein complexes used antibody pull-down, and this caused a high background of peptides arising from the antibodies used. The original TAP tag consists of tandem protein A domains, linked via a tobacco etch-virus (TEV) protease site to a calmodulin-binding protein (CBP) domain. In practice, the protein of interest is expressed as a TAP-tagged fusion protein, and a cell extract is made after expression of the fusion. The tagged protein, along with complexed proteins, is absorbed on to IgG beads via the protein A domains, washed, and then released by TEV protease treatment. A second round of purification involves binding to calmodulin beads in the presence of Ca^{2+} , followed by washing and release by a buffer containing EGTA to remove the Ca^{2+} , and dissociate the calmodulin complex.

Subsequently, the TAP tag has been made more general by removing a nuclear localisation signal, and has been adapted for use in plant cells by removing a cryptic

splice site, polyadenylation sites and AT- or GC-rich regions, and by the inclusion of castor bean catalase intron 1 for improved expression in plants (Rohila et al. 2004, 2006). Two GATEWAY-compatible binary vectors, NTAPi and CTAPi, were constructed and initially tested using *Agrobacterium*-mediated transformation for transient expression in *Nicotiana benthamiana* leaves. Recently, in a project studying protein kinase signaling networks in cereal leaves, transgenic stable transformants expressing 41 TAP-tagged rice protein kinases were produced and used for subsequent analysis of interacting proteins in rice plants (Rohila et al. 2006). In total, all 41 rice kinases were purified, and 23 of these were isolated as complexes with one or more interacting proteins (Rohila et al. 2006).

However, the NTAPi and CTAPi tags also have some disadvantages. First, many plant proteins have CBDs and therefore will be co-purified with the tagged protein (Reddy et al. 2002). We regularly observe co-purification of elongation factor (EF) 1-alpha proteins, from a small EF-Tu/EF-1A subfamily of identical genes At1g07920, At1g07930, At1g07940, At5g60390 (an example present in Table 16.1), which have a CBD, in protein mixtures pulled down by a number of different NTAPi- and CTAPi-tagged proteins. Two rice EF-homologous proteins were reported to be among the recurring cellular proteins in the TAP procedure by Rohila et al. (2006). Second, the proteolytic treatment by TEV protease involves incubation at 16°C, which may lead to proteolysis by endogenous proteases, because at this stage general protease inhibitors cannot be added to the reaction buffer because they would inhibit the TEV protease. The exception is E-64, but this inhibits only cysteine proteases.

These problems have been addressed recently by development of an alternative tandem affinity purification tag for the isolation of plant protein complexes, called pC-TAPa (Rubio et al. 2005). This construct is also a Gateway-compatible vector, allowing convenient recombination of ORFs from pre-existing Gateway entry clones. The pC-TAPa tag consists of two protein A binding domains, a 3C HRV protease site, six histidine repeats (6-His) and nine myc epitope repeats (myc tag). An advantage of 3C HRV protease is that this enzyme is active at a wide range of temperatures from as low as 4°C (although this requires an increase of either the amount of the enzyme in the reaction mixture, or in incubation time, compared to the relatively fast cleavage at 16°C). The second purification step can use either the 6-His tag or the myc tag, but 6-His tag Ni affinity chromatography has been reported to be more efficient than myc-tag epitope immunoprecipitation (Rubio et al. 2005).

16.7.2 TAP Tag Purification of Nuclear Transport Factor

We have used NTAPi-fusion for purification and proteomic analysis of proteins interacting with nuclear transport factor p15h2 (At1g27970) using transient expression in *Arabidopsis* cell culture, as described previously (Koroleva et al. 2005). This produced relatively high levels of expression of this particular protein. Another significant change made to the original protocol was using TCA

Table 16.1 List of interacting proteins identified by proteomic analysis of p15h2-NTAPi pull-down

| Gene ID | Function | Mass | Peptides | | Score | Other match | Comment |
|-----------|-----------------------------|--------|----------|----------|-------|---|--------------|
| | | | (#) | (unique) | | | |
| At3g04120 | GAPDH | 37,005 | 23; | 6 | 1,504 | | Bait |
| At1g27970 | p15h2 | 14,107 | 16; | 3 | 867 | | Ribosomal |
| At5g09510 | 40S Ribosomal protein S1-54 | 17,061 | 20; | 3 | 733 | At1g04270 | Ribosomal |
| At2g01250 | 60S Ribosomal protein L7-1 | 28,210 | 26; | 2 | 611 | | Ribosomal |
| At1g22780 | 40S Ribosomal protein S18 | 17,591 | 5; | 1 | 591 | At1g34030; At4g09800 | Ribosomal |
| At5g17330 | Glutamate decarboxylase | 57,429 | 20; | 6 | 584 | | Contaminant? |
| At2g44120 | 60S Ribosomal protein L7 | 27,979 | 25; | 2 | 497 | At2g44120; At3g13580 | Ribosomal |
| At5g23860 | Tubulin beta-8 chain | 51,258 | 10; | 3 | 345 | At5g62690; At5g62700; Q93Y15 | Tubulin |
| At1g51060 | Histone H2A | 13,926 | 8; | 2 | 332 | At4g27230; At5g54640 | Histone |
| At1g07920 | Elongation factor 1-alpha | 49,813 | 6; | 2 | 281 | At1g07930; At1g07940; At5g60390; etc | - |
| At1g07790 | Histone H2B | 16,392 | 4; | 1 | 218 | At2g28720; At2g37470; At3g45980; etc | Histone |
| At4g34670 | 40S Ribosomal protein S3A | 29,956 | 3; | 1 | 191 | | Ribosomal |
| At2g29550 | Tubulin beta-7 chain | 51,341 | 7; | 3 | 190 | | Tubulin |
| At2g37270 | 40S Ribosomal protein S5-1 | 23,090 | 3; | 1 | 183 | At3g11940 | Ribosomal |
| At1g09590 | 60S Ribosomal protein L21 | 18,698 | 6; | 2 | 150 | At1g09690; At1g57660; At1g57860 | Ribosomal |
| At5g10860 | CBS domain; cAMP-binding | 22,829 | 4; | 3 | 139 | | - |
| At2g19740 | 60S Ribosomal protein L31 | 13,750 | 4; | 2 | 127 | At4g26230; At5g56710 | Ribosomal |

| | | | | | |
|-----------|---------------------------------------|---------|------|-----|---|
| At3g02080 | 40S Ribosomal protein S19-1 | 15,875 | 4; 2 | 125 | Ribosomal |
| At2g09990 | 40S Ribosomal protein S16 | 16,849 | 6; 2 | 124 | Ribosomal |
| At1g27310 | p15h3 (NTF2) | 13,633 | 2; 1 | 106 | Heterodimer? |
| At1g23290 | 60S Ribosomal protein L27a-2 | 16,339 | 4; 1 | 104 | Ribosomal |
| At3g25520 | 60S Ribosomal protein L5 | 34,394 | 2; 1 | 101 | Ribosomal |
| At2g47610 | 60S Ribosomal protein L7a | 29,339 | 3; 1 | 99 | Ribosomal |
| At3g48170 | Betaine aldehyde dehydrogenase | 55,681 | 5; 2 | 94 | Contaminant? |
| At5g11110 | Sucrose-phosphate synthase | 100,178 | 4; 1 | 94 | Contaminant? |
| At5g17920 | Methionine synthase | 84,646 | 2; 2 | 90 | Contaminant? |
| At1g48830 | 40S Ribosomal protein S7 | 21,909 | 3; 1 | 86 | Ribosomal |
| At1g58380 | 40S Ribosomal protein S2 | 30,950 | 6; 1 | 81 | Ribosomal |
| | | | | | At1g58684; At1g58983; At1g59359; At3g57490; At2g41840 |
| At5g20980 | Methionine synthase | 90,993 | 2; 2 | 68 | Contaminant |
| At1g79330 | Metacaspase 7 | 45,217 | 1; 1 | 68 | Contaminant? |
| At4g31700 | Ribosomal protein S6 | 28,578 | 3; 1 | 68 | Ribosomal |
| At1g65960 | Glutamate decarboxylase | 56,504 | 9; 5 | 68 | Contaminant? |
| At2g36070 | Mitochondrial translocator TIM44-2 | 53,784 | 4; 1 | 66 | - |
| At1g67430 | 60S Ribosomal protein L17-2 | 20,013 | 2; 1 | 63 | Ribosomal |
| At1g22360 | UDP-Glucose glucosyltransferase | 55,014 | 1; 1 | 60 | Contaminant? |
| At3g09680 | 40S Ribosomal protein S23-1 | 15,931 | 1; 1 | 52 | Ribosomal |
| At5g15520 | 40S Ribosomal protein S19-2 | 15,864 | 9; 1 | 51 | Ribosomal |
| At1g72150 | SEC 14 Phosphoglycerate transferase | 64,121 | 1; 1 | 44 | Contaminant? |

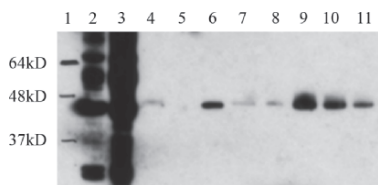


Fig. 16.2 Purification steps of p15h2-NTAPI and interacting proteins for proteomic analysis. Western blot with rabbit anti-CBP primary antibody and anti-rabbit secondary antibody. Lanes: 1 Protein relative molecular mass (Mr) markers, 2 initial cell extract, 3 10x extract after IgG bead adsorption, 4 wash from IgG beads, 5 tobacco etch-virus (TEV) wash, 6 TEV eluate, 7 wash through from calmodulin column, 8–11 fractions eluted from calmodulin column

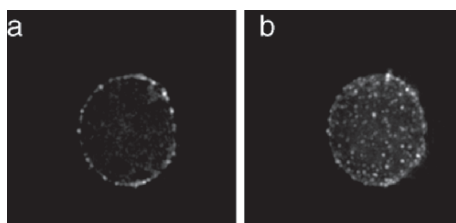


Fig. 16.3 Localisation of p15h2 by green fluorescent protein (GFP) fusion. Numerous small foci at the nuclear periphery are labelled, consistent with localisation to the nuclear pores and with the function of p15 in nuclear export. **a** A single optical section showing the foci labelled at the nuclear periphery. **b** 3D projection from the entire focal section stack through the nucleus shown in **a**

protein precipitation and tryptic digestion of the pellet, instead of separating proteins on a gel and cutting out bands with subsequent digestion. This modification of the procedure avoided losses of yield at the stage of gel separation and extraction of peptides from the gel, which can be 50% or more. It also minimised the time required for MS analysis, as 39 proteins (Table 16.1) were identified simultaneously during a single 1 h LC run followed by nano ESI Q-TOF analysis. Several fractions were set aside from the affinity purification steps for subsequent Western blot analysis, which showed the molecular mass for the fusion protein band to be much higher than expected for this small 14 kDa protein (Fig 16.2). This suggested that the protein was isolated in its homo- or hetero-dimeric form. Another nuclear transport factor, p15h3 (At1g27310), was present among the co-purified proteins (Table 16.1), and so it is likely that these proteins form a stable hetero-dimer or higher order complex. Half of the identified proteins (Table 16.1) were components of the large and small subunits of the ribosome, which would be expected for a nuclear transport factor. We expressed a GFP fusion of p15h2 ORF in Arabidopsis suspension culture and observed a striking pattern of localisation on the surface of nuclear envelope, which probably corresponds to nuclear pores with which p15h2 is associated (Fig 16.3).

From our experience with several other TAP-tagged proteins, the most essential factors for the successful isolation of protein complexes are the abundance and stability of the complex, a conclusion also drawn by Rohila et al (2006).

16.8 Reliability of Proteomic Data

Proteomic analysis is an exciting new technique that can provide large data sets in short periods of time. However, when data from different large-scale projects are compared, certain groups of proteins appear to be ubiquitous despite pre-proteomic sample fractionation techniques. So an important challenge is how to set up criteria to distinguish between true positive identifications and false positives. A consortium has been established to develop general criteria for the publication and exchange of MS data and database search results (Kaiser 2002), and guidelines have been developed to assist researchers in the publication of protein identification from MS data (Carr et al. 2004). A method to estimate the rate of false positive identifications using a reverse database (Peng et al. 2003) has provided a technique for reporting the stringency of the search parameters (Cargile et al. 2004). This approach was applied to the analysis of nucleolar proteins in the most recent publication of the nucleolar proteome (Andersen et al. 2005). The criterion used in this study was set up as at least two matching peptides per protein, a mass accuracy within 3 ppm, a MASCOT score for individual peptides of more than 20, and a delta score of more than 5. The threshold of statistical significance in MASCOT searches is a generally accepted criterion for protein identifications. However, researchers working with unsequenced or incompletely sequenced genomes have to perform searches against expressed sequence tag (EST) databases and sometimes have to consider single peptide matches.

Although some aberrant proteins can be distinguished by the use of replicates, this is insufficient to deal with 'persistent contaminants' – high abundance proteins, or those with a tendency to aggregate with nucleoli during extraction (cytoplasmic metabolic enzymes seem especially prone to this). In fact, as far as proteomic analysis is concerned, these proteins are genuine components of the starting preparation and it is only from the cell biological point of view that they are contaminants. Proteomic analysis is unlikely to resolve this problem, although comparison with other studies and databases of common contaminants can help to identify suspect proteins.

The best solution is to use a completely different technique to confirm the sub-cellular location of proteins identified. The most direct technique is visualisation of cellular location by fluorescent tagging or immunofluorescence, either by sampling a small sample of proteins whose localisation is unconfirmed (Andersen et al. 2002) or by the use of high-throughput approaches such as the Gateway system to analyse a large proportion of the proteins identified (Pendle et al. 2005). This also allows a distinction to be drawn between proteins that are largely nucleolar, and those that are also found in other cellular compartments.

16.9 Data Analysis and Databases

The great advantage of high-throughput techniques is that single sets of experiments produce large volumes of data, but this can be a mixed blessing, especially in the absence of generally agreed strategies for ensuring data quality. Hence, the

utility of proteomic data to the scientific community depends on curation of data into easily navigable databases, and the use of regularly updated formats that allow an assessment of the biological relevance of any given protein or proteins within a proteome.

Following the identification of the proteins from *Arabidopsis* nucleoli, the results were classified by probable protein function and placed online at the *Arabidopsis* Nucleolar Protein Database (<http://bioinf.scri.sari.ac.uk/cgi-bin/atnopdb/home>; Brown et al. 2005). Potential human and yeast orthologues of the *Arabidopsis* nucleolar proteins were identified by reciprocal BLAST search, and comparisons with the human nucleolar proteome were performed. A database of the human nucleolar proteome is also available, and is sufficiently well characterised for use as a test system for novel MS techniques (Vollmer et al. 2006). Over 700 proteins are listed, 500 dynamically characterised, with orthologues annotated from yeast, *Drosophila* and *Caenorhabditis elegans* (<http://www.lamondlab.com/NOPdb/>; Leung et al. 2003). Both databases are cross-referenced to the PubMed systems and the Lamond NOPdb database also includes raw data, i.e. sequenced peptides from tandem MS/MS. As Coute et al. (2006) note, the data on the basis of which such identifications were made may need to be provided to ensure valid comparisons between different experiments or systems (such as the full list of peptides made available following recent analysis of the human acetylproteome; Kim et al. 2006) in a manner analogous to the 'MIAME' standards (http://www.mged.org/Workgroups/MIAME/miame_1.1.html), which allow direct comparisons between microarray data.

Proteomic analysis of isolated organelles cannot determine whether a protein is specific to the organelle in question, or whether it is also located in other parts of the cell. Neither can it determine the distribution of that protein within the organelle. All of this information provides important guides to function. For these reasons, and also to assess the rate of false positives, about half of the proteins identified in our original *Arabidopsis* nucleolar proteome (those for which we could obtain ORFs at the time) were transiently expressed as GFP-fusions (Pendle et al. 2005). These data are also provided in the Atnopdb database, including specificity for sub-nucleolar compartments and presence in other parts of the nucleus and cytoplasm. Attempts to understand the complexities of protein localisation could be aided by the use of databases that compile and compare localisation data from different sources. Unfortunately, neither the cross-species Organelle DB (<http://organelledb.lsi.umich.edu>; Wiwatwattana and Kumar 2005) nor the *Arabidopsis*-specific GFP/MS-based "sub-cellular location database for *Arabidopsis* proteins (SUBA)" at www.suba.bcs.uwa.edu.au (Heazlewood and Millar 2003) or GFP-based database at <http://aztec.stanford.edu/gfp/> (Li et al. 2006) specify sub-nuclear regions.

Even after reciprocal BLAST search, 37 *Arabidopsis* nucleolar proteins are of unknown function due to the lack of characterised orthologues. The presence of protein domains of known function within such proteins can help both in suggesting potential functions and in the broader study of the nucleolus in evolution. For example, an analysis of protein domains of the human nucleolar proteome concluded that the core functions of the nucleolus (i.e. those connected to ribosome

biosynthesis) were of archaeal origin, but that the many eukaryote-specific domains suggested that the nucleolus had undergone a massive subsequent enlargement, driving the evolution of, for example, RNA helicases (Staub et al. 2004). Accordingly, the human proteome database is searchable by protein domain, as well as by gene ontology.

In the examples cited above, genome sequences can be used to identify the proteins detected by MS. This is not possible in organisms with as yet unsequenced genomes, such as wheat, and alternative strategies will be needed to present MS data in a useable manner. Full use will have to be made of EST databases, and it will be necessary to ensure that such libraries are as amenable to automated searching by MASCOT or related software as are genome databases. BLAST search against the genomes of sequenced relatives such as rice will also be important. Although considered the 'model grass', rice is only distantly related to wheat (diverging 130–240 million years ago; Crane et al. 1995), and the proposed sequencing of the *Brachypodium* genome may be important for searching for homologues of wheat proteins within a more closely related species.

16.10 Future Perspectives

The presumed goal of proteomics is to catalogue all the different proteins that can be expressed in a given organism along with all the PTMs of each protein, to quantify the expression and modification levels and to determine the way these levels change in different cell types and cell stages. Cell biologists would add to this the determination of the sub-cellular and sub-organellar location and the dynamics of all the cell's proteins. We are a long way from achieving these ambitious goals, if indeed they are achievable or if they are still considered worth achieving when they become technically possible. One major problem is the sheer number of possibilities; most mammalian genes are subject to alternative splicing, and alternative splicing is likely to be much more widespread in plants than currently expected. Each resulting polypeptide is then potentially subject to hundreds of PTMs. In addition to the enormous number of possible molecular species arising from each gene, the dynamic range in concentrations of proteins is also enormous, ranging from as low as one or two molecules to many billions per cell. This is currently one of the most significant limitations in proteomics technology; low abundance species are simply swamped by the overpowering concentrations of common proteins.

Nevertheless, huge improvements have been made recently both in the experimental methods and in the instrumental flexibility and sensitivity for proteomics. Each new generation of MS machines has greater sensitivity, speed and throughput. The result is that current technologies are proving highly productive for samples of medium complexity, such as nucleoli, which contain a few hundred different proteins. In many cases, these 'partial proteomes' seem more accessible than total cellular proteomes and, at least in the short term, the information that this type of analysis can provide seems more useful. Many practical problems of displaying,

interpreting and curating the information remain, as well as issues of reproducibility and control of artefacts, but these technical problems are likely to be alleviated as the technology matures. As with other types of large-scale bioinformatics data, accessibility and interoperability of different types of data is still a problem. In our experience, the power of proteomics analysis has to be seen in the context of the complementary cell biology techniques for localisation of the proteins identified. Proteomics will clearly be most powerful when closely integrated with other techniques in multi-disciplinary approaches to biology.

Acknowledgements We are grateful to Andrew Bottrill and Mike Naldrett (John Innes Centre) for their help with proteomic analysis. Financial support for this research was provided by the UK Biotechnology and Biological Sciences Research Council (BBSRC).

References

- Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. *Nature* 422:198–207
- Amanchy R, Kalume DE, Iwahori A, Zhong J, Pandey A (2005) Phosphoproteome analysis of HeLa cells using stable isotope labeling with amino acids in cell culture (SILAC). *J Proteome Res* 4:1661–1671
- Andersen JS, Lyon CE, Fox AH, Leung AK, Lam YW, Steen H, Mann M, Lamond AI (2002) Directed proteomic analysis of the human nucleolus. *Curr Biol* 12:1–11
- Andersen JS, Lam YW, Leung AK, Ong SE, Lyon CE, Lamond AI, Mann M (2005) Nucleolar proteome dynamics. *Nature* 433:77–83
- Bae MS, Cho EJ, Choi EY, Park OK (2003) Analysis of the Arabidopsis nuclear proteome and its response to cold stress. *Plant J* 36:652–663
- Blagoev B, Ong SE, Kratchmarova I, Mann M (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat Biotechnol* 22:1139–1145
- Bodnar WM, Blackburn RK, Krise JM, Moseley MA (2003) Exploiting the complementary nature of LC/MALDI/MS/MS and LC/ESI/MS/MS for increased proteome coverage. *J Am Soc Mass Spectrom* 14:971–979
- Brown JW, Shaw PJ, Shaw P, Marshall DF (2005) Arabidopsis nucleolar protein database (AtNoPDB). *Nucleic Acids Res* 33:D633–636
- Cargile BJ, Bundy JL, Stephenson JL Jr (2004) Potential for false positive identifications from large databases through tandem mass spectrometry. *J Proteome Res* 3:1082–1085
- Carr S, Aebersold R, Baldwin M, Burlingame A, Clauser K, Nesvizhskii A (2004) The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data. *Mol Cell Proteomics* 3:531–533
- Chen SL, Huddleston MJ, Shou WY, Deshaies RJ, Annan RS, Carr SA (2002) Mass spectrometry-based methods for phosphorylation site mapping of hyperphosphorylated proteins applied to Net1, a regulator of exit from mitosis in yeast. *Mol Cell Proteomics* 1:186–196
- Conrad-Webb H, Butow RA (1995) A polymerase switch in the synthesis of ribosomal-RNA in *Saccharomyces cerevisiae*. *Mol Cell Biol* 15:2420–2428
- Coute Y, Burgess JA, Diaz JJ, Chichester C, Lisacek F, Greco A, Sanchez JC (2006) Deciphering the human nucleolar proteome. *Mass Spectrom Rev* 25:215–234
- Crane PR, Friis EM, Pedersen KR (1995) The origin and early diversification of angiosperms. *Nature* 374:27–33
- Cutler SR, Ehrhardt DW, Griffiths JS, Somerville CR (2000) Random GFP::cDNA fusions enable visualization of subcellular structures in cells of Arabidopsis at a high frequency. *Proc Natl Acad Sci USA* 97:3718–3723

- Dunkley TP, Dupree P, Watson RB, Lilley KS (2004a) The use of isotope-coded affinity tags (ICAT) to study organelle proteomes in *Arabidopsis thaliana*. *Biochem Soc Trans* 32:520–523
- Dunkley TP, Watson R, Griffin JL, Dupree P, Lilley KS (2004b) Localization of organelle proteins by isotope tagging (LOPIT). *Mol Cell Proteomics* 3:1128–1134
- Earley K, Lawrence RJ, Pontes O, Reuther R, Enciso AJ, Silva M, Neves N, Gross M, Viegas W, Pikaard CS (2006) Erasure of histone acetylation by *Arabidopsis* HDA6 mediates large-scale gene silencing in nucleolar dominance. *Genes Dev* 20:1283–1293
- Elortza F, Nuhse TS, Foster LJ, Stensballe A, Peck SC, Jensen ON (2003) Proteomic analysis of glycosylphosphatidylinositol-anchored membrane proteins. *Mol Cell Proteomics* 2:1261–1270
- Escobar NM, Haupt S, Thow G, Boevink P, Chapman S, Oparka K (2003) High-throughput viral expression of cDNA-green fluorescent protein fusions reveals novel subcellular addresses and identifies unique proteins that interact with plasmodesmata. *Plant Cell* 15:1507–1523
- Fox AH, Lam YW, Leung AKL, Lyon CE, Andersen J, Mann M, Lamond AI (2002) Paraspeckles: a novel nuclear domain. *Curr Biol* 12:13–25
- Fujiwara T, Suzuki S, Kanno M, Sugiyama H, Takahashi H, Tanaka J (2006) Mapping a nucleolar targeting sequence of an RNA binding nucleolar protein, Nop25. *Exp Cell Res* 312:1703–1712
- Gruhler A, Schulze WX, Matthiesen R, Mann M, Jensen ON (2005) Stable isotope labeling of *Arabidopsis thaliana* cells and quantitative proteomics by mass spectrometry. *Mol Cell Proteomics* 4:1697–1709
- Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 17:994–999
- Haupt S, Stroganova T, Ryabov E, Kim SH, Fraser G, Duncan G, Mayo MA, Barker H, Taliany M (2005) Nucleolar localization of potato leafroll virus capsid proteins. *J Gen Virol* 86:2891–2896
- Heazlewood JL, Millar AH (2003) Integrated plant proteomics – putting the green genomes to work. *Funct Plant Biol* 30:471–482
- Hernandez-Verdun D, Roussel P, Gebrane-Younes J (2002) Emerging concepts of nucleolar assembly. *J Cell Sci* 115:2265–2270
- Hiscox JA (2002) The nucleolus – a gateway to viral infection? *Arch Virol* 147:1077–1089
- Hwang SI, Lundgren DH, Mayya V, Rezaul K, Cowan AE, Eng JK, Han DK (2006) Systematic characterization of nuclear proteome during apoptosis: a quantitative proteomic study by differential extraction and stable isotope labeling. *Mol Cell Proteomics* 5:1131–1145
- Kaiser J (2002) Proteomics. Public-private group maps out initiatives. *Science* 296:827
- Khan MM, Komatsu S (2004) Rice proteomics: recent developments and analysis of nuclear proteins. *Phytochemistry* 65:1671–1681
- Kim SC, Sprun R, Chen Y, Xu YD, Ball H, Pei JM, Cheng TL, Kho Y, Xiao H, Xiao L, Grishin NV, White M, Yang XJ, Zhao YM (2006) Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol Cell* 23:607–618
- Kim SH, Ryabov EV, Brown JWS, Taliany M (2004) Involvement of the nucleolus in plant virus systemic infection. *Biochem Soc Trans* 32:557–560
- Koroleva OA, Tomlinson ML, Leader D, Shaw P, Doonan JH (2005) High-throughput protein localization in *Arabidopsis* using *Agrobacterium*-mediated transient expression of GFP-ORF fusions. *Plant J* 41:162–174
- Lapeyre B, Amalric F, Ghaffari SH, Rao SVV, Dumbar TS, Olson MOJ (1986) Protein and cDNA sequence of a glycine-rich, dimethylarginine-containing region located near the carboxyl-terminal end of nucleolin (C23 and 100 kDa). *J Biol Chem* 261:9167–9173
- Leung AKL, Andersen JS, Mann M, Lamond AI (2003) Bioinformatic analysis of the nucleolus. *Biochem J* 376:553–569
- Li J, Steen H, Gygi SP (2003) Protein profiling with cleavable isotope-coded affinity tag (cICAT) reagents: the yeast salinity stress response. *Mol Cell Proteomics* 2:1198–1204
- Li SJ, Ehrhardt DW, Rhee SY (2006) Systematic analysis of *Arabidopsis* organelles and a protein localization database for facilitating fluorescent tagging of full-length *Arabidopsis* proteins. *Plant Physiol* 141:527–539

- Lilley KS, Dupree P (2006) Methods of quantitative proteomics and their application to plant organelle characterization. *J Exp Bot* 57:1493–1499
- Maggi LB, Weber JD (2005) Nucleolar adaptation in human cancer. *Cancer Invest* 23:599–608
- Matsuzaki M, Misumi O, Shin-I T, Maruyama S, Takahara M, Miyagishima SY, Mori T, Nishida K, Yagisawa F, Yoshida Y, Nishimura Y, Nakao S, Kobayashi T, Momoyama Y, Higashiyama T, Minoda A, Sano M, Nomoto H, Oishi K, Hayashi H, Ohta F, Nishizaka S, Haga S, Miura S, Morishita T, Kabeya Y, Terasawa K, Suzuki Y, Ishii Y, Asakawa S, Takano H, Ohta N, Kuroiwa H, Tanaka K, Shimizu N, Sugano S, Sato N, Nozaki H, Ogasawara N, Kohara Y, Kuroiwa T (2004) Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* 428:653–657
- Mayer C, Grummt I (2005) Cellular stress and nucleolar function. *Cell Cycle* 4:1036–1038
- Melese T, Xue Z (1995) The nucleolus – an organelle formed by the act of building a ribosome. *Curr Opin Cell Biol* 7:319–324
- Mo YY, Yu YN, Shen ZY, Beck WT (2002) Nucleolar delocalization of human topoisomerase I in response to topotecan correlates with sumoylation of the protein. *J Biol Chem* 277:2958–2964
- Najbauer J, Johnson BA, Young AL, Aswad DW (1993) Peptides with sequences similar to glycine, arginine-rich motifs in proteins interacting with RNA are efficiently recognized by methyltransferase(s) modifying arginine in numerous proteins. *J Biol Chem* 268:10501–10509
- Nightingale KP, O'Neill LP, Turner BM (2006) Histone modifications: signalling receptors and potential elements of a heritable genetic code. *Curr Opin Genet Dev* 16:125–136
- Olsen JV, Mann M (2004) Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc Natl Acad Sci USA* 101:13417–13422
- Olson MOJ, Dunder M (2005) The moving parts of the nucleolus. *Histochem Cell Biol* 123:203–216
- Omer AD, Lowe TM, Russell AG, Ebhardt H, Eddy SR, Dennis PP (2000) Homologs of small nucleolar RNAs in archaea. *Science* 288:517–522
- Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 1:376–386
- Panse VG, Kressler D, Pauli A, Petfalski E, Gnadig M, Tollervey D, Hurt E (2006) Formation and nuclear export of preribosomes are functionally linked to the small-ubiquitin-related modifier pathway. *Traffic* 7:1311–1321
- Pendle AF, Clark GP, Boon R, Lewandowska D, Lam YW, Andersen J, Mann M, Lamond AI, Brown JW, Shaw PJ (2005) Proteomic analysis of the Arabidopsis nucleolus suggests novel nucleolar functions. *Mol Biol Cell* 16:260–269
- Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res* 2:43–50
- Phair RD, Misteli T (2000) High mobility of proteins in the mammalian cell nucleus. *Nature* 404:604–609
- Pontes O, Li CF, Nunes PC, Haag J, Ream T, Vitins A, Jacobsen SE, Pikaard CS (2006) The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell* 126:79–92
- Raska I, Shaw PJ, Cmarko D (2006) Structure and function of the nucleolus in the spotlight. *Curr Opin Cell Biol* 18:325–334
- Reddy VS, Ali GS, Reddy AS (2002) Genes encoding calmodulin-binding proteins in the Arabidopsis genome. *J Biol Chem* 277:9840–9852
- Reed ML, Dove BK, Jackson RM, Collins R, Brooks G, Hiscox JA (2006) Delineation and modeling of a nucleolar retention signal in the coronavirus nucleocapsid protein. *Traffic* 7:833–848
- Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* 17:1030–1032

- Rohila JS, Chen M, Cerny R, Fromm ME (2004) Improved tandem affinity purification tag and methods for isolation of protein heterocomplexes from plants. *Plant J* 38:172–181
- Rohila JS, Chen M, Chen S, Chen J, Cerny R, Dardick C, Canlas P, Xu X, Gribskov M, Kanrar S, Zhu JK, Ronald P, Fromm ME (2006) Protein–protein interactions of tandem affinity purification-tagged protein kinases in rice. *Plant J* 46:1–13
- Rubbi CP, Milner J (2003) Disruption of the nucleolus mediates stabilization of p53 in response to DNA damage and other stresses. *EMBO J* 22:6068–6077
- Rubio V, Shen Y, Saijo Y, Liu Y, Gusmaroli G, Dinesh-Kumar SP, Deng XW (2005) An alternative tandem affinity purification strategy applied to Arabidopsis protein complex isolation. *Plant J* 41:767–778
- Scherl A, Coute Y, Deon C, Calle A, Kindbeiter K, Sanchez JC, Greco A, Hochstrasser D, Diaz JJ (2002) Functional proteomic analysis of human nucleolus. *Mol Biol Cell* 13:4100–4109
- Shadforth IP, Dunkley TP, Lilley KS, Bessant C (2005) i-Tracker: for quantitative proteomics using iTRAQ. *BMC Genomics* 6:145
- Simpson JC, Wellenreuther R, Poustka A, Pepperkok R, Wiemann S (2000) Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO Rep* 1:287–292
- Song ZY, Wu M (2005) Identification of a novel nucleolar localization signal and degradation signal in Survivin-deltaEx3: a potential link between nucleolus and degradation. *Oncogene* 24:2723–2734
- Staub E, Fiziev P, Rosenthal A, Hinzmann B (2004) Insights into the evolution of the nucleolus by an analysis of its protein domain repertoire. *Bioessays* 26:567–581
- Sutherland HGE, Mumford GK, Newton K, Ford LV, Farrall R, Dellaire G, Caceres JF, Bickmore WA (2001) Large-scale identification of mammalian proteins localized to nuclear sub-compartments. *Human Mol Genet* 10:1995–2011
- Tian GW, Mohanty A, Chary SN, Li S, Paap B, Drakakaki G, Kopec CD, Li J, Ehrhardt D, Jackson D, Rhee SY, Raikhel NV, Citovsky V (2004) High-throughput fluorescent tagging of full-length Arabidopsis gene products in planta. *Plant Physiol* 135:25–38
- Unlu M, Morgan ME, Minden JS (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 18:2071–2077
- Vollmer M, Horth P, Rozing G, Coute Y, Grimm R, Hochstrasser D, Sanchez JC (2006) Multi-dimensional HPLUMS of the nucleolar proteome using HPLC-chip/MS. *J Sep Sci* 29:499–509
- Wiwatwattana N, Kumar A (2005) Organelle DB: a cross-species database of protein localization and function. *Nucleic Acids Res* 33:D598–D604
- Wu WW, Wang G, Baek SJ, Shen RF (2006) Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF. *J Proteome Res* 5:651–658
- Xu C, Henry PA, Setya A, Henry MF (2003) In vivo analysis of nucleolar proteins modified by the yeast arginine methyltransferase Hmt1/Rmt1p. *RNA* 9:746–759
- Yuan XJ, Zhou YG, Casanova E, Chai MQ, Kiss E, Grone HJ, Schutz G, Grummt I (2005) Genetic inactivation of the transcription factor TIF-I leads to nucleolar disruption, cell cycle arrest, and p53-mediated apoptosis. *Mol Cell* 19:77–87
- Zieske LR (2006) A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies. *J Exp Bot* 57:1501–1508