transactions

Proteomic Complex Detection using Sedimentation (ProCoDeS): screening for proteins in stable complexes and their candidate interaction partners

Marcelo P. Segura, Kathryn S. Lilley and Paul Dupree[1]
Department of Biochemistry, University Of Cambridge, Tennis Court Road, Cambridge CB2 1QW, U.K.

Abstract

Over the last few years, our view of cellular organization has changed from one in which enzymes and proteins usually act independently to the situation at present where we commonly accept that many, if not all, enzymes act in close association with others. Co-precipitation using an antibody against a test protein is the standard assay for the identification of members of protein complexes [Musso, Zhang and Emili (2007) Chem. Rev. 107, 3585–3600]. The introduction of TAP (tandem affinity purification) tagging enhanced original approaches in order to analyse protein complexes on a larger scale with reduced false discoveries of interacting partners due to more efficient purification of complexes. However, this technique has some limitations as a high-throughput tool for systems biology: the requirement for genetic manipulation to express the tagged protein excludes studies of non-transformable organisms and intact tissue. In those cases where TAP is applicable, a considerable amount of work is required to generate the baits and to optimize experimental conditions. A technique developed in our laboratories, ProCoDeS (Proteomic Complex Detection using Sedimentation), focuses on the detection of endogenous complexes. Protein samples are separated by centrifugation and then different fractions from the resulting gradient are analysed using quantitative MS. The identification of possible protein partners is based on statistical analysis of the co-fractionation of proteins, without any need for purification of individual complexes. The prospects of ProCoDeS and similar techniques based on quantitative MS for measurement of protein complex composition are reviewed in the present article.

Limitations to current affinity-purification methods

From the perspective of method scaling, the generation of specific antibodies required for affinity purification of complexes becomes rapidly unattainable as the number of surveyed proteins increases. A more universal approach which overcomes the need for large-scale antibody production involves the expression of engineered bait fusion proteins with generic peptide tags. TAP (tandem affinity purification) can be used to recover any bait and its binding partner proteins through two sequential affinity methods [1,2]. The dependence on genetic modification restricts the application of TAP methods to organisms where transformation protocols have been established and are amenable for implementation on a large scale, e.g. single-cell organisms and cell cultures. Even for those species where genetic manipulation is available, there are situations in which cell cultures are not a good surrogate for an entire organism. If we are specifically trying to address the variability of complexes across several tissues, conditions or subjects (particularly humans), then native samples may need to be analysed.

Affinity-purification methods emphasize the recovery of pure complexes and the development of strategies to capture and to select complexes from a sample in a high-throughput manner. The large-scale determination of protein complexes with TAP [3–5] seemed to have paved the way for the discovery of the full set of complexes of yeast and other organisms. However, the high rate of false interactions and the production of tagged proteins raise new issues. The spurious interaction of background proteins with antibodies, affinity matrices or the bait protein is well documented [6]. The identification of false-positive interactions has been partly addressed by the use of repeated purification of complexes using different baits and the inclusion of controls without baits to enable the identification of 'sticky' proteins that frequently contaminate the affinity-purification experiments. The destabilization of the native interaction with the tag is another potential issue which has been addressed in some cases by the use of alternative C- or N-terminal tags. Spurious interactions can thus be corrected to a certain extent with multiple controls and experiment repetition. A 'socio-affinity' index for pairs of proteins, derived by Gavin et al. [5], is related to the probability of a particular interaction from

Key words: blue native gel electrophoresis, high-throughput screen, Pearson correlation, protein–protein interaction, Proteomic Complex Detection using Sedimentation (ProCoDeS), tandem affinity purification (TAP).

Abbreviations used: BN, blue native gel electrophoresis; ICAT, isotope-coded affinity tag; iTRAQ, isobaric tag for relative and absolute quantification; MS/MS, tandem MS; TAP, tandem affinity purification.

[1]To whom correspondence should be addressed (email p.dupree@bioc.cam.ac.uk).

Biochemical Society Transactions www.biochemsoctrans.org

their observed frequency in reciprocal experiments. In effect, the high-throughput nature of the tagging experiments leads to a prediction of protein–protein interactions and associated confidence levels. The cost of generating more accurate maps of interactions can be high. For the determination of protein interactions in yeast, 6466 clones were created, each of them carrying a single epitope-tagged protein [5]. Without the assistance of automated procedures, even a fraction of such a study is beyond the scope of most research groups.

Quantitative approaches to native protein complex characterization

Proteomic profiling provides an alternative approach to the investigation of protein-protein interactions. Using quantitative MS methods coupled with statistical analysis, it is possible to interrogate samples containing protein mixtures and to predict possible protein–protein interactions. Such techniques are beginning to emerge, but show great promise and rely heavily on robust quantitative proteomics that enable accurate protein profiling. Such methodologies involve large-scale detection and quantification of complexes using MS methods. The most common approaches to MS quantification of proteins involve the use of either differential stable isotopes or label-free technologies. In both cases, the abundance of a protein is calculated on the basis of MS measurements of peptides derived from proteolysis of proteins, generally using trypsin as the protease.

Stable isotope labelling involves quantification using differential incorporation of stable isotopes either *in vivo* or *in vitro*. *In vivo* labelling can be achieved in several ways. One method involves the growth of cultures in the presence of a defined medium containing a heavy elemental isotope, typically ^{15}N [7]. Samples grown in the presence of the natural isotope and the heavy isotope can be pooled and then reduced to peptides. The relative abundance of a peptide generated from a protein within cultures being compared is then calculated by measuring ion intensities of the 'light' and 'heavy' versions of the same peptide. A variant of the previous method is SILAC (stable isotope labelling of amino acids in culture), where isotopically modified amino acids are incorporated into the protein during growth [8–11].

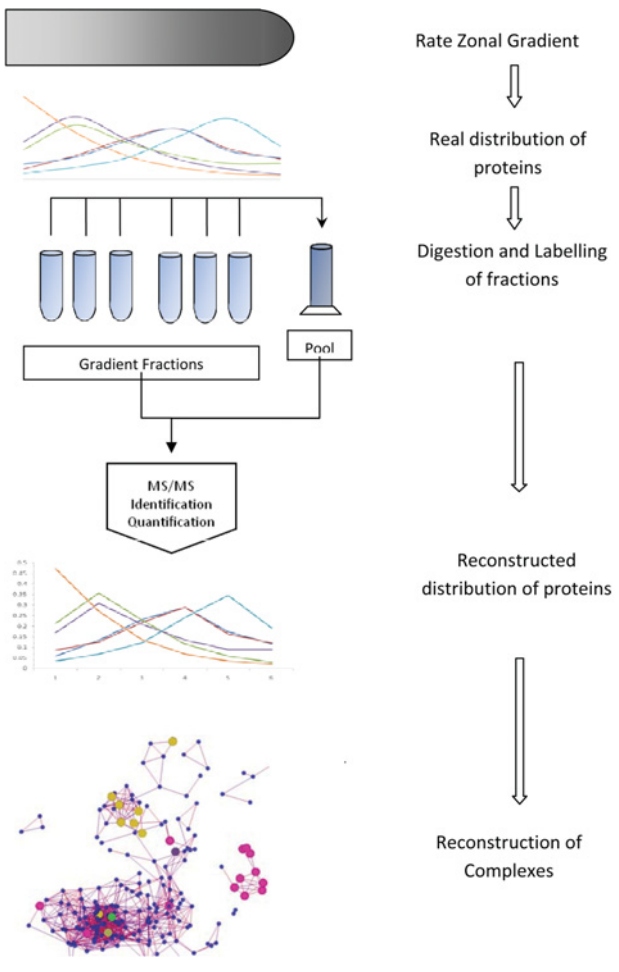
In vitro tagging versions of differential stable isotope labelling have also been used to great effect within quantitative proteomics. ICATs (isotope-coded affinity tags) can be used to chemically label cysteine residues of proteins before digestion to peptides [12–14]. The label exists in lighter or heavier form such that discrimination between peptides arising from two different samples is possible during MS analysis. The most commonly used tagging system, however, involves the use of the multiplexed set of four or eight isotopic tags of the iTRAQ (isobaric tag for relative and absolute quantification) system, which labels primary amino groups within peptides generated from extracted proteins [15]. Since the iTRAQ tags are isobaric, differentially labelled versions of a peptide appear as a single precursor ion peak. When an iTRAQ-labelled peptide is subjected to collision-induced dissociation in MS/MS (tandem MS) mode, the iTRAQ tags release diagnostic low-mass reporter ions that are used for quantification.

Label-free quantification is based entirely on peak intensity measurements of peptides detected by MS or on the number of ions per protein (spectral counts) detected in a MS experiment [16].

We have developed a technique, ProCoDeS (Proteomic Complex Detection using Sedimentation), for the identification of membrane protein complexes (Figure 1) [17]. ProCoDeS is designed to identify, using quantitative MS analysis, proteins in tissue homogenates that more quickly sediment in a rate-zonal ultracentrifugation gradient and are therefore likely to be in high-molecular-mass complexes.

Figure 1 | Workflow of protein complex identification with ProCoDeS

Membrane protein samples are separated through a rate zonal gradient. The quantification of the distribution of proteins is achieved through MS/MS with isotopic labels. The resulting protein traces can be used to discover proteins that sediment as complexes and also to estimate the most likely complex partner for each protein in the experiment.



Co-migrating proteins in the gradient can then be identified, leading to predictions of possible protein complex compositions. Consecutive fractions of the gradient are quantified against a common pool using MS. This approach has been validated by showing that members of the different complexes of the mitochondrial respiratory chain of *Arabidopsis thaliana* co-migrate in groups according to their complex membership. Hartman et al. [17] demonstrated that protein profiles, as determined by quantitative MS using ICAT peptide labelling, could be used to estimate the peak of protein sedimentation in the gradient of approx. 200 proteins in an analogous fashion to a chromatogram. More recently, we have employed Pearson correlation as a measure of similarity, which rendered similar predictions about possible protein interactions in the sample (M.P. Segura and P. Dupree, unpublished work). Unlike many previous methods to study protein complexes, the detection of proteins in a complex in ProCoDeS does not rely on the actual recovery of the complex from the sample and therefore no high level of purification is necessary. Many complexes in the sample can be detected in a single experiment. The number of fraction samples analysed and the number of proteins identified can be increased with the use of the multiplexing capability of iTRAQ (N.T. Hartman, K.S. Lilley and P. Dupree, unpublished work). A very important advantage in ProCoDeS is the possibility to process native samples without transformation of cells or organisms. The avoidance of expression of a fusion protein avoids the problems of artefacts derived from such overexpression. The interrogation of the sample is instead limited by the sensitivity of MS and the native protein complex abundance.

Recently, several techniques have been described that use similar principles to ProCoDeS. Dong et al. [18] used quantitative MS profiling to quantify soluble proteins from *Escherichia coli* extracts. Proteins were sequentially separated through anion exchange and gel-filtration columns, and proteins in the fractions were analysed by MALDI (matrix-assisted laser-desorption ionization)–TOF (time-of-flight) MS/MS of iTRAQ-labelled peptides. Overall, from approx. 100 polypeptides studied, the quantification of fractionation of 37 proteins in 13 complexes was achieved, including RNA polymerase and pyruvate dehydrogenase. A very high proportion of these proteins in complexes (95%) had very similar elution profiles, with an overlapping maximum intensity in the same or contiguous fractions. This result accounts for approx. 88% of the interactions defined by TAP experiments in *E. coli*. In order to develop an automated system of classification, a value of Pearson correlation equal or higher than 0.92 was used as a criterion to group protein profiles in a single complex. More than half of the complex standards were grouped correctly using this criterion.

In a second approach, homogenates of human cells were separated by ultracentrifugation, and the resulting fractions were analysed by MS using label-free quantification [19], resulting in profiles of over 3000 proteins. Among the identified complexes were the core components of RNA polymerase II, the TCP1 (T-complex polypeptide 1) chaperone complex and NADH dehydrogenase. The identification of complexes directly through correlation of protein profiles was not very effective. A correlation as low as 0.4 or above, was found for 63% of the known protein partners. However, the authors showed that the data from this proteomic experiment could be combined with other sources of information such as co-expression in order to produce a more accurate map of probable interactions [19]. The low ability of protein profiles to identify additional complex partners could be partly explained by the MS quantification system based on spectral counts. Label-free methods have increasing popularity, but may give more imprecise estimation of protein quantities than isotopic labelling methods because, without pooling of labelled samples early in the workflow, the technical variability is not shared [20].

BN (blue native gel electrophoresis) is another technique that has been widely used for the analysis of complexes [21,22], but its usefulness as a high-throughput tool is limited because of the dependence of BN on visual selection of protein complexes. An enhanced version of this technique was recently presented by Wessels et al. [23], where a whole lane is divided into thin slices and then characterized by shotgun proteomics. BN coupled to quantitative MS proved useful to identify complexes 1–5 from the respiratory chain of human kidney cells. A Pearson coefficient of 0.7 was used again as the grouping criterion for proteins. As in the study by Ramani et al. [19], protein profiles were measured using a label-free approach based on counting of the unique peptides detected during MS analysis.

Even if the grouping of proteins into complexes can be achieved through visual comparison of the protein profiles, large-scale analysis requires an objective quantitative method for automation and scaling of the technique. In most cases, Pearson correlation has been used as a measure of similarity of profiles, but the cut-off values described for the different studies vary significantly. Our experience with ProCoDeS supports the need for high correlation, 0.9 or more, in order to perform a satisfactory assignment of proteins into known complexes. The classification based on the reconstruction of peaks by Hartman et al. [17] or the pre-filtering based on apex co-elution before correlation as demonstrated by Dong et al. [18] show that other aspects of the data could be incorporated in the analysis. The fortuitous elution of non-related protein together with a real complex is one of the aspects that could require more attention in protein complex correlation profiling. The inclusion of additional separation steps could resolve 'opportunistic' co-elution.

The mixed quality of results observed for previously studied protein complexes has revealed an interesting aspect of protein profiling. Experiments performed so far show that the extent of co-elution and overlap of traces can be actually very high and Pearson coefficient can capture this similarity efficiently. However, for some known complexes, one or more of the subunits are clearly eluted in a different fraction. A typical example is the NADH dehydrogenase complex, which appears as a multi-peaking complex in elution profiles of human, plant and other samples [17,18]. Rather than

a problem, this multiplicity of peaks in different species suggests the existence of true complex isoforms, which in the case of NADH dehydrogenase are supported by the literature [24,25]. The discovery of all the isoforms of a complex can be easily detected with protein profiling. In contrast, affinity methods may only show an 'average' form, where mostly core components would be identified, and accessory components in low frequency could be mistakenly identified as contaminants. It may also be difficult to deduce the composition of such variable complexes from affinity-tagging experiments.

Currently, the proteomic techniques can give only preliminary information about protein complexes. Results can be used for the generation of new hypotheses, but additional information is mandatory in order to confirm the interaction. Comparison with other experimental information predicting protein interactions from TAP, yeast two-hybrid, gene co-expression and other experiments remains important. Additionally, enrichment of GO (gene ontology) and key terms and other similar ideas are now broadly used [19,26]. Despite being 'softer' information, the use of data and text-mining tools not only helps to validate the techniques, but provides a summarized biological description of the results. Central repositories of scientific information can play an important role in compiling such information supporting protein–protein interactions.

Conclusions

The successful proof of concept of protein complex profiling has been presented with a variety of technical implementations. The field of protein complex profiling is still in its infancy, but the potential increase of sample throughput is an enticing goal that could be achieved with these techniques. In the near future, the protein profiling could be greatly enhanced by an increase in the sensitivity of MS or lowered costs that could allow increased resolution in the analysis of profiles.

Funding

M.P.S. is grateful for a scholarship from CONICYT (Comisión Nacional de Investigación Científica y Tecnológica) BIRF (Banco Internacional de Fomento y Reconstrucción), Chile, and The Cambridge Trusts.

References

1 Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M. and Séraphin, B. (1999) A generic protein purification method for protein complex characterization and proteome exploration. Nat. Biotech. **17**, 1030–1032
2 Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M. and Séraphin, B. (2001) The tandem affinity purification (tap) method: a general procedure of protein complex purification. Methods **24**, 218–229
3 Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.-L., Millar, A., Taylor, P., Bennett, K., Boutilier, K. et al. (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature **415**, 180–183
4 Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P. et al. (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. Nature **440**, 637–643
5 Gavin, A.-C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.-M., Cruciat, C.-M. et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature **415**, 141–147
6 Wodak, S.J., Pu, S., Vlasblom, J. and Séraphin, B. (2009) Challenges and rewards of interaction proteomics. Mol. Cell. Proteomics **8**, 3–18
7 Benschop, J.J., Mohammed, S., O'Flaherty, M., Heck, A. J.R., Slijper, M. and Menke, F.L.H. (2007) Quantitative phosphoproteomics of early elicitor signaling in *Arabidopsis*. Mol. Cell. Proteomics **6**, 1198–1214
8 Lund, R., Leth-Larsen, R., Jensen, O.N. and Ditzel, H.J. (2009) Efficient isolation and quantitative proteomic analysis of cancer cell plasma membrane proteins for identification of metastasis-associated cell surface markers. J. Proteome Res. **8**, 3078–3090
9 Kantardjieff, A., Jacob, N.M., Yee, J.C., Epstein, E., Kok, Y.-J., Philp, R., Betenbaugh, M. and Hu, W.-S. (2010) Transcriptome and proteome analysis of Chinese hamster ovary cells under low temperature and butyrate treatment. J. Biotechnol. **145**, 143–159
10 Oppermann, F.S., Gnad, F., Olsen, J.V., Hornberger, R., Greff, Z., Kéri, G., Mann, M. and Daub, H. (2009) Large-scale proteomics analysis of the human kinome. Mol. Cell. Proteomics **8**, 1751–1764
11 Selbach, M. and Mann, M. (2006) Protein interaction screening by quantitative immunoprecipitation combined with knockdown (QUICK). Nat. Methods **3**, 981–983
12 Chen, R., Brentnall, T.A., Pan, S., Cooke, K., Moyes, K.W., Lane, Z., Crispin, D.A., Goodlett, D.R., Aebersold, R. and Bronner, M.P. (2007) Quantitative proteomics analysis reveals that proteins differentially expressed in chronic pancreatitis are also frequently involved in pancreatic cancer. Mol. Cell. Proteomics **6**, 1331–1342
13 Sun, Q., Hu, C., Hu, J., Li, S. and Zhu, Y. (2009) Quantitative proteomic analysis of CMS-related changes in Honglian CMS rice anther. Protein J. **28**, 341–348
14 Miles, G.P., Samuel, M.A., Ranish, J.A., Donohoe, S.M., Sperrazzo, G.M. and Ellis, B.E. (2009) Quantitative proteomics identifies oxidant-induced, AtMPK6-dependent changes in *Arabidopsis thaliana* protein profiles. Plant Signaling Behav. **4**, 497–505
15 Ross, P.L., Huang, Y.L.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S. et al. (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. Mol. Cell. Proteomics **3**, 1154–1169
16 Old, W.M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K.G., Mendoza, A., Sevinsky, J.R., Resing, K.A. and Ahn, N.G. (2005) Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol. Cell. Proteomics **4**, 1487–1502
17 Hartman, N.T., Sicilia, F., Lilley, K.S. and Dupree, P. (2007) Proteomic complex detection using sedimentation. Anal. Chem. **79**, 2078–2083
18 Dong, M., Yang, L.L., Williams, K., Fisher, S.J., Hall, S.C., Biggin, M.D., Jin, J. and Witkowska, H.E. (2008) A "tagless" strategy for identification of stable protein complexes genome-wide by multidimensional orthogonal chromatographic separation and iTRAQ reagent tracking. J. Proteome Res. **7**, 1836–1849
19 Ramani, A.K., Li, Z., Hart, G.T., Carlson, M.W., Boutz, D.R. and Marcotte, E.M. (2008) A map of human protein interactions derived from co-expression of human mRNAs and their orthologs. Mol. Syst. Biol. **4**, 180
20 Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J. and Mann, M. (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol. Cell. Proteomics **4**, 1265–1272
21 Reisinger, V. and Eichacker, L.A. (2006) Analysis of membrane protein complexes by blue native PAGE. Proteomics **6**, (Suppl. 2), 6–15
22 Reisinger, V. and Eichacker, L.A. (2008) Isolation of membrane protein complexes by blue native electrophoresis. Methods Mol. Biol. **424**, 423–431
23 Wessels, H.J.C.T., Vogel, R.O., van den Heuvel, L., Smeitink, J.A., Rodenburg, R.J., Nijtmans, L.G. and Farhoud, M.H. (2009) LC-MS/MS as an alternative for SDS-PAGE in blue native analysis of protein complexes. Proteomics **9**, 4221–4228

24 Heazlewood, J.L., Howell, K.A. and Millar, A.H. (2003) Mitochondrial complex I from *Arabidopsis* and rice: orthologs of mammalian and fungal components coupled with plant-specific subunits. Biochim. Biophys. Acta **1604**, 159–169
25 Meyer, E.H., Taylor, N.L. and Millar, A.H. (2008) Resolving and identifying protein components of plant mitochondrial respiratory complexes using three dimensions of gel electrophoresis. J. Proteome Res. **7**, 786–794
26 Musso, G.A., Zhang, Z. and Emili, A. (2007) Experimental and computational procedures for the assessment of protein complexes on a genome-wide scale. Chem. Rev. **107**, 3585–3600

Received 10 January 2010
doi:10.1042/BST0380923