

REVIEW

Isolation and characterization of plant protein complexes by mass spectrometry

Delphine Pflieger^{1*}, Jean Bigeard^{2*} and Heribert Hirt²

¹ Laboratoire Analyse et Modélisation pour la Biologie et l'Environnement, UMR CNRS 8587, Université d'Evry Val d'Essonne, Evry, France

² URGV Plant Genomics, INRA/CNRS/University of Evry Val d'Essonne, Rue Gaston Crémieux, Evry, France

The components that enable cells and organisms to fulfill a plethora of chemical and physical reactions, including their ability to metabolize, replicate, repair and communicate with their environment are mostly based on the functioning of highly complex cellular machines which are to a large extent composed of proteins. With the development of MS techniques compatible with the analysis of minute amounts of biological material, it has become more and more feasible to dissect the composition and modification of these protein machineries. Indeed, new purification methods of protein complexes followed by MS analysis together with the genomic sequencing of various organisms – and in particular of crop species – now provide unforeseen insight to understand biological processes at a molecular level. We here review the current state of the art of *in vivo* protein complex isolation and their MS-based analytical characterization, emphasizing on the tandem affinity purification approach.

Received: October 12, 2010

Revised: January 15, 2011

Accepted: January 31, 2011

Keywords:

MS / Plant proteomics / Protein complexes / Tandem affinity purification

1 Introduction

A multitude of cellular activities are carried out by specific molecular machineries of which proteins constitute core components. The existence, role and functioning of such modules started to be addressed about 60 years ago. In the late 1940s already, the notion of 'multienzyme complex' was defined to describe an 'organized mosaic of enzymes' that catalyze consecutive reactions in metabolic pathways [1]; later such units were dubbed 'metabolons' [2]. In these modules, metabolic intermediates are transferred by 'metabolite channeling' between consecutively required enzymes, without diffusing into the bulk phase of the cell. A number of biosynthetic pathways in plants involve such enzyme assemblies [3–5]. In other modules, one enzyme may play a

role of structural or regulatory subunits for its partner enzymes, as exemplified by the cysteine synthase holoenzyme in *Arabidopsis thaliana* [6]. To isolate the complexes, diverse physical and chemical methods were used to study the macromolecular interactions within functional complexes, among which ultracentrifugation, chromatography and electrophoresis [7]. Surface plasmon resonance and nuclear magnetic resonance were used to assess the internal protein bonds and their possible correlations with protein structure [8]. Affinity purification has become a technique of choice to purify complexes involving one particular protein whose interaction partners are unknown. The latter technique was in particular used to study the mechanism of ordered protein assembly until reconstituting a functional structure *in vitro* [9]; obtaining such fully active molecular machines has since then been described to be feasible in a cell-free protein synthesis system [10].

The present review focuses on the characterization of protein complexes *in vivo*, by the combination of affinity purification and mass spectrometry (MS) analysis. *In vitro* or heterologous system approaches are not reported here

Correspondence: Dr. Heribert Hirt, URGV Plant Genomics, UMR INRA 1165/ERL CNRS 8196/University of Evry Val d'Essonne, 2 rue Gaston Crémieux, 91057 Evry, France

E-mail: heribert.hirt@univie.ac.at

Fax: +33-1-60874510

Abbreviations: CBP, calmodulin-binding peptide; CSN, COP9 signalosome; PK, protein kinase; TAP, tandem affinity purification; TBP, TATA-box binding protein; TEV, tobacco etch virus

*These authors contributed equally to this work.

Colour Online: See the article online to view Fig. 1 in colour.

although they are also an important source of data. The strategies of single-step and two-step affinity purification are first presented, with a special emphasis on studies using two-step affinity purification in plants. They are then illustrated by outstanding examples of their implementation in large-scale analyses in various organisms and cell types. The need to characterize in-depth protein complexes is finally discussed according to the following aspects: the dynamics of complex composition, the post-translational modifications (PTMs) of proteins, the stoichiometry of protein complex components and the direct protein–protein contacts. We mostly limited our bibliography to approaches involving the MS analysis of enzymatically digested protein complexes. However, illuminating reports described the analysis of intact protein assemblies by MS, thus providing precise insight into the subunit architecture of complexes (stoichiometry of associated proteins, exchange of monomers between complexes, etc.). Such studies were reviewed in detail elsewhere [11] and we only briefly deal with them at the very end of this review.

2 Isolation and analysis of protein complexes

Obtaining a description of multi-protein complexes corresponding to their actual *in vivo* state requires isolating them in close-to-physiological conditions before further MS analysis. This can be performed by diverse methods such as

native fractionation on sucrose gradients [12], chromatography [13–15], Blue Native (BN)-PAGE [16–18] or multi-dimensional electrophoresis [19]. Nonetheless, affinity purification has been favored by the majority of studies aiming to characterize protein complexes. Two main generic affinity-based methods can provide an efficient way to purify complexes from crude extracts, namely epitope-tagging [20] (single-step affinity purification), and tandem affinity purification (TAP) [21, 22] (two-step affinity purification) (Fig. 1).

2.1 Single-step affinity purification techniques

In single-step affinity purification, the bait protein is expressed in a tagged form, the tag being a peptide or a protein fusion that possesses high affinity for a ligand linked to a solid support [23]: the ligand may be a metal ion [24, 25], a biomolecule, peptide or sugar [26, 27] or a whole protein [28, 29]. A frequent drawback of single-step affinity purification is the large number of co-purifying contaminant proteins, which often non-specifically bind to the affinity matrix or the antibodies used. This approach is widely used in plant research and identifies putative interacting proteins. Just to mention one study, MS analysis of purified epitope-tagged 26S proteasomes from *A. thaliana* [30] revealed that the core protease-regulatory particle is actually composed of a heterogeneous set of particles being made up of several proteasome-associated proteins, which probably assist in protein complex assembly and regulation.

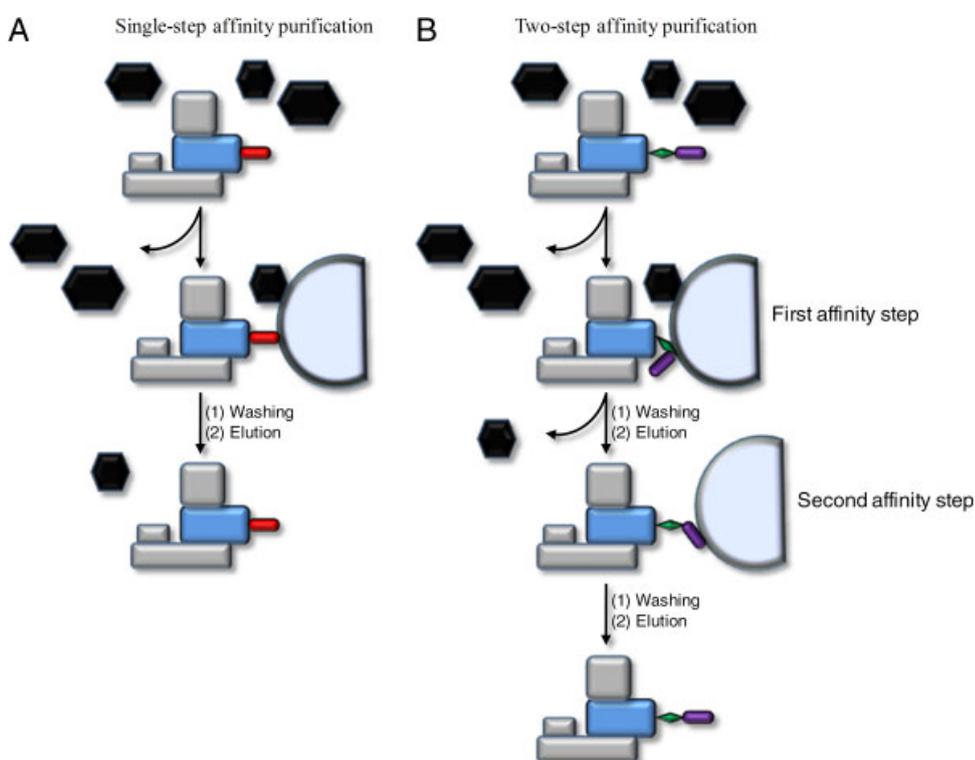


Figure 1. Two main affinity-based methods to purify protein complexes. (A) Single-step affinity purification. (B) Two-step affinity purification. Grey boxes represent partner proteins of the tagged bait protein and black boxes represent contaminant proteins; the blue box and the small colored boxes correspond to the bait protein and its fused tags, respectively.

2.2 Two-step affinity purification techniques

To get higher certainty in the identity of the interacting partners of a protein of interest, two-step affinity purification techniques were developed. These techniques can be grouped under the label 'TAP'. In comparison to one-step purification, the TAP protocol tends to yield cleaner samples but transient or weak interactions can be lost due to the multiple incubation and washing steps [31]. Initially, the TAP approach was developed to isolate and identify protein complexes in the yeast *Saccharomyces cerevisiae* [21]. In its original form, a bait protein is fused to different tags, which are sequentially used to isolate a highly purified bait protein and its associated protein partners. The three tags are two IgG-binding units of protein A of *Staphylococcus aureus* (ProtA), a tobacco etch virus (TEV) protease cleavage site and a calmodulin-binding peptide (CBP). This method proved to be successful and a first large-scale study was published a few years later in yeast, where 1739 genes were processed leading notably to the purification of 589 protein assemblies [32]. Later on, Krogan et al. processed 4562 different TAP-tagged proteins in yeast [33]. About 2357 purifications were successful demonstrating notably 7123 protein–protein interactions involving 2708 proteins. The TAP approach was also applied and developed in other organisms, including bacteria [34], mammalian cells [35] and insect cells [36].

For plants, the first work using the TAP approach was published by Rohila et al. in 2004 [37], i.e. five years after the development of the technique in yeast. Actually, beforehand, some teams already tried to analyze protein complexes using methods close to the TAP. This was the case of Rivas et al. in 2002 who fused the TAP tags from Rigaut et al. to the C-terminus of the tomato Cf-9 gene and transiently expressed the construct in *Nicotiana benthamiana* leaves [38]. However, although the authors could show that the Cf-9 protein is part of a membrane complex consisting of an additional glycoprotein partner(s), they did not take advantage of the full potential of the combination of tags as they only used the ProtA tag and the TEV protease cleavage site. In the same way, Zhong et al. could identify 86 different proteins that were putatively part of complexes associated with the TATA-box binding protein (TBP) in rice [39]. The authors used rice suspension cell cultures expressing TBP fused to a biotin peptide and a TEV protease cleavage site. Rohila et al. were thereby the first to apply a standard TAP approach in plants [37]. The authors characterized and improved for plants a system developed initially in yeast. For example, they detected and mutated a nuclear localization signal (NLS) in the CBP region of the N-terminal TAP tag; they modified the duplicated nucleotide sequence coding for the protein A domain to reduce recombination and the possibility of repeat-induced gene silencing. After all these modifications and optimizations, they transiently expressed a hybrid glucocorticoid receptor fused to this new synthetic TAP cassette in tobacco leaves and could identify

two interacting proteins, HSP70 and HSP90, using an additional step of formaldehyde cross-linking. These two HSPs were known to interact with the glucocorticoid receptor *in vivo* in mammalian cells and *in vitro* in plants, which confirmed the feasibility of a whole TAP procedure in plants.

Several parameters have to be examined for the success of a TAP approach. The possibility to identify a protein complex to which the bait protein belongs depends on the inherent properties of the complex under study. The stability of the complex, its abundance and its subcellular localization are specific characteristics of each complex, indicating that the success of a TAP approach is hardly predictable and hence highly empirical. Having considered these inescapable physicochemical aspects, several ways of performing a TAP approach apply.

First, the choice of the TAP cassette, meaning the combination of tags, is an important parameter. As we will see in the examples described thereafter, different TAP cassettes have been used in plant studies. Actually, several of them do not correspond to the TAP cassette developed by Rigaut and colleagues in 1999. For a detailed description of the commonly used tag combinations for TAP, see the review of Li [40].

Then, the promoter that is used for expression of the bait protein has to be chosen carefully. The amount of expressed bait protein is a crucial factor in the success of the approach. Overall, it can be stated that the more a protein is expressed the higher will be the amount of the purified bait protein. However, even if this is usually true for the bait protein, the protein partners in the complex(es) will usually not adapt their expression and/or abundance to the level of the bait protein. Moreover, if the biological function of the bait protein necessitates a tight regulation, overexpression will very probably lead to phenotypes different from the wild-type.

In the same way, the genetic background needs to be considered. The alternatives are usually a wild-type background or a bait protein mutant background. The complementation of the mutant phenotype is a good indicator that the TAP-tagged protein is functional and that the protein complex(es) are functional as well. Moreover, working in a mutant background avoids the competition between the endogenous protein and the tagged protein for the formation of the complex(es).

Finally, the expression system and/or type of culture system of the biological material are other important aspects to examine. Several possibilities exist: cell cultures, transient in planta expression or stable plant transformation. The use of stable plant transformants is of course closer to physiological conditions, but as we will see in one of the examples below, the use of cell cultures also has some advantages. The success of the purification depends on a combination of all these parameters. Ideally, a TAP approach should enable to highly purify a protein and its associated partners while respecting their physiological characteristics (expression

level, stoichiometry, dynamics, stability, subcellular localization and function). Below are several examples that illustrate some TAP approaches that were realized successfully in targeted or larger scale studies.

2.3 The COP9 signalosome (CSN) complex in *Arabidopsis*

Rubio et al. were the first to use stably transformed *Arabidopsis* transgenic lines to study the CSN complex, a highly conserved multi-protein complex consisting of eight subunits and playing a key role in the ubiquitin/26S proteasome proteolytic regulatory pathway [41]. The authors generated a modified TAP cassette by replacing the TEV protease cleavage site with the more specific sequence recognized by the human rhinovirus 3C (3C) protease, which is active at low temperatures. They also replaced the CBP by a six-histidine-repeat motif (6xHis) and a nine-myc-peptide repeat (9xmyc) for the second affinity purification step. All these modifications allowed the TAP procedure to be efficiently performed at 4°C, allowing more stable complex isolation. They generated transgenic plant lines with 31 TAP-tagged proteins involved in light signaling and the ubiquitin/26S proteasome pathway, but their work focused on the study of the CSN3 subunit. The TAP-CSN3 protein was under the control of two copies of the CaMV 35S promoter. The protein fusion complemented the lethal *csn3* phenotype: CSN3-TAP-containing transgenic plants in the mutant background (CSN3-TAP *csn3/csn3*) overcame the lethal *csn3* phenotype, although their development was delayed when compared with transgenic plants in a heterozygous mutant background (CSN3-TAP *CSN3/csn3*). The protein fusion was therefore at least partially functional. By gel filtration assays, the authors could show that the TAP-CSN3 protein is part of a larger protein complex, as it is the case for the endogenous CSN3 protein. Finally, using specific-subunit antibodies and partial MS analysis of the CSN3-TAP purified sample, the authors showed the presence of six of seven other CSN subunits tested, demonstrating the efficiency of this whole TAP approach for the study of this CSN multi-protein complex.

2.4 Protein kinase (PK) complexes in rice

Michael E. Fromm's team published two papers where PK complexes from rice were studied by a TAP approach [42, 43]. As mentioned previously, this team was the first to successfully use TAP in plants [37]. In these two articles, the authors overall analyzed 129 PKs from rice. They fused the PK cDNAs in-frame with an N-terminal TAP cassette under the control of the maize ubiquitin promoter to allow constitutive expression of the tagged proteins in transgenic rice plants. After the whole TAP process, they could recover 83 TAP-tagged PKs (64%) to identify 29 complexes (35% of

recovered PKs). Overall, 22% of the 129 PKs analyzed by TAP led to the identification of protein complexes. At a first glance, this percentage may seem low. However, for most of the PKs analyzed the authors performed only a single TAP analysis using a standard protocol. Some biological repetitions of the experiment and optimizations of the protocol according to the properties of the protein under study would probably increase the success rate of this medium-scale approach. The protein complexes may not always be abundant and/or stable enough to be recovered efficiently by this method. In addition, the authors compared the results obtained by the TAP approach with yeast two-hybrid screening. They screened 93 of the 129 PKs in a yeast two-hybrid system against a library of rice cDNA prey vectors. Only four proteins interacted with one or more of the same proteins in both techniques. The same interacting proteins found by both methods represented 3% of the total interacting proteins. The authors concluded this percentage was similar to that found in a comparison of protein interaction methods in yeast.

2.5 The cell cycle machinery in *Arabidopsis*

A few years ago, Geert De Jaeger's team developed a novel TAP method that was applied to study the cell cycle machinery in *A. thaliana* [44]. The authors chose cell suspension cultures for several reasons: they constitute an unlimited supply of plant material and their ploidy level of 8n provides more protein for assembly into protein complexes; as undifferentiated cells, they divide actively and can be easily synchronized. The authors actually developed a whole TAP platform from the cloning of the genes of interest to MS analysis of the purified complexes. In this article the authors used the same TAP cassette as Rohila et al. described previously [37], but they later developed new TAP cassettes [45]. They also compared the expression of some tagged proteins expressed under the control of their own promoter or the CaMV 35S promoter and compared the accumulation of some tagged proteins and the corresponding endogenous proteins. They always observed a higher accumulation of the tagged proteins expressed under the strong CaMV 35S promoter, but they observed different patterns of accumulation between the endogenous and the tagged proteins. The authors applied the whole procedure to six known core cell cycle proteins. They could identify 42 protein–protein molecular associations: 28 were new interactions and 14 already known interactions [44]. This team recently published another article where this TAP approach was applied to approximately 100 cell cycle proteins, most of them belonging to the cell cycle core list [46]. This impressive work permitted to map the cell cycle interactome in *Arabidopsis*, providing new insights into the cell division mechanisms in plants.

The TAP approach proved to be successful in several cases but publications using this method are still relatively

rare in plant research compared with other organisms. An explanation could be the generally low purification yield of the tagged protein when stable plant transformants are used. Besides, to maintain fragile protein–protein interactions within the complex, covalent cross-linking between the bait and its partners can be carried out before or during cell lysis [37, 47, 48]. It is finally worth mentioning that affinity purification sometimes fails to isolate target proteins or to co-purify interaction partners. This problem is often due to the fact that the tag may interfere with bait folding or affect the interaction with the partners. However, as can be seen from the large-scale studies described below, the success rate of complex purifications is high.

3 Large-scale identification of affinity-purified complex components using MS

TAP and MS have been combined for the large-scale analysis of protein complexes within cells such as *Escherichia coli* [49], *S. cerevisiae* [32, 33, 50] or *Drosophila melanogaster* [51]. Medium-scale analysis was also performed in whole plants and plant cell cultures as described above. Similarly, single-step affinity purification proved successful in selecting hundreds of tagged proteins, together with potential interaction partners in vivo, from *E. coli* [52], yeast [53] and human cells [54]. Recent medium-scale studies using affinity purification methods deciphered in detail the interaction web around specific classes of proteins. Because other high-throughput studies were not exhaustive in mapping the interactions around kinases and phosphatases, protein complexes formed around these two types of enzymes were specifically studied to build a densely connected interaction network in yeast [55]. Hutchins et al. combined different techniques, among which TAP, to characterize 100 mitotic protein complexes in human cells, thus shedding light on the mechanisms implicated in chromosome segregation [56].

Affinity-purified samples are often separated by SDS-PAGE from which either the lanes are cut into slices over the entire separation range [32, 50], or only bands are taken which are specifically detected in the samples of interest [49]. Upon protease digestion, protein identification is then obtained by peptide mass fingerprinting (PMF) using a MALDI-TOF instrument and/or by LC-MS/MS analysis for peptide sequence information [32, 33, 50, 52–54]. The combination of PMF and sequence data allows higher stringency on protein identification, whereas LC-MS/MS analyses are particularly useful in analyzing co-migrating proteins in the same gel slice. As an alternative to, or associated with the SDS-PAGE-PMF approach, affinity-purified samples can be digested before proteolytic peptide mixtures are analyzed by LC-MS/MS [33, 49]. The latter gel-free strategy is less time consuming and allows higher peptide recovery. This method is more suitable when

analyzing TAP-purified samples, whose limited complexity usually allows nearly complete identification of its components. Since single-step IPs can provide tens or hundreds of proteins, direct LC-MS/MS analysis of whole digested samples precludes reliable identification of bait partners: the MS instrument duty cycle may not allow fragmenting all eluted peptides and partner finding may resemble searching a needle in a haystack. Here, differential (semi-quantitative) MS analysis appears indispensable to discriminate true bait interaction partners from background contaminants (see Section 4).

Large-scale approaches cannot systematically confirm the interaction by orthogonal assays, such as co-localization, co-immunoprecipitation or RNA interference assays. To determine the experimental background, purifications from mock-transformed control strains are often performed [32, 52, 54]. In addition, recurrently occurring proteins are usually excluded as background proteins if they fulfill the following criteria: (i) the proteins are co-purified with more than 2.5–5% of the baits; (ii) the proteins are major structural components (e.g. ribosomal proteins, histones, tubulins, etc.) [32, 33, 50, 53, 54]. On the other hand, protein interactions are deemed reliable when present in multiple biological repeats and/or found in reciprocal sample purifications targeting different complex constituents [49, 54]. Besides using such simple thresholds, statistical models were more recently designed: Breitkreutz et al. thus compared the number of peptides identifying a putative bait partner to a probability distribution to determine its likelihood of being a true interactor [55].

4 In-depth characterization of affinity-purified protein complexes

Previous large-scale protein complex analyses distinguished bait partners from contaminants by empirical arguments. However, some interaction partners can also be lost, when detected in numerous affinity-purified samples (e.g. ribosomal proteins might be valid interactors) [57]. This problem can be solved by measuring the relative abundance of all proteins in a given complex as compared with a control prepared in parallel. Such measurements require semi-quantitative MS analysis methods. In addition, previous experiments provided a static view of the protein complex in a given cell type under certain conditions (growth, etc). Knowledge on the dynamics of the modular protein organization is of high value, allowing us to address the following questions: what is the spatial and temporal plasticity of each complex? What are the PTMs in the components? Are the PTMs correlated with the variable complex composition? Which interactions are stable and which are dynamic or transient? What is the stoichiometry of the partners in the complex? Which of the constitutive proteins are in physical contact?

4.1 Dynamics of protein complex composition and PTMs

Below we provide an overview of procedures to remove contaminants and monitor the variable composition and modifications of protein complexes. Several procedures were designed to allow statistically retrieving the components of a partially purified complex from the majority of background proteins. The core idea consists of purifying side by side two samples, one being obtained from cells expressing the tagged protein of interest, and the other one from cells expressing an empty vector. The two samples are then subjected to semi-quantitative MS analysis for estimating the relative abundance of all identified proteins. First, the two samples can be differentially labeled with isotope-coded reagents, in light (L) or heavy (H) forms, mixed and finally analyzed by LC-MS/MS in their proteolyzed form. Contaminants are present in equal amounts in both samples, then their proteolytic peptides all lead to relative abundance ratios of theoretical value 1. In contrast, the bait partners are enriched in the sample of interest, and their proteolytic peptides yield H/L ratios statistically appearing as outliers from the main ratio distribution centered on 1. Variations in complex composition, triggered by different treatments or cell states, can be similarly assessed. The original studies describing such strategies based on semi-quantitative MS used differential ICAT [58], or SILAC labeling [59]. Numerous reports utilized such strategies to monitor changes in protein complex compositions [60–62], whereas the vast majority of studies purified a tagged version of the protein of interest, Mann et al. tested a combination of SILAC labeling, RNA interference, immuno-purification of the endogenous protein and semi-quantitative MS to identify interaction partners of baits in mammalian cells [63]. This method allowed to alleviate the biases introduced by the tag or bait over-expression.

Differential labeling of affinity-purified samples followed by semi-quantitative MS analysis renders single-step IP attractive in the scope of large-scale protein interaction studies, due to their shorter preparation times and possibly better preserved weaker interactions. However, the use of labeling reagents represents a significant financial investment which encouraged the development of strategies relying on label-free comparative MS. Considering the fact that the total number of peptides identifying a given protein correlates with its abundance in a sample, Washburn et al. defined and used normalized spectral abundance factors to determine constituents of a chromatin remodeling complex and build a probabilistic local interaction network [64]. This approach is feasible if proteins are repetitively identified at a relatively constant number of peptides, which requires that the complexity of the analyzed sample does not overwhelm the duty cycle of the LC-MS system. An alternative label-free strategy consists of reconstituting the MS-detected chromatographic elution peak of

every peptide in the different compared samples: the ratios of peak areas between sample of interest and control are averaged over all peptides identifying a given protein, to yield the relative abundance of that protein. Such an approach was employed to determine the actual complex constituents and their variations upon stimulus [65]. Although a priori highly flexible and amenable to large-scale protein complex analysis, this label-free strategy suffers from insufficiently robust software (improper extraction of chromatographic peak areas, for example due to overlapping isotopic patterns or to non-linear drifts in retention times) and requires a very careful statistical treatment of semi-quantitative data.

To characterize protein complexes more deeply, it can be interesting to distinguish between stably associated and dynamically exchanged protein constituents. An astute strategy to address this question was recently described [66–68]. Protein samples are affinity purified from control cells and cells expressing the tagged bait that were differentially SILAC labeled. Affinity selection is performed either before or after pooling of the cell extracts. Proteins whose abundance changes between both protocols appear to be dynamically exchanged bait interactors. A refined description of the transcription complexes containing TBPs showed that one partner was exchanged during affinity purification, except during cell mitosis [68].

Finally, because biologically significant interactions are dynamic and regulated by covalent modifications, analytical workflows were developed to enhance detection of post-translationally modified regions of complex constituents, among which phosphorylated sequences have received more attention [59, 69, 70]. To enhance the detection of phosphorylated sequences within digested complexes, immobilized metal affinity chromatography (IMAC) and strong cation exchange (SCX) fractionation are frequently used. Thus, Smolka et al. could identify and quantify both dynamic protein associations and phosphorylations following DNA damage [71]. TiO₂ enrichment of phosphopeptides was more recently introduced as a powerful alternative and was for example utilized to study the cell-cycle dependency of TBP-containing complexes and their phosphorylations [72]. Beside protein phosphorylation, the impact of other PTMs on protein complex formation was specifically probed. Among them, methylation and acetylation of histone proteins have lately gained increased attention. Vermeulen et al. thus very recently reported a detailed interaction screen of the activating and repressive trimethyl marks on histones H3 and H4 [73]. A more complete picture of protein complexes would finally be attained if all aspects – determination of changes in complex constituents upon stimulation and identification of static/variable PTMs of all types – could be analyzed. Links to pathways might thus be highlighted (e.g. connections with particular kinases, with the proteasome pathway, etc.). Given the complexity of PTMs, this remains, however, a challenge.

4.2 Stoichiometry and direct protein–protein contacts in protein complexes

The vast majority of studies aiming to characterize protein complexes remained elusive as to which partners are in direct physical contact or which stoichiometry exists in the protein complex. The identity of proteins in direct interaction within a complex and their binding interfaces can be probed using cross-linking agents. A variety of chemical reagents targeting different protein functions have been designed and successfully used to get insights into the three-dimensional structure of particular protein complexes [74]. However, only very few attempts were made to test cross-linking on large data sets: the challenge resides in great part in the fact that cross-linked species are of low stoichiometry relative to unmodified peptides. Rinner et al. developed a strategy that allows the identification of cross-linked peptides from total cell lysates [75]. The main features of the procedure consist of the use of isotopically coded cross-linking agents, to more efficiently point to peptides bearing this tag, and a specifically designed search engine xQuest. Alternatively, a procedure using newly designed cross-linkers, whose structure contains specific MS-cleavable bonds, a mass-encoded reporter group and an affinity tag, was applied to a non-targeted analysis of *Shewanella oneidensis* [76]. These recent procedures may open the way to a more systematic study of direct protein–protein interactions in partially purified complexes containing hundreds of different proteins.

The stoichiometry of protein complexes is an additional sparsely addressed question. A few procedures have been used to determine the absolute abundance of some specific proteins: they rely on the use of heavy-isotope-coded peptides [77, 78], a concatenated protein made up of tryptic peptides of the target proteins [79, 80] or heavy-labeled full-length proteins of interest [81, 82]. These strategies require, for each targeted protein, to synthesize isotope-coded sequences and/or determining a few proteotypic peptides providing its specific identification. They thus appear hardly applicable to large-scale studies of complexes. In contrast, a promising strategy toward this goal was recently described: complex affinity purification is performed while decorating the bait with a tag containing a tryptic peptide sequence; a heavy form of that peptide is introduced after sample isolation from the cell extract [83]. After measuring the absolute amount of an affinity-purified bait protein, the MS signals detected for its proteolytic peptides can be used as a calibration curve to deduce its amount as prey in other affinity-purified samples. This elegant strategy requires targeting several members of the same complex as bait; this constraint was already reached in large-scale studies [32]. Holzmann et al. described an alternative cost-effective strategy that utilizes synthesized peptides corresponding to the association of a constant non-naturally occurring tryptic sequence (equalizer peptide) and tryptic sequences from the complex constituents. After digestion,

the equalizer peptide enables mixing all standard peptides in equimolar ratios. Finally, differential labeling of this reference mixture and of the digested protein complex of interest with mTRAQ reagents allows determining complex stoichiometry [84].

4.3 Analysis of intact complexes by MS

We chose to focus in this review on strategies of protein complex characterization, which rely on the classical proteomic analysis of the samples of interest in a proteolyzed form. However, it is worth mentioning that the aspects described above of dynamic protein complex composition, complex stoichiometry and direct protein–protein contacts were elegantly and very successfully addressed by the MS analysis of whole intact protein complexes, as extensively reviewed in [11]. Various protein complexes, of high purity and composed of proteins of known identity, could thus be characterized. Such MS analyses allowed determining the stoichiometry of multi-protein complexes made up of identical subunits and highlighting the polydisperse nature of complexes. The dynamic interactions, the exchange of subunits and the assembly process of protein complexes could be monitored in real-time during MS analysis. The dissociation patterns observed in MS/MS spectra inform on the subunit stoichiometry. However, when performing MS/MS of the whole complex under study, only peripheral subunits are released. To get further insights into the interactions between subunits, chemical cross-linking or the partial disruption of complexes using chaotropic agents to generate various subcomplexes can be obtained in solution before MS and MS/MS analyses. Although these approaches require the complex of interest to be highly purified from biologically non-related proteins, it could be applied to the analysis of several protein complexes purified by TAP from *S. cerevisiae* to establish detailed three-dimensional interaction maps [85]. Once the constituents of a protein complex have been identified by the classical proteomic analysis, as described above, a refined knowledge of its structure (stoichiometry of the partners, direct contacts, proteins located at the periphery/in the core of the assembly) may then be obtained from MS analysis of the intact complex.

5 Concluding remarks

Protein complexes are part of almost all the processes occurring in cells. Their identification and characterization are hence necessary to fully understand the chemical and physical reactions that finally constitute cellular dynamics and homeostasis. In the last two decades, new technical and instrumental advances permitted huge progress in the study of protein complexes in model organisms, including plants. Even if the picture is far from exhaustive, the identification of proteins and their physiological partners has well advanced. However, the dynamic characterization of protein

complexes, their PTM identification, and above all the structure–function relation of these protein assemblies are still poorly understood. This is particularly true for organisms like crop plants where much effort is necessary to answer all these questions.

The authors have declared no conflict of interest.

6 References

- [1] Green, D. E., Studies in organized enzyme systems. *Harvey Lect.* 1958, 177–227.
- [2] Srere, P. A., Macromolecular interactions: tracing the roots. *Trends Biochem. Sci.* 2000, 25, 150–153.
- [3] Yanofsky, C., Rachmeler, M., The exclusion of free indole as an intermediate in the biosynthesis of tryptophan in *Neurospora crassa*. *Biochim. Biophys. Acta* 1958, 28, 640–641.
- [4] Winkel, B. S., Metabolic channeling in plants. *Annu. Rev. Plant Biol.* 2004, 55, 85–107.
- [5] Graham, J. W., Williams, T. C., Morgan, M., Fernie, A. R. et al., Glycolytic enzymes associate dynamically with mitochondria in response to respiratory demand and support substrate channeling. *Plant Cell* 2007, 19, 3723–3738.
- [6] Droux, M., Ruffet, M. L., Douce, R., Job, D., Interactions between serine acetyltransferase and *O*-acetylserine (thiol) lyase in higher plants – structural and kinetic properties of the free and bound enzymes. *Eur. J. Biochem.* 1998, 255, 235–245.
- [7] Gontero, B., Cardenas, M. L., Ricard, J., A functional five-enzyme complex of chloroplasts involved in the Calvin cycle. *Eur. J. Biochem.* 1988, 173, 437–443.
- [8] Graciet, E., Gans, P., Wedel, N., Lebreton, S. et al., The small protein CP12: a protein linker for supramolecular complex assembly. *Biochemistry* 2003, 42, 8163–8170.
- [9] Suzuki, J. Y., Ytterberg, A. J., Beardslee, T. A., Allison, L. A. et al., Affinity purification of the tobacco plastid RNA polymerase and in vitro reconstitution of the holoenzyme. *Plant J.* 2004, 40, 164–172.
- [10] Asahara, H., Chong, S., In vitro genetic reconstruction of bacterial transcription initiation by coupled synthesis and detection of RNA polymerase holoenzyme. *Nucleic Acids Res.* 2010, 38, e141.
- [11] Sharon, M., Robinson, C. V., The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu. Rev. Biochem.* 2007, 76, 167–193.
- [12] Ramani, A. K., Li, Z. H., Hart, G. T., Carlson, M. W. et al., A map of human protein interactions derived from co-expression of human mRNAs and their orthologs. *Mol. Syst. Biol.* 2008, 4.
- [13] Dong, M., Yang, L. L., Williams, K., Fisher, S. J. et al., A “tagless” strategy for identification of stable protein complexes genome-wide by multidimensional orthogonal chromatographic separation and iTRAQ reagent tracking. *J. Proteome Res* 2008, 7, 1836–1849.
- [14] Liu, X. P., Yang, W. C., Gao, Q., Regnier, F., Toward chromatographic analysis of interacting protein networks. *J. Chromatogr.* 2008, 1178, 24–32.
- [15] Olinares, P. D., Ponnala, L., van Wijk, K. J., Megadalton complexes in the chloroplast stroma of *Arabidopsis thaliana* characterized by size exclusion chromatography, mass spectrometry, and hierarchical clustering. *Mol. Cell. Proteomics* 2010, 9, 1594–1615.
- [16] Remmerie, N., Roef, L., Van De Slijke, E., Van Leene, J. et al., A bioanalytical method for the proteome wide display and analysis of protein complexes from whole plant cell lysates. *Proteomics* 2009, 9, 598–609.
- [17] Kikuchi, S., Oishi, M., Hirabayashi, Y., Lee, D. W. et al., A 1-megadalton translocation complex containing Tic20 and Tic21 mediates chloroplast protein import at the inner envelope membrane. *Plant Cell* 2009, 21, 1781–1797.
- [18] Balabaskaran Nina, P., Dudkina, N. V., Kane, L. A., van Eyk, J. E. et al., Highly divergent mitochondrial ATP synthase complexes in *Tetrahymena thermophila*. *PLoS Biol.* 2010, 8, e1000418.
- [19] Wang, X., Chen, G., Liu, H., Zhao, Z., Li, Z., Four-dimensional orthogonal electrophoresis system for screening protein complexes and protein–protein interactions combined with mass spectrometry. *J. Proteome Res.*
- [20] Munro, S., Pelham, H. R. B., Use of peptide tagging to detect proteins expressed from cloned genes: deletion mapping functional domains of *Drosophila* hsp70. *EMBO J.* 1984, 3, 3087–3093.
- [21] Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M. et al., A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* 1999, 17, 1030–1032.
- [22] Kocher, T., Superti-Furga G: mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat. Methods* 2007, 4, 807–815.
- [23] Terpe, K., Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* 2003, 60, 523–533.
- [24] Hochuli, E., Bannwarth, W., Dobeli, H., Gentz, R., Stuber, D., Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. *Bio-Technology* 1988, 6, 1321–1325.
- [25] Hochuli, E., Dobeli, H., Schacher, A., New metal chelate adsorbent selective for proteins and peptides containing neighboring histidine-residues. *J. Chromatogr.* 1987, 411, 177–184.
- [26] Smith, D. B., Johnson, K. S., Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione *S*-transferase. *Gene* 1988, 67, 31–40.
- [27] Diguan, C., Li, P., Riggs, P. D., Inouye, H., Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. *Gene* 1988, 67, 21–30.
- [28] StofkoHahn, R. E., Carr, D. W., Scott, J. D., A single step purification for recombinant proteins – characterization of a microtubule associated protein (map-2) fragment which

- associates with the type-II camp-dependent protein-kinase. *FEBS Lett.* 1992, *302*, 274–278.
- [29] Voss, S., Skerra, A., Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the strep-tag II peptide and improved performance in recombinant protein purification. *Protein Eng.* 1997, *10*, 975–982.
- [30] Book, A. J., Gladman, N. P., Lee, S. S., Scalf, M. et al., Affinity purification of the *Arabidopsis* 26S proteasome reveals a diverse array of plant proteolytic complexes. *J. Biol. Chem.* 2010, *285*, 25554–25569.
- [31] Goudreault, M., D'Ambrosio, L. M., Kean, M. J., Mullin, M. J. et al., A PP2A phosphatase high density interaction network identifies a novel striatin-interacting phosphatase and kinase complex linked to the cerebral cavernous malformation 3 (CCM3) protein. *Mol. Cell. Proteomics* 2009, *8*, 157–171.
- [32] Gavin, A. C., Bosche, M., Krause, R., Grandi, P. et al., Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 2002, *415*, 141–147.
- [33] Krogan, N. J., Cagney, G., Yu, H., Zhong, G. et al., Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 2006, *440*, 637–643.
- [34] Gully, D., Moinier, D., Loiseau, L., Bouveret, E., New partners of acyl carrier protein detected in *Escherichia coli* by tandem affinity purification. *FEBS Lett.* 2003, *548*, 90–96.
- [35] Knuesel, M., Wan, Y., Xiao, Z., Holinger, E. et al., Identification of novel protein–protein interactions using a versatile mammalian tandem affinity purification expression system. *Mol. Cell. Proteomics* 2003, *2*, 1225–1233.
- [36] Forler, D., Kocher, T., Rode, M., Gentzel, M. et al., An efficient protein complex purification method for functional proteomics in higher eukaryotes. *Nat. Biotechnol.* 2003, *21*, 89–92.
- [37] Rohila, J. S., Chen, M., Cerny, R., Fromm, M. E., Improved tandem affinity purification tag and methods for isolation of protein heterocomplexes from plants. *Plant J.* 2004, *38*, 172–181.
- [38] Rivas, S., Romeis, T., Jones, J. D., The Cf-9 disease resistance protein is present in an approximately 420-kilodalton heteromultimeric membrane-associated complex at one molecule per complex. *Plant Cell* 2002, *14*, 689–702.
- [39] Zhong, J., Haynes, P. A., Zhang, S., Yang, X. et al., Development of a system for the study of protein–protein interactions in planta: characterization of a TATA-box binding protein complex in *Oryza sativa*. *J. Proteome Res.* 2003, *2*, 514–522.
- [40] Li, Y., Commonly used tag combinations for tandem affinity purification. *Biotechnol. Appl. Biochem.* 2010, *55*, 73–83.
- [41] Rubio, V., Shen, Y., Saijo, Y., Liu, Y. et al., An alternative tandem affinity purification strategy applied to *Arabidopsis* protein complex isolation. *Plant J.* 2005, *41*, 767–778.
- [42] Rohila, J. S., Chen, M., Chen, S., Chen, J. et al., Protein–protein interactions of tandem affinity purification-tagged protein kinases in rice. *Plant J.* 2006, *46*, 1–13.
- [43] Rohila, J. S., Chen, M., Chen, S., Chen, J. et al., Protein–protein interactions of tandem affinity purified protein kinases from rice. *PLoS One* 2009, *4*, e6685.
- [44] Van Leene, J., Stals, H., Eeckhout, D., Persiau, G. et al., A tandem affinity purification-based technology platform to study the cell cycle interactome in *Arabidopsis thaliana*. *Mol. Cell. Proteomics* 2007, *6*, 1226–1238.
- [45] Van Leene, J., Witters, E., Inze, D., De Jaeger, G., Boosting tandem affinity purification of plant protein complexes. *Trends Plant Sci.* 2008, *13*, 517–520.
- [46] Van Leene, J., Hollunder, J., Eeckhout, D., Persiau, G. et al., Targeted interactomics reveals a complex core cell cycle machinery in *Arabidopsis thaliana*. *Mol. Syst. Biol.* 2010, *6*, 397.
- [47] Stingl, K., Schauer, K., Ecobichon, C., Labigne, A. et al., In vivo interactome of helicobacter pylori urease revealed by tandem affinity purification. *Mol. Cell. Proteomics* 2008, *7*, 2429–2441.
- [48] Nittis, T., Guittat, L., LeDuc, R. D., Dao, B. et al., Revealing novel telomere proteins using in vivo cross-linking, tandem affinity purification, and label-free quantitative LC-FTICR-MS. *Mol. Cell. Proteomics* 2010, *9*, 1144–1156.
- [49] Butland, G., Peregrin-Alvarez, J. M., Li, J., Yang, W. et al., Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* 2005, *433*, 531–537.
- [50] Gavin, A. C., Aloy, P., Grandi, P., Krause, R. et al., Proteome survey reveals modularity of the yeast cell machinery. *Nature* 2006, *440*, 631–636.
- [51] Giot, L., Bader, J. S., Brouwer, C., Chaudhuri, A. et al., A protein interaction map of *Drosophila melanogaster*. *Science* 2003, *302*, 1727–1736.
- [52] Arifuzzaman, M., Maeda, M., Itoh, A., Nishikata, K. et al., Large-scale identification of protein–protein interaction of *Escherichia coli* K-12. *Genome Res.* 2006, *16*, 686–691.
- [53] Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D. et al., Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 2002, *415*, 180–183.
- [54] Ewing, R. M., Chu, P., Elisma, F., Li, H. et al., Large-scale mapping of human protein–protein interactions by mass spectrometry. *Mol. Syst. Biol.* 2007, *3*, 89.
- [55] Breitkreutz, A., Choi, H., Sharom, J. R., Boucher, L. et al., A global protein kinase and phosphatase interaction network in yeast. *Science* 2010, *328*, 1043–1046.
- [56] Hutchins, J. R., Toyoda, Y., Hegemann, B., Poser, I. et al., Systematic analysis of human protein complexes identifies chromosome segregation proteins. *Science* 2010, *328*, 593–599.
- [57] Bouwmeester, T., Bauch, A., Ruffner, H., Angrand, P. O. et al., A physical and functional map of the human TNF- α /NF- κ B signal transduction pathway. *Nat. Cell Biol.* 2004, *6*, 97–105.
- [58] Ranish, J. A., Yi, E. C., Leslie, D. M., Purvine, S. O. et al., The study of macromolecular complexes by quantitative proteomics. *Nat. Genet.* 2003, *33*, 349–355.

- [59] Blagoev, B., Kratchmarova, I., Ong, S. E., Nielsen, M. et al., A proteomics strategy to elucidate functional protein–protein interactions applied to EGF signaling. *Nat. Biotechnol.* 2003, 21, 315–318.
- [60] Brand, M., Ranish, J. A., Kummer, N. T., Hamilton, J. et al., Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics. *Nat. Struct. Mol. Biol.* 2004, 11, 73–80.
- [61] Corvey, C., Koetter, P., Beckhaus, T., Hack, J. et al., Carbon source-dependent assembly of the Snf1p kinase complex in *Candida albicans*. *J. Biol. Chem.* 2005, 280, 25323–25330.
- [62] Kaake, R. M., Milenkovic, T., Przulj, N., Kaiser, P., Huang, L., Characterization of cell cycle specific protein interaction networks of the yeast 26S proteasome complex by the QTAX strategy. *J. Proteome Res.* 2010, 9, 2016–2029.
- [63] Selbach, M., Mann, M., Protein interaction screening by quantitative immunoprecipitation combined with knock-down (QUICK). *Nat. Methods* 2006, 3, 981–983.
- [64] Sardi, M. E., Cai, Y., Jin, J., Swanson, S. K. et al., Probabilistic assembly of human protein interaction networks from label-free quantitative proteomics. *Proc. Natl. Acad. Sci. USA* 2008, 105, 1454–1459.
- [65] Rinner, O., Mueller, L. N., Hubalek, M., Muller, M. et al., An integrated mass spectrometric and computational framework for the analysis of protein interaction networks. *Nat. Biotechnol.* 2007, 25, 345–352.
- [66] Kito, K., Kawaguchi, N., Okada, S., Ito, T., Discrimination between stable and dynamic components of protein complexes by means of quantitative proteomics. *Proteomics* 2008, 8, 2366–2370.
- [67] Wang, X. R., Huang, L., Identifying dynamic interactors of protein complexes by quantitative mass spectrometry. *Mol. Cell. Proteomics* 2008, 7, 46–57.
- [68] Mousson, F., Kolkman, A., Pijnappel, W., Timmers, H. T. M., Heck, A. J. R., Quantitative proteomics reveals regulation of dynamic components within TATA-binding protein (TBP) transcription complexes. *Mol. Cell. Proteomics* 2008, 7, 845–852.
- [69] Pflieger, D., Junger, M. A., Muller, M., Rinner, O. et al., Quantitative proteomic analysis of protein complexes: concurrent identification of interactors and their state of phosphorylation. *Mol. Cell. Proteomics* 2008, 7, 326–346.
- [70] Przybylski, C., Junger, M. A., Aubertin, J., Radvanyi, F. et al., Quantitative analysis of protein complex constituents and their phosphorylation states on a LTQ-Orbitrap Instrument. *J. Proteome Res.* 2010.
- [71] Smolka, M. B., Albuquerque, C. P., Chen, S. H., Schmidt, K. H. et al., Dynamic changes in protein–protein interaction and protein phosphorylation probed with amine-reactive isotope tag. *Mol. Cell. Proteomics* 2005, 4, 1358–1369.
- [72] Pijnappel, W. P., Kolkman, A., Baltissen, M. P., Heck, A. J., Timmers, H. M., Quantitative mass spectrometry of TATA binding protein-containing complexes and subunit phosphorylations during the cell cycle. *Proteome Sci.* 2009, 7, 46.
- [73] Vermeulen, M., Eberl, H. C., Matarese, F., Marks, H. et al., Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* 2010, 142, 967–980.
- [74] Petrotchenko, E. V., Borchers, C. H., Crosslinking combined with mass spectrometry for structural proteomics. *Mass. Spectrom. Rev.* 2010.
- [75] Rinner, O., Seebacher, J., Walzthoeni, T., Mueller, L. N. et al., Identification of cross-linked peptides from large sequence databases. *Nat. Methods* 2008, 5, 315–318.
- [76] Zhang, H., Tang, X., Munske, G. R., Tolic, N. et al., Identification of protein–protein interactions and topologies in living cells with chemical cross-linking and mass spectrometry. *Mol. Cell. Proteomics* 2009, 8, 409–420.
- [77] Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., Gygi, S. P., Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. USA* 2003, 100, 6940–6945.
- [78] Schmidt, C., Lenz, C., Grote, M., Luhrmann, R., Urlaub, H., Determination of protein stoichiometry within protein complexes using absolute quantification and multiple reaction monitoring. *Anal. Chem.* 2010, 82, 2784–2796.
- [79] Kito, K., Ota, K., Fujita, T., Ito, T., A synthetic protein approach toward accurate mass spectrometric quantification of component stoichiometry of multiprotein complexes. *J. Proteome Res.* 2007, 6, 792–800.
- [80] Pratt, J. M., Simpson, D. M., Doherty, M. K., Rivers, J. et al., Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nat. Protoc.* 2006, 1, 1029–1043.
- [81] Brun, V., Dupuis, A., Adrait, A., Marcellin, M. et al., Isotope-labeled protein standards: toward absolute quantitative proteomics. *Mol. Cell. Proteomics* 2007, 6, 2139–2149.
- [82] Hanke, S., Besir, H., Oesterhelt, D., Mann, M., Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level. *J. Proteome Res.* 2008, 7, 1118–1130.
- [83] Wepf, A., Glatter, T., Schmidt, A., Aebersold, R., Gstaiger, M., Quantitative interaction proteomics using mass spectrometry. *Nat. Methods* 2009, 6, 203–205.
- [84] Holzmann, J., Pichler, P., Madalinski, M., Kurzbauer, R., Mechtler, K., Stoichiometry determination of the MP1-p14 complex using a novel and cost-efficient method to produce an equimolar mixture of standard peptides. *Anal. Chem.* 2009, 81, 10254–10261.
- [85] Hernandez, H., Dziembowski, A., Taverner, T., Seraphin, B., Robinson, C. V., Subunit architecture of multimeric complexes isolated directly from cells. *EMBO Rep.* 2006, 7, 605–610.