Interaction proteomics: characterization of protein complexes using tandem affinity purification–mass spectrometry

Pamela Völkel, Perrine Le Faou and Pierre-Olivier Angrand[1]

Chromatinomics, Interdisciplinary Research Institute, Université de Lille 1 – Sciences et Technologies/CNRS USR 3078, Parc Scientifique de la Haute Borne, 50 Avenue Halley, F-59658 Villeneuve d'Ascq, France

Abstract

Most cellular processes are carried out by a multitude of proteins that assemble into multimeric complexes. Thus a precise understanding of the biological pathways that control cellular events relies on the identification and on the biochemical characterization of the proteins involved in such multimeric assemblies. Advances in MS have made possible the identification of multisubunit protein complexes isolated from cell lysates with high sensitivity and accuracy, whereas the TAP (tandem affinity purification) methodology efficiently isolates native protein complexes from cells for proteomics analysis. TAP is a generic method based on the sequential utilization of two affinity tags to purify protein assemblies. During the first purification step, the Protein A moiety of the TAP tag is bound to IgG beads, and protein components associated with the TAP-tagged protein are retrieved by TEV (tobacco etch virus) protease cleavage. This enzyme is a sequence-specific protease cleaving a seven-amino-acid recognition site located between the first and second tags. In the second affinity step, the protein complex is immobilized to calmodulin-coated beads via the CBP (calmodulin-binding peptide) of the TAP tag. The CBP–calmodulin interaction is calcium-dependent and calcium-chelating agents are used in the second elution step to release the final protein complex preparation used for protein identification by MS. The TAP–MS approach has proven to efficiently permit the characterization of protein complexes from bacteria, yeast and mammalian cells, as well as from multicellular organisms such as *Caenorhabditis elegans*, *Drosophila* and mice.

Introduction

More than the proteins themselves, multimeric protein complexes constitute the fundamental and functional units of the proteome [1,2]. Indeed, most cellular processes require protein–protein interactions or the assemblies of large protein complexes. Protein complexes (or molecular machines) are dynamic entities that assemble in an ordered time- and space-dependent fashion to store and relay various cellular signals or to contribute to the cellular architecture. Protein complexes are more than the sum of their components. For example, enzymes very often interact with regulatory subunits or with scaffolding proteins that are required for their activity, subcellular localization or substrate specificity. Different enzymatic activities could also be associated within the same protein complexes. To co-ordinate the various cellular functions efficiently and precisely, many protein complexes interact with one another. Moreover, the same proteins could be associated within different molecular machines that exert distinct cellular functions. Thus the study of protein complex components is essential to unravel the molecular nature of the cellular programmes. In addition, important insights can be revealed by the characterization of the molecular environment of proteins with known functions.

Since protein complexes correspond to a key level of the information encoded by individual genes, the description of the interactome is one of the challenges of the post-genomic era [3]. The composition of protein complexes is not easily inferred by studying pairwise interactions on a large scale. In contrast, the development of protein complex purification strategies combined with protein identification by MS/MS (tandem MS) has emerged as a powerful approach to decipher the interactome. Among the different protein complex purification strategies, protein tagging based on the TAP (tandem affinity purification) tag has proven particularly successful for molecular machine analyses [4,5].

The TAP–MS procedure

The TAP method is an epitope-tagging purification strategy based on a dual purification that represents a major improvement in sample purification for further protein identification by MS. In the original system, the TAP tag fused in-frame with the protein of interest is composed of two tags separated by a TEV (tobacco etch virus) cleavage site: the proximal tag is the CBP (calmodulin-binding peptide),

www.biochemsoctrans.org

Biochemical Society Transactions
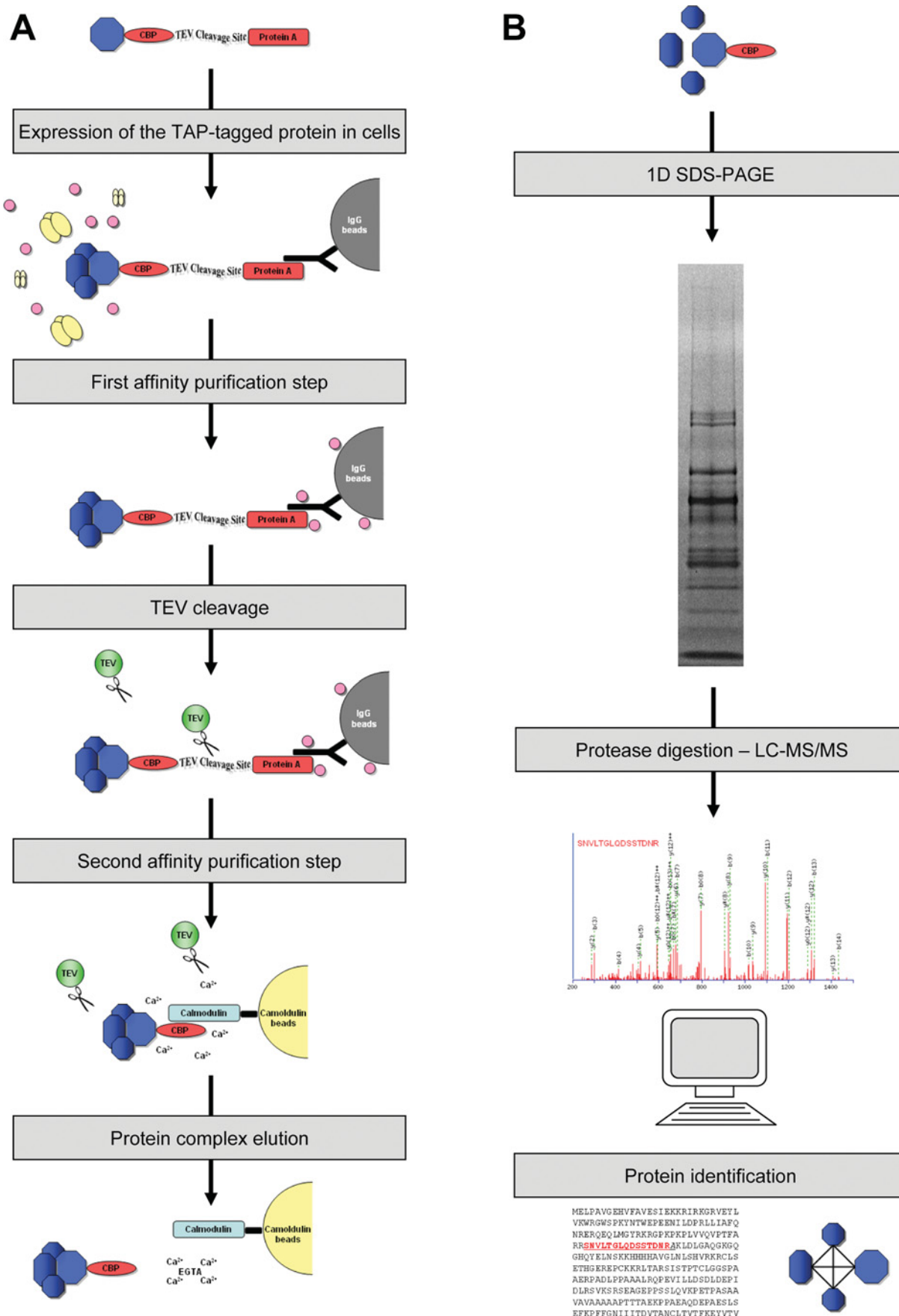
Figure 1 | For legend see facing page

Figure 1 | **Schematic representation of the TAP–MS methodology**

(**A**) Protein complex purification based on TAP. A TAP-tagged protein is expressed in cells and assembles with endogenous proteins. The TAP-tagged associated protein complex is sequentially purified in a two-step affinity procedure. In the first step, protein complexes immobilized on IgG-containing resin are specifically eluted by TEV protease cleavage and in the second step by lowering the calcium-dependent affinity of CBP to calmodulin using chelating agents. (**B**) Protein identification by MS. Proteins from the complexes are separated by one-dimensional SDS/PAGE, excised from the gel and digested with trypsin. The peptides generated are partially sequenced by MS/MS. Partial amino acid sequence and mass datasets allow the protein identification through database searches, and protein–protein interaction networks are deduced.

whereas the distal tag is composed of the IgG-binding part of the *Staphylococcus aureus* Protein A [4]. The TAP-tagged protein is expressed in cells to associate in protein complexes with endogenous components and is recovered together with its binding partners via a TAP procedure (Figure 1A). First, the TAP-tagged protein is immobilized on IgG–agarose beads via the Protein A moiety of the TAP tag, and gently washed. Protein complexes associated with the TAP-tagged protein are then specifically released from the resin by TEV protease cleavage. TEV is a sequence-specific protease cleaving a seven-amino-acid recognition site which is relatively rare in the mammalian proteome. Most of the proteins bound non-specifically to the IgG beads are left behind because the binding to resin and the release of the tagged protein from it are based on different chemistries (affinity compared with protease cleavage). In the second affinity step, the CBP tag is bound to calmodulin-coated beads in the presence of calcium. TEV protease and further contaminants are washed out. Finally, the recombinant protein and its interacting partners are then specifically retrieved by calcium chelation, since the CBP–calmodulin interaction is calcium-dependent.

Purified components are then separated by one-dimensional SDS/PAGE (Figure 1B) because the complexity of the protein samples after TAP is high enough to require a gel separation, but, in principle, two-dimensional gel electrophoresis is not necessary. Proteins are visualized by staining, excised from the gel and cleaved into peptides by trypsin or the endoproteinase Lys-C.

As far as purified protein complexes are considered, MS/MS combined with online nano-LC (liquid chromatography) peptide separation remains a reliable, robust and efficient approach [6,7]. Online nano-LC–MS is particularly adapted to samples containing an excess of contaminant proteins such as actin or tubulin, since it allows the identification of proteins that constitute as little as 2–5% of the total protein mixture. In MS/MS, the separated peptide ions enter the mass spectrometer and their mass-to-charge (m/z) ratios are measured in real time (in the effective range of the MS, approx. 200–2000 Da). Then, peptides are selected for fragmentation within the spectrometer. During this process, peptides are forced to collide with inert gas molecules leading to cleavage of peptide bonds and generating a set of fragments that differ in length by one amino acid each. The resulting peptide fragments are again measured to produce a series of signals (MS/MS spectra) which correspond in

mass to adjacent amino acid segments. Information on peptide sequence, peptide mass and fragment mass can be queried against theoretical spectra generated from protein sequences in appropriate protein, EST (expressed sequence tag) or genomic databases using peptide sequencing software (Mascot, SEQUEST, etc.). Thus, even if, in many cases, only part of the amino acid sequence can be read, the searches lead to the identification of the corresponding protein [8–11]. Each peptide analysed identifies a given protein deduced from the peptide sequence. Analysis of different peptides in the digest mixture can confirm the identification of a protein or identify a different protein that is present in the mixture.

The TAP–MS approach, originally described for protein complex analysis in yeast [4], has also been shown to permit efficient protein complex characterization from *Escherichia coli* [12], trypasonomatids [13] and mammalian cells [14,15], as well as from multicellular organisms such as plants [16,17], *Caenorhabditis elegans* [18], *Drosophila* [19] and mice [20].

Application fields of TAP–MS methodology

TAP–MS leads to the identification and characterization of protein complexes that are formed *in vivo*, and the two successive purification steps confer a high degree of specificity on the technology. Stringent washes are not required, and less stable multiprotein assemblies may be recovered together with a reduced number and a lower amount of contaminating proteins. More importantly, TAP–MS is a generic and reproducible method. Protein complex purification by TAP does not rely on the availability of specific antibodies and is potentially applicable to any protein, as far as its corresponding cDNA can be cloned. Moreover, the same optimized protocol is used for the purifications of different protein complexes, making the list of background contaminants across purifications easy to establish. However, in some cases, a purification contaminant protein could be a true component of a given protein complex, and more stringent statistical analyses of the MS data (number of peptides identified, etc.) are required to determine whether a protein identified is a contaminant or an interacting partner.

Because of its numerous strengths, TAP–MS appears to be a method of choice to map protein–protein interactions in large-scale experiments in *E. coli* [21] and in yeast [22,23]. Without approaching the scale of the bacterial and yeast data, the TAP–MS technology has been used efficiently in

mammalian cells to map smaller protein networks such as the Par interactome involved in cell polarity [14] or the interaction network around the human transcription machinery [24], as well as to decipher biological signalling pathways such as the TNFα (tumour necrosis factor α)/NF-κB (nuclear factor κB) [25] or Wnt/β-catenin [26] pathways. In the case of the TNFα/NF-κB signalling pathway, Bouwmeester et al. [25] applied the TAP–MS methodology to compare protein complex compositions under different physiological conditions (stimulated or non-stimulated by TNFα). Finally, the application of the TAP–MS approach to transgenic mice enables the comparison of protein complex organization between different tissues and allowed the characterization of novel interacting partners not previously identified in cells in culture [20].

Thus TAP–MS emerges as a powerful tool in protein network studies for systems biology in a number of different organisms and in different physiological conditions.

Further developments of the TAP–MS technology

Despite a number of strengths, the TAP–MS approach has several limitations associated with the epitope-tagging technologies in general. In particular, the fusion of the TAP tag (approx. 21 kDa) to a protein may affect its folding, activity or interaction with other cellular components. In a large-scale study in yeast, Gavin et al. [22] determined that, in 18% of the cases when essential genes were TAP-tagged, viable strains were not obtained, indicating that C-terminal TAP tagging can impair protein function. Tagging the proteins at the other end might often overcome the problem, but the development of smaller tags, based on new combinations of epitope tags is an alternative way to tackle this issue. In particular, a novel tag (SF-TAP) composed of the streptavidin II tag and the FLAG tag reduces the size of the TAP tag to 4.6 kDa [27]. In this case, protein complex recovery is not based on the addition of a protease and a chelating agent, but on the presence of metabolites. In the first purification step, the streptavidin–biotin binding is disrupted by desthiobiotin, and final protein complexes are eluted with FLAG octapetides in the second step. Moreover, B-Raf was significantly less expressed when tagged with the original TAP tag in comparison with when expressed in fusion with the SF-TAP tag. Even though this phenomenon was demonstrated with only one tagged protein (B-Raf), this example could indicate a possible instability of proteins when fused to large tags.

TAP is successfully used to purify protein complexes from mammalian cells in culture. However, yields from the tandem purifications remain low, and consequently large amounts of starting material are required. In order to increase purification yields, Bürckstümmer et al. [28] developed an alternative tag, named GS-TAP tag, composed of two IgG-binding units of the Protein G from *Streptococcus* sp., a TEV cleavage site, and SBP (streptavidin-binding peptide). The GS-TAP tag produces a 10-fold increase in efficiency when compared with the conventional TAP tag. The procedure allows purification

of protein complexes that were not amenable to TAP and use of less starting material, leading to higher success rates and possibly identifies less abundant protein assemblies. The technological improvements conferred by the GS-TAP tag are also confirmed by studies performed in *Drosophila* embryos [29].

The TAP method may fail to detect transient interactions, low stoichiometric complexes and/or those interactions occurring only in specific physiological states under-represented in exponentially growing cells. In addition, gentle extraction conditions associated with the procedure renders the purification of protein complexes strongly associated with cell structures such as chromatin difficult to recover [30]. The development of a HB-TAP (histidine-biotin) tag for two-step purification under fully denaturing conditions combined with *in vivo* cross-linking might overcome some of these current limitations of the technology [31].

In *E. coli* and yeast, homologous recombination allows fast and efficient ORF (open reading frame) tagging, and subsequent expression of the tagged protein at physiological levels under the control of their endogenous promoters [21,22]. In contrast, strategies based on exogenous expression of the TAP-tagged cDNA were mainly used for protein complex studies in mammalian systems. It is quite reasonable to expect that changes in expression of some proteins may alter the stoichiometry of the corresponding protein complexes, making the identification of low-abundant binding interactors more difficult. In particular, high expression of the TAP-tagged protein using transient expression often results in the recovery of large quantities of chaperones and heat-shock proteins presumably associated with the misfolded overexpressed TAP-tagged protein. Very often, rather than generating false interactions, overexpression of the TAP-tagged proteins leads to the masking of binding partners of low abundance. Thus stable expression using weak or inducible promoters inserted by virus-mediated gene transfer to reduce the number of copies by adjusting the multiplicity of infection, would be favoured in order to identify more easily the binding partners. Ultimately, the expression of the TAP-tagged protein at real endogenous levels can be achieved by the generation of TAP-knockin mouse lines [32]. An interesting alternative to TAP-knockin mice also comes from the usage of large genomic clones such as BACs (bacterial artificial chromosomes). Transgenes based on BACs are often large enough to contain complete genes with all their endogenous regulatory sequences. In addition, the development of efficient DNA engineering methods enables rapid and robust modifications of these large DNA sequences, including high-throughput TAP tagging for protein complex purifications from cells in culture and mice [33].

Conclusions

By dramatically reducing the levels of background proteins in the purifications, TAP is a generic method that allows an efficient and accurate purification of protein complexes for identification by MS. In recent years, the TAP–MS strategy

has been employed successfully to perform large-scale protein interaction mapping in lower organisms and to decipher smaller interactomes and signalling pathways in mammals, therefore making this technology a central tool in systems biology.

References

1 Alberts, B. (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologist. Cell **92**, 291–294

2 Gavin, A.C. and Superti-Furga, G. (2003) Protein complexes and proteome organization from yeast to man. Curr. Opin. Chem. Biol. **7**, 21–27

3 Vidal, M. (2001) A biological atlas of functional maps. Cell **104**, 333–339

4 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. Biotechnol. **17**, 1030–1032

5 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 for complex purification. Methods **24**, 218–229

6 McCormack, A.L., Schieltz, D.M., Goode, B., Yang, S., Barnes, G., Drubin, D. and Yates, 3rd, J.R. (1997) Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. Anal. Chem. **69**, 767–776

7 Bauer, A. and Kuster, B. (2003) Affinity purification–mass spectrometry: powerful tools for the characterization of protein complexes. Eur. J. Biochem. **270**, 570–578

8 Mann, M. and Wilm, M. (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. Anal. Chem. **66**, 4390–4399

9 Eng, J.K., McCormack, A.L. and Yates, 3rd, J.R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. **5**, 976–989

10 Kuster, B., Mortensen, P., Andersen, J.S. and Mann, M. (2001) Mass spectrometry allows direct identification of protein in large genomes. Proteomics **1**, 641–650

11 Nesvizhskii, A.I. and Aebersold, R. (2004) Analysis, statistical validation and dissemination of large-scale proteomics datasets generated by tandem MS. Drug Discov. Today **9**, 173–181

12 Guly, D., Moinier, D., Loiseau, L. and Bouveret, E. (2003) New partners of acyl carrier protein detected in Escherichia coli detected by tandem affinity purification. FEBS Lett. **548**, 90–96

13 Estevez, A.M., Kempf, T. and Clayton, C. (2001) The exosome of Trypanosoma brucei. EMBO J. **20**, 3831–3839

14 Brajenovic, M., Joberty, G., Kuster, B., Bouwmeester, T. and Drewes, G. (2004) Comprehensive proteomic analysis of human Par protein complexes reveals an interconnected protein network. J. Biol. Chem. **279**, 12804–12811

15 Westermarck, J., Weiss, C., Saffrich, R., Kast, J., Musti, A.M., Wessely, M., Ansorge, W., Séraphin, B., Wilm, M., Valdez, B.C and Bohmann, D. (2002) The DEXD/H-box RNA helicase RHII/Gu is a co-factor for c-Jun-activated transcription. EMBO J. **21**, 451–460

16 Rohila, J.S., Chen, M., Cerny, R. and Fromm, M.E. (2004) Improved tandem affinity purification tag and methods for isolation of protein heterocomplexes from plants. Plant J. **38**, 172–181

17 Van Leene, J., Stals, H., Eeckhout, D., Persiau, G., Van De Slijke, E., Van Isterdael, G., De Clercq, A., Bonnet, E., Laukens, K., Remmerie, N. et al. (2007) A tandem affinity purification-based technology platform to study the cell cycle interactome in Arabidopsis thaliana. Mol. Cell. Proteomics **6**, 1226–1238

18 Polanowska, J., Martin, J.S., Fisher, R., Scopa, T., Rae, I. and Boulton, S.J. (2004) Tandem immunoaffinity purification of protein complexes from Caenorhabditis elegans. BioTechniques **36**, 778–780

19 Veraksa, A., Bauer, A. and Artavanis-Tsakonas, S. (2005) Analyzing protein complexes in Drosophila with tandem affinity purification–mass spectrometry. Dev. Dyn. **232**, 827–834

20 Angrand, P.O., Segura, I., Völkel, P., Ghidelli, S., Terry, R., Brajenovic, M., Vintersten, K., Klein, R., Superti-Furga, G., Drewes, G. et al. (2006) Transgenic mouse proteomics identifies new 14-3-3 associated proteins involved in cytoskeletal rearrangements and cell signaling. Mol. Cell. Proteomics **5**, 2211–2227

21 Butland, G., Peregrín-Alvarez, J.M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N. et al. (2005) Interaction network containing conserved and essential protein complexes in Escherichia coli. Nature **433**, 531–537

22 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

23 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

24 Jeronimo, C., Forget, D., Bouchard, A., Li, Q., Chua, G., Poitras, C., Thérien, C., Bergeron, D., Bourassa, S., Greenblatt, J. et al. (2007) Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. Mol. Cell **27**, 262–274

25 Bouwmeester, T., Bauch, A., Ruffner, H., Angrand, P.O., Bergamini, G., Croughton, K., Cruciat, C., Eberhard, D., Gagneur, J., Ghidelli, S. et al. (2004) A physical and functional map of the human TNF-α/NF-κB signal transduction pathway. Nat. Cell Biol. **6**, 97–105

26 Major, M.B., Camp, N.D., Berndt, J.D., Yi, X., Goldenberg, S.J., Hubbert, C., Biechele, T.L., Gingras, A.C., Zheng, N., Maccoss, M.J. et al. (2007) Wilms tumor suppressor WTX negatively regulates WNT/β-catenin signaling. Science **316**, 1043–1046

27 Gloeckner, C.J., Boldt, K., Schumacher, A., Roepman, R. and Ueffing, M. (2007) A novel tandem affinity purification strategy for the efficient isolation and characterisation of native protein complexes. Proteomics **7**, 4228–4234

28 Bürckstümmer, T., Bennett, K.L., Preradovic, A., Schütze, G., Hantschel, O., Superti-Furga, G. and Bauch, A. (2006) An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. Nat. Methods **3**, 1013–1019

29 Kyriakakis, P., Tipping, M., Abed, L. and Veraksa, A. (2008) Tandem affinity purification in Drosophila: the advantages of the GS-TAP system. Fly **2**, 229–235

30 Souza, P.P., Völkel, P., Trinel, D., Vandamme, J., Rosnoblet, C., Héliot, L. and Angrand, P.O. (2009) The histone methyltransferase SUV420H2 and heterochromatin proteins HP1 interact but show different dynamic behaviours. BMC Cell Biol. **10**, 41

31 Tagwerker, C., Flick, K., Cui, M., Guerrero, C., Dou, Y., Auer, B., Baldi, P., Huang, L. and Kaiser, P. (2006) A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with in vivo cross-linking. Mol. Cell. Proteomics **5**, 737–748

32 Fernández, E., Collins, M.O., Uren, R.T., Kopanitsa, M.V., Komiyama, N.H., Croning, M.D.R., Zografos, L., Armstrong, J.D., Choudhary, J.S. and Grant, S.G.N. (2009) Targeted tandem affinity purification of PSD-95 recovers core postsynaptic complexes and schizophrenia susceptibility proteins. Mol. Syst. Biol. **5**, 269

33 Poser, I., Sarov, M., Hutchins, J.R., Hériché, J.K., Toyoda, Y., Pozniakovsky, A., Weigl, D., Nitzsche, A., Hegemann, B., Bird, A.W. et al. (2008) BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. Nat. Methods **5**, 409–415