REVIEW

Methods for the detection and analysis of protein–protein interactions

Tord Berggård[1], *Sara Linse*[1] *and Peter James*[2]

[1] Department of Biophysical Chemistry, Lund University, Lund, Sweden
[2] Department of Protein Technology, Lund University, Lund, Sweden

Received: February 6, 2007
Revised: April 12, 2007
Accepted: April 12, 2007

A large number of methods have been developed over the years to study protein–protein interactions. Many of these techniques are now available to the nonspecialist researcher thanks to new affordable instruments and/or resource centres. A typical protein–protein interaction study usually starts with an initial screen for novel binding partners. We start this review by describing three techniques that can be used for this purpose: (i) affinity-tagged proteins (ii) the two-hybrid system and (iii) some quantitative proteomic techniques that can be used in combination with, *e.g.*, affinity chromatography and coimmunoprecipitation for screening of protein–protein interactions. We then describe some public protein–protein interaction databases that can be searched to identify previously reported interactions for a given bait protein. Four strategies for validation of protein–protein interactions are presented: confocal microscopy for intracellular colocalization of proteins, coimmunoprecipitation, surface plasmon resonance (SPR) and spectroscopic studies. Throughout the review we focus particularly on the advantages and limitations of each method.

Keywords:
Affinity tags / Protein complexes / Protein–protein interactions / Protein purification

1 Introduction

The human genome consists of 20 000–30 000 genes coding for over 500 000 different proteins of which more than 10 000 proteins can be produced by the cell at any given time (the cellular 'proteome'). Is has been estimated that over 80% of proteins do not operate alone but in complexes. These protein–protein interactions are regulated by several mechanisms. For example, metal-binding or PTMs can lead to conformational changes that alter the affinity, co-operativity and kinetic parameters of the interaction. Many protein–protein interactions are part of larger cellular networks of protein–protein interactions. It is believed that the cellular network of protein–protein interactions are built up by highly connected protein nodes (so called hubs) and many poorly connected nodes. Each node receives inputs and generates one or more specific outputs in a manner similar to computational units. Examples of important protein complexes are the spliceosome, the ribosome and the nuclear pore complex. The basic architecture of the protein–protein interaction network is similar in all cells. Thus, hubs essential for cell survival are the same, but cell-specific differences can be found at the regulatory level. It has been shown previously that knock-out of a protein which has a central role in many networks tends to be lethal [1]. This phenomenon has been observed in many organisms [2–4] and is commonly referred to as the centrality-lethality rule. The current view that most, if not all, cellular proteins are directly or indirectly coupled in a large

Correspondence: Dr. Tord Berggård, Department of Biophysical Chemistry, Lund University, P. O. Box 124, SE-221 00 Lund, Sweden
E-mail: tord.berggard@bpc.lu.se
Fax: +46-46-2221495

Abbreviations: co-IP, coimmunoprecipitation; SPR, surface plasmon resonance; **TAP,** tandem affinity purification; **TEV,** tobacco etch virus; **Y2H,** yeast two-hybrid

cellular protein–protein interaction network has implications for the way we define cellular pathways. Thus, a given pathway (*e.g.*, vesicle trafficking, apoptosis or cell cycle control) can be regarded as a subsystem that is highly interconnected to other pathways.

We have during the last 10 years witnessed a tremendous development within the field of proteomics. Many innovative methods for the identification and characterization of protein–protein interactions have been presented and many of them are currently in use in laboratories around the world. The technology development has in many cases been fuelled by the extraordinary advances in MS, which has made the identification of proteins a relatively simple task. In some cases macromolecular complexes such as ribosomes and exosomes have been purified and analysed directly in the mass spectrometer [5–8]. Thus mass spectrometric techniques can be used not only to identify individual proteins, but also to characterize biological assemblies. A number of large-scale studies have been presented, using *e.g.*, two-hybrid screens [9–15] and coaffinity purification followed by MS [16–19] to detect protein–protein interactions on a genome-wide scale. Somewhat surprisingly, only a small number of the interactions are supported by more than one method [20]. Estimates of 40–80% false negatives and 30–60% false positives and have been assigned to high-throughput studies that have used two-hybrid techniques, affinity-based techniques or computational approaches [20–22]. The poor overlap can be explained partly by the fact that many different methods have been used. However, even within subsets of protein–protein interactions identified using the same method, the overlap can be poor (*e.g.*, compare the results from the yeast two-hybrid (Y2H) screens from Ito *et al.* [10] and Uetz *et al.* [9]). It has been estimated that due to a high false-positive rate, current yeast and human interaction maps are roughly only 50 and 10% complete, respectively [23]. Although a number of methods have been developed lately to combat false-positive discoveries [24] it is clearly important to experimentally validate protein–protein interactions by several methods. In this review, we describe some methods that are commonly used to identify and verify protein–protein interactions. We focus particularly on the advantages and disadvantages of each method.

2 Use of affinity tags for purification of protein complexes *in vivo*

In vivo affinity fusion-based protein purification takes advantage of the selective binding of a genetically fused affinity tag. First, cells are transfected with a plasmid coding for a bait protein fused to the tag. After an appropriate expression period, cells are lysed and the tagged bait, together with bound proteins, is isolated using a specific chemical or biological ligand linked to a solid support. Eluted proteins are then separated by gel-electrophoresis and specifically bound proteins (*i.e.*, proteins absent from the control) are identified by MS. Compared with other techniques, fusion-based affinity protein purification is an excellent method to purify and identify multiprotein complexes. Many other widely used techniques, such as the Y2H system, cannot detect interactions involving more than two proteins (see ref. [25]).

As the tagged protein is expressed *in vivo* it can undergo PTMs. This has important implications for the purification of protein complexes from mammalian cells because PTMs, such as phosphorylation, are often used by regulatory proteins to increase or decrease the affinity for their target proteins. Furthermore, two proteins that interact *in vitro* may not be expressed in the same cells (or cellular subcompartments) and therefore the interaction observed *in vitro* may be nonphysiological. By expressing the bait protein directly in cells it is allowed to be directed to its correct subcellular location and to associate with its physiological targets. However, during cell lysis, the tagged protein is present in a mixture of both physiological and nonphysiological targets. Therefore, nonphysiological targets may sometimes be present in the lysate and may be incorporated in the complex. This may be a problem, especially if the physiological complex is composed of proteins interacting with fast kinetics (high K_{on}, high K_{off}) and/or low affinity, and thus more easily replaced by nonphysiological ones.

Affinity-based methods are biased towards proteins that interact with high affinity and with slow kinetics of dissociation. Thus affinity chromatography-based procedures may not be optimal for the detection of transient protein interactions, especially if stringent rinsing procedures are used. For example, transient complexes involved in post-translational control of protein activity, may escape detection. What is the reason for the bias towards high affinity interacting proteins? First, one must consider that the intracellular milieu is very different from that in the test tube. Protein–protein interactions are not designed to occur in dilute buffers such as those that researchers use in the laboratory. Instead all protein–protein interactions inside a cell occur in a concentrated mixture of macromolecules. In fact the protein concentration in the cytosol may be as high as that of some protein crystals [26]. The high intracellular protein concentration influences the rate at which molecules diffuse in the cell and it leads to competition for water (referred to as macromolecular crowding). As a consequence, the binding affinity between proteins in a complex may be much higher in the crowded environment inside a cell compared to two proteins in a buffer. This may affect the retrieval of weakly interacting proteins in experiments relying on affinity-based purification of tagged proteins: It is possible that weakly interacting proteins are indeed bound to the tagged protein inside the cell, but dissociate after cell lysis and affinity purification of the tagged protein. Surprisingly, the fact that affinities between interacting proteins may be higher *in vivo* than *in vitro* (because of macromolecular crowding effects) has not been discussed very much in the literature on methods for purification and identification of protein–protein interactions.

High-throughput interaction data using tags are heavily biased towards proteins of high abundance [20]. If we consider a protein that is represented by just a few molecules in each cell, it may be very difficult to obtain enough material for identification by MS. The gel-band corresponding to the protein of interest may not be seen on the gel or it may be 'drowned' by other gel-bands corresponding to more abundant proteins. Methods such as two-hybrid systems [20] or proteome-chips are less biased towards proteins of high abundance because these methods are less influenced by the cellular expression level of a particular protein.

The tagged bait protein is often slightly overexpressed, as compared to the expression level of the endogenous counterpart. If the expression level of the bait is too high, the assembly of the protein complexes may not occur under physiological conditions [27, 28]. This may result in the isolation of large amounts of chaperones and heat shock proteins, which presumably interact with misfolded bait proteins [29]. Furthermore, overexpression can sometimes cause cytotoxicity.

As the bait protein is fused to an exogenous tag, this means that an artificial protein is introduced into the cell. Is some cases, the tag may cause problems and it is therefore recommended that both N- and C-terminally tagged proteins are used in each experiment. Examples of tag-specific problems are: (i) if the tag becomes buried inside the protein complex, the complex may not be retrieved on the affinity column. (ii) The tag may abolish or decrease the affinity between individual components of the complex. (iii) The tag may affect the intracellular localization of the bait protein. For example, some proteins are directed to the plasma membrane by myristoylation at the N-terminus [30]. If a tag is fused to the N-terminus of such proteins, they are likely to become incorrectly localized and as a result they may not bind to their physiological targets.

Stringent washes using buffers containing detergents or high salt concentrations are sometimes used to reduce contaminating proteins. However, stringent washes may also reduce the binding of low-affinity targets. In many cases it is an advantage to purify binding partners for several baits at the same time because it is easier to identify nonspecifically interacting proteins. Proteins that are identified repeatedly are likely to be contaminants or, alternatively, they may be promiscuous proteins (*e.g.*, chaperones) which are irrelevant in a given experiment. A common factor for contaminating proteins such as ribosomal proteins, metabolic enzymes and chaperones, is that they are often present in large amounts in the cell. Thus, a researcher may find it useful to generate databases of abundant proteins in the cell line(s) in which the tagged protein is being expressed. Schirle *et al.* [31] used LC-MS/MS to identify the core proteome of several cell lines: HEK293 (embryonic kidney cells), SKNBE2 (neuroblastoma), SW480 (colon carcinoma), HeLa (cervical adenocarcinoma), HeLaS3 (a clonal derivative of HeLa), and HepG2 (hepatocellular carcinoma) cells as well as a nuclear preparation from HEK293 cells. All proteins identified in the cell lines were then compiled into abundance lists based on the sum of all MASCOT scores [32] with which a particular protein was identified. Since abundant proteins are often contaminants, the abundance lists can be helpful when discriminating between specific *vs.* nonspecific binding partners in the cell lines mentioned above. Another powerful alternative for the identification of nonspecifically bound proteins is the use of quantitative proteomics which is described later in this review.

3 Tandem affinity purification (TAP)-tags

One of the best known methods to purify protein complexes is the TAP-method. The basic concept of TAP is the use of a so called TAP-tag, which is fused to a specific protein (Fig. 1). The TAP-tag consists of two sequential affinity tags spaced by a cleavage site of tobacco etch virus (TEV) protease [33]. The TEV protease cleaves a sequence (Glu-X-X-Tyr-X-Gln/Ser) which is very uncommon in mammalian proteins. The use of TEV protease therefore minimizes the risk of cleaving the bait protein and/or associated proteins.

An important advantage of the TAP-technique is that the amount of nonspecific binding is reduced as compared to other affinity-based techniques. This is achieved by the use of two purification steps. First, the TAP-tagged protein is expressed in a suitable cell line or organism and allowed to associate to its endogenous targets. After lysis of the cells, the TAP-tagged protein is allowed to bind *via* the first part of the TAP-tag (*e.g.*, protein A) to a specific column (*e.g.*, immobilized Igs). After rinsing, TEV-proteinase is added and the TAP-tag is cleaved, leaving the first affinity tag on the column. The bait protein, still fused to the second part of the TAP-tag (*e.g.*, a calmodulin-binding peptide) is then bound to a second column (*e.g.*, calmodulin-coated beads), which is rinsed and eluted (*e.g.*, by a buffer containing EDTA). The TAP-tag strategy has several advantages compared with other methods. For example, post-translationally modified and fully processed proteins are used as entry points and complex formation occurs *in vivo*. Although the TAP-tag method is highly sensitive and selective, a potential problem with the method is that the increased purity comes at a price; protein–protein interactions of a more transient nature may be lost during the series of purification steps [34]. Another problem is that a relatively large amount of starting material is required, which makes purification and identification of low-abundance binding partners a difficult task. Due to the problems with low yields, the TAP-method has been used only with limited success in mammalian cells. Recently, however, an efficient TAP-tag based on protein G and the streptavidin-binding peptide (GS-TAP) was developed for the purification of protein complexes from mammalian cells (Fig. 1) [35]. The GS-TAP-tag exhibits two important advantages in comparison to other TAP tags, such as the yeast TAP tag (yTAP) [33]: (i) relative to the yTAP the bait recovery is increased by a factor of ten and (ii) if desired, it is possible to

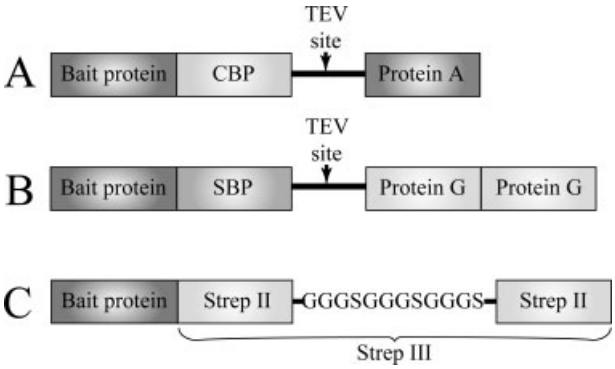

Figure 1. Schematic representations of some affinity tags used for protein–protein interaction studies. (A) Structure of the original TAP-tag. A bait protein is fused to tag consisting of a calmodulin-binding peptide (CBP), a TEV protease cleavage site and the IgG-binding moiety of *Staphylococucus aureus* protein A. (B) Structure of the GS-TAP-tag. A bait protein is fused to tag consisting of the streptavidin-binding peptide (SBP), a TEV protease cleavage site and protein G. (C) Structure of *Strep*-tag III. The *Strep*-tag III consists of two sequentially arranged *Strep*-tag II (Strep II) sequences (WSHPQFEK), separated by a linker region (GGGSGGGSGGGS).

perform single-step elution by streptavidin followed by elution by boiling or with biotin. These improvements make it possible to use the GS-TAP method for the characterization of protein complexes from mammalian cells that are not easily cultivated, such as immune cells and neuronal cells. In a normal experiment, the number of cells required for expression ranges from 5×10^7 to 1×10^9, and the purification procedure recovers around 5% of the bait present in the lysate assuming two-step elution. If single-step elution is chosen the recovery can be increased at the expense of specificity [35]. The recovery can be further increased if 'non-specific' elution by boiling is used instead of 'specific' elution using biotin. Thus the method appears to have the potential to experimentally achieve a balance between recovery and specificity for a given bait protein.

The TAP-method has also been used to identify protein–protein interactions from transgenic mice. In this study, mouse lines expressing TAP-tagged 14-3-3zeta protein were generated and protein interactions were determined [36]. This work points to the possibility of using transgenic techniques to study tissue-specific protein–protein interactions.

4 Strep-tag III

Sometimes the use of single tags is preferred to the use of TAP-tags because it may increase the yield of low-affinity binding targets. However, the amount of background is usually also increased. Examples of commonly used tags are the Flag-tag, the Myc-tag, the LAP-tag, or the HA-tag [37]. Junttila *et al.* [38] developed a tag that can be used as a 'single tag' alternative to TAP-tags for the purification of protein complexes from mammalian cells. This tag was called the *Strep*-tag III and has been commercialized by IBA (Göttingen, Germany) under the name One-STrEP-tag. It is a modified variant of the *Strep*-tag II which binds with high affinity to an engineered streptavidin derivative called *Strep*-Tactin (Fig. 1). The elution of the *Strep*-tag III fusion protein is achieved by the addition of biotin or a derivative of biotin called desthiobiotin, which binds to *Strep*-Tactin in competition with the *Strep*-tag III [39, 40]. The *Strep*-tag III is a short peptide tag and as such it is presumed to minimally interfere with complex formation. The tag has been adapted to a higher detergent compatibility, which is important for optimizing binding and washing conditions for protein–protein interactions. Junttila *et al.* [38] demonstrated the usefulness of the *Strep*-tag II by purifying both known and novel interacting proteins for PP2A from a cultured mammalian cancer cell line (human fibrosarcoma) [38].

5 Identification of protein–protein interactions by quantitative proteomics

One of the major challenges when working with protein–protein interactions is to distinguish specific from unspecific binding. This is particularly important when using MS as the final identification method because MS-based protein identification has become so sensitive that any protein–protein interaction screen will result in a large number of identified contaminant proteins. As discussed above, more specific and stringent purification such as TAP can decrease nonspecific binding. However, weakly interacting proteins may be lost [34] and it is sometimes preferred to use methods that preserve weak interactions (*e.g.*, single tag approaches, coimmunoprecipitation (co-IP) or classical affinity chromatography). Such methods, however, often produce significantly higher background and small amounts of specifically binding proteins may be masked by more abundant proteins that bind in a nonspecific manner. When using methods that increase recovery at the expense of specificity it is important to identify contaminating proteins. An excellent tool for the identification of false-positive interactions is quantitative proteomics. Quantitative proteomics is the common name for a set of techniques to differentially label proteins with stable isotopes and compare their relative abundance in different samples using MS. Two different types of labelling can be distinguished: 'metabolic' or 'chemical' labelling (Fig. 2). In metabolic labelling, stable isotopes are incorporated, *e.g.*, by introducing amino acids containing stable isotopes in the growth media of living cells. In chemical modification-based approaches, heavy and light chemical agents are bound to reactive amino acids of proteins 'postharvest'. In a typical experiment, two different samples are used. Proteins from the first sample are labelled with a light isotope mass tag, whereas proteins from the second sample are labelled with a heavy isotope tag. The two samples are then mixed, proteins

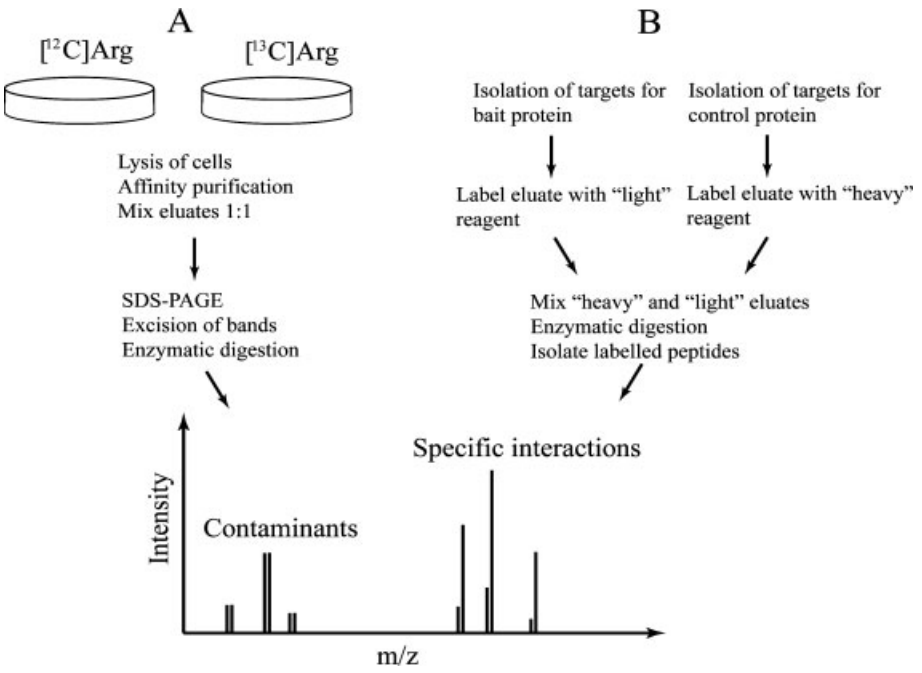

Figure 2. Quantitative proteomics for identifying nonspecifically bound proteins in protein–protein interaction experiments (A) Metabolic labelling: SILAC. Cells are grown in medium containing 'light' or 'heavy' amino acids which are metabolically incorporated into the newly synthesized proteins of cells. After cell-lysis, binding partners for a bait protein and a control protein are purified. Depending on the nature of the interaction between bait and target, lysates may be mixed before affinity purification of protein complexes (see ref. [86]) or the eluates may be mixed after affinity purification (see ref. [44]). Mixed eluates are then subjected to SDS-PAGE and excised bands are enzymatically digested and subjected to LC-MS/MS analysis. Peptides originating from nonspecifically bound proteins are in roughly the same abundance in both isotopic forms and they are identified based on the fact that these peptides display similar intensities in their 'heavy' and 'light' forms. On the other hand, peptides originating from specifically bound proteins display more intense peaks in one of the isotopic states ('heavy' state in the figure). After relative quantification in MS mode, peptide identification is obtained by MS/MS. (B) Chemical labelling: ICAT. Binding partners for a bait protein and a control protein are purified and the two samples are labelled by light and heavy ICAT reagents which are bound to Cys residues or other reactive groups. The two samples are mixed, enzymatically digested and ICAT labelled peptides are separated from unlabelled peptides. As with the SILAC approach, light and heavy versions of the same peptide can then be identified based on the mass difference and the relative abundances of the peptides can be determined by relative quantification in MS mode and peptide identification by MS/MS.

are enzymatically digested and analysed by MS. Peptides containing heavy and light peptides can be distinguished in a mass spectrum (parent ion scan, no peptide fragmentation) based on their mass difference. The method is considered 'quantitative' because the ion abundance ratio between the heavy- and the light-labelled peptide will reflect the actual abundance ratio of this peptide from two different samples. Peptides are then subjected to MS/MS for identification.

SILAC [41] is a metabolic labelling technique which can be used in protein–protein interaction screens, but also for the screening of peptide motif-based interactions (Fig. 2) (for reviews, see ref. [42, 43]). SILAC requires the use of isotope labelled amino acids, which is not necessarily expensive if labelled amino acids are bought in large quantities from a supplier [43]. Culturing cells using media supplemented with labelled amino acids is straightforward and not very different from normal cell culturing [42]. However, cell lines need to be adapted to dialyzed serum. As with chemical labelling techniques, access to a proteomics resource centre, as well as expertise in the handling of the mass spectrometric equipment is required. An open source programme, MSQuant, is available for interpreting the results [44].

Isotope-coded affinity tag (ICAT) is a chemical labelling technique that can be used to detect quantitative changes in the composition of complexes as well as quantitative changes in the abundance of protein complexes (Fig. 2) [29, 45]. An advantage of chemical labelling techniques is that one is not dependent on cell lines. As the labelling can be performed at any time, body fluids, biopsy material and tissue homogenates and can be used.

6 Chemical crosslinking

Another classical approach for determining protein–protein interaction has been by chemical crosslinking. One of the first approaches used to map large complexes such as the

ribosome, was the use of homobifunctional reagents, such as di-*N*-hydryoxysuccinimide esters to crosslink proteins *via* lysine groups or trauts reagent which converts lysine residues into SH-groups which are then crosslinked by oxidation. Further variations have included the development of photoactivatable reagents to allow crosslinking to be studied in a dynamic manner. These reagents are generally much less selective and allow crosslinking independent of the sequence of the binding domain. More complex reagents allow photocrosslinking and carry a ligand such as biotin to allow recovery of the crosslinked products. A comprehensive overview of these type of reagents is provided in the Pierce Applications Handbook with original references. New reagents are being synthesized all the time, the latest being isotopically labelled reagents. These are synthesized in labelled and unlabelled forms with zero and four deuteriums, respectively. This generates a specific isotope pattern which allows the crosslinked peptides to be rapidly identified even at low levels. Protein crosslinking is often used to 'freeze' and capture transient and/or low-affinity interactions (for a review, see ref. [46]). For example, TAP was recently combined with formaldehyde crosslinking to detect transient protein–protein interactions *in vivo* [47].

7 The two-hybrid system

The two-hybrid system is one of the most widely used methods to screen or confirm protein–protein interactions. It is based on the fact that many eukaryotic transcription factors, such as the yeast enhancer *Gal4*, are composed of two functionally distinct domains that mediate transcriptional activation and DNA binding respectively. The original method, which was developed in yeast, is described in the legend to Fig. 3. Many variants of the original Y2H system have been developed and the method has been modified and scaled up to focus on genome-wide interaction mapping (see ref. [9, 10, 15, 48]). The advantages of the Y2H system are that it is simple to set up, it is relatively inexpensive to use, it requires little optimization and it detects protein interactions *in vivo*. Transient and weak interactions, which are often important in signalling cascades, are more easily detected in two-hybrid systems since the genetic reporter gene strategy results in significant signal amplification [49]. There are also several disadvantages with the Y2H system. For example, only binary interactions can be studied and the method generates a large number of false positives. The exact rate of false-positive results is not known, but it has been estimated that as high as 50% of the identified interactions may be unreliable [50]. Therefore, once two interacting partners are identified by Y2H it is important to verify the interaction by other methods. Interactions between mammalian proteins that are dependent on PTMs and/or that are incorrectly folded in yeast cannot be studied by the original Y2H-system. To overcome this problem, mammalian variants of the Y2H system have been developed [51–53]. The original Y2H system requires that the bait-BD- and target-AD-fusion proteins are targeted to the yeast nucleus. Therefore interactions between proteins that contain strong(er) signals for targeting to other compartments may not be detected. Moreover, proteins that contain strongly hydrophobic domains, such as integral membrane proteins, may aggregate in the aqueous environment in the nucleus. To overcome this problem, cytoplasm-based Y2H variants [54, 55] and systems that allow the study of interactions between integral membrane proteins [56, 57] have been developed.

8 Public protein–protein interaction databases

After the initial screen for protein–protein interactions a list of potential targets has been generated. The next step is to search the literature to find out if there are any proteins in the list that has been detected by other investigators as binding partners to the bait protein. Usually, the researcher begins by searching PubMed (http://www.ncbi.nlm.nih.gov/entrez). However, many useful public databases for protein–protein interaction data also exist, including DIP (http://dip.doe-mbi.ucla.edu/), IntAct (http://www.ebi.ac.uk/intact), MINT (http://mint.bio.uniroma2.it/mint/Welcome.do) and MIPS (http://mips.gsf.de/genre/proj/mpact/). In Sections 8.1 and 8.2, we will mention a few databases that we have found particularly useful.

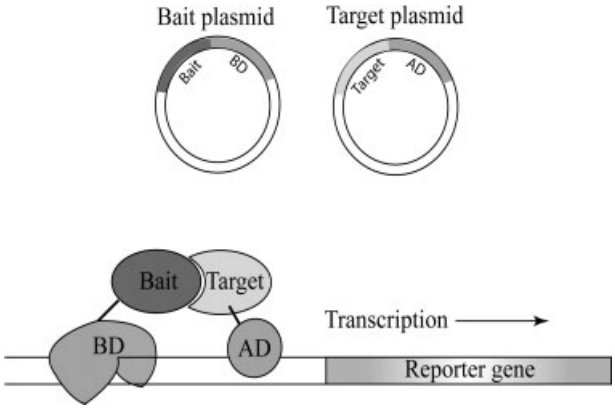

Figure 3. The yeast-two-hydrid system. Many eukaryotic transcriptional activators have two distinct functional domains, one that binds to a promoter DNA sequence (BD) and one that activates transcription (AD). In the two-hybrid assay, two plasmids are constructed. The first plasmid encodes a bait protein fused to a BD. The second plasmid encodes a target protein (or proteins encoded by an expression library) fused to an AD. The two fusion constructs are then coexpressed in the same cell. If the bait- and target proteins interact the two halves (BD and AD) will be brought together forming an intact and functional transcriptional activator. The intact transcriptional activator is then able to induce transcription of a downstream reporter gene.

8.1 iHOP

The iHOP (Information Hyperlinked over Proteins) database (http://www.ihop-net.org/UniPub/iHOP/) can be searched to identify previously reported interactions in PubMed for a protein of interest [58]. It is more efficient than the use of conventional keyword searches in PubMed.

8.2 IntAct

IntAct (http://www.ebi.ac.uk/intact.) is a freely available protein-interaction database which also contains analysis tools for protein interaction data [59]. All interactions are derived from literature mining or direct user submissions. At the time of the writing of this article, IntAct contained over 110 000 interactions. Interactions that derive from large-scale studies are usually not retrieved by keyword searches in PubMed, but are found in the IntAct-database. For example, we recently identified SNAP-25 as a binding partner for secretagogin, a poorly characterized EF-hand protein [60]. PubMed searches did not indicate the presence of any other identified targets. However, when searching IntAct, we found that secretagogin was identified in a large-scale two-hybrid screen as a binding partner to SNAP-23 [15]. SNAP-23 and SNAP-25 are related proteins that are essential for regulated exocytosis in different cell types [61]. By applying the rule of 'guilt by association' these results suggest that secretagogin may also be involved in regulated exocytosis.

9 Verification of interactions: confocal microscopy, coimmunoprecipitation, surface plasmon resonance and spectroscopic studies

A potential target may pass all control-experiments (*e.g.*, it is not identified from blank columns and it does not display promiscuous binding to other proteins). However, the interaction may still be indirect *via* a common interaction partner or it may simply be unphysiological. It is therefore important to verify the binding by additional methods. In some cases the original method can be used. For example, if a tagging approach is used, identified interactions can be systematically verified by tagging each of the interacting protein(s) and repeating the procedure. If the original bait protein is identified in the second round, the binding is less likely to be an artefact.

9.1 Confocal microscopy

Colocalization is defined as the presence of two or more different molecules residing at the same physical location in a cell. If proteins interact *in vivo* they are expected to be colocalized or, at least, they will display overlapping distribution within the cell. The intracellular localization of two (or more) proteins can be studied by confocal microscopy (For a review, see ref. [62]). Cells are first transfected with expression plasmids that encode the first protein fused to a specific tag and the second protein fused to another tag. Cells are then fixed and incubated with primary antibodies directed against the tags of the first and the second protein, respectively. Finally the specimen is incubated with secondary antibodies labelled by different fluorophores (*e.g.*, Cy2 and Cy3). By taking advantage of the fact that the two fluorophores display different emission maxima, the intracellular localization of the proteins can be monitored: If the two proteins are colocalized, the fluorescent probes will also be colocalized. This is represented in the image generated by the confocal microscope by pixels containing both colour contributions. If information on the organelle-specific localization of the protein complex is desired, a number of organelle-specific probes or of mAbs is available.

Overexpression and/or tagging may sometimes cause a protein to become incorrectly localized. Therefore, as an alternative to transient expression of tagged proteins, the localization of the endogenous proteins can be studied. This, however, requires the availability of highly specific primary antibodies raised against the proteins of interest. After fixing of the cells and incubation with primary antibodies, the cells are incubated with two fluorophore labelled secondary antibodies and detected in the same manner as described above. A potential problem with detection of endogenous proteins is that the components of the protein complex may not be expressed sufficiently in the cell lines studied.

9.2 Co-IP

One of the most commonly used methods for verification of protein–protein interactions is co-IP. In a typical experiment, bait complexes are captured from, *e.g.*, a cell lysate using a specific antibody. The antibody is then immobilized using protein A or protein G covalently attached to sepharose beads. After washing of the beads, the antibody, the bait and proteins associated to the bait are eluted (*e.g.*, by boiling). The bound proteins can then be identified by MS or by immunoblotting. Co-IP experiments usually generate significant background and it is therefore important to conduct parallel negative controls. Co-IP experiments can be carried out in several ways: (i) Co-IP from cell-lines or tissues expressing their endogenous proteins can be performed. The advantage of this approach is that endogenous protein complexes are studied. Therefore, any artificial effects of affinity tags or overexpression are avoided. The disadvantage is that highly specific antibodies are required. (ii) It is also possible to use cells transfected with a plasmid encoding a tagged bait protein (for a review, see ref. [63]). An antibody directed against the tag (instead of against the bait protein) can then be used in the co-IP experiments. An advantage of this approach is that one can be relatively confident that the antibody directed against the tag is specific and does not cross react with other proteins. Furthermore, epitope-tagged proteins can often be eluted by incubation with competing peptides, or other small

molecules, instead of boiling. Such specific elution often reduces the amount of contaminating proteins in the eluate. (iii) Alternatively, one can perform co-IP experiments using cells transfected with tagged versions of two putative interaction partners.

Several commercial co-IP kits for mammalian protein expression systems are available (*e.g.*, kits for HA- or c-Myc-tagged proteins from Pierce). Many commercial kits use spin-cup devices that increase washing efficiency, covalent binding of antibodies to the protein A/G support to minimize antibody interference and elution conditions that allow the immobilized antibody to be recycled several times.

9.3 Surface plasmon resonance (SPR) studies

SPR experiments offer several advantages for the verification and study of protein–protein interactions [64, 65]. Main advantages are that very small amounts of sample are needed (μg or sub-μg may suffice), and that no labelling is required. In addition, the method can provide information not only on affinities but also on the rates of association and dissociation, which can be equally important biologically, for example in regulatory networks. The method is generic for all kinds of proteins, because it relies on the phenomenon of SPR attributed to thin metal films (mainly gold or silver) and the signal recorded (the angle of minimum reflected light) depends on the refractive index close to the surface. The generality extends to all kinds of molecules, so not only protein–protein, but also protein–nucleic acid [66] or protein–ligand [65, 67] interactions may be verified and studied and recently the technique was applied to protein–nanoparticle interactions [68]. The response is proportional to the mass bound at the surface, *i.e.*, the number of molecules bound times their molecular weight. By this technique, one of the interaction partners is immobilized on a surface over a gold film and the other partner is injected over the surface. The amount bound is monitored continuously as a function of time, and after switching to buffer flow, the dissociation of the complex can be monitored (Fig. 4). The dissociation rate constant(s) is obtained from fitting a single or multiple (if more than one type of interaction occurs) exponential decay to the dissociation data. If the concentration of the injected protein is known, the association rate constant(s) can be obtained from fitting to the association phase data. The data analysis is described in more detail elsewhere [69, 70].

There are several kinds of surfaces commercially available for covalent coupling of proteins and the most common one is coated with a carboxylated dextrane matrix. Coupling of the protein to the surface can be *via* primary amino groups, for example, N-terminus and lysine side-chains. More specific coupling can be achieved *via*, for example, a single cysteine residue, natural or engineered. This eliminates the heterogeneity in coupled protein which often results upon amine coupling when multiple lysine residues are present, and the resulting data can be more precisely described by functions assuming a single type of binding site

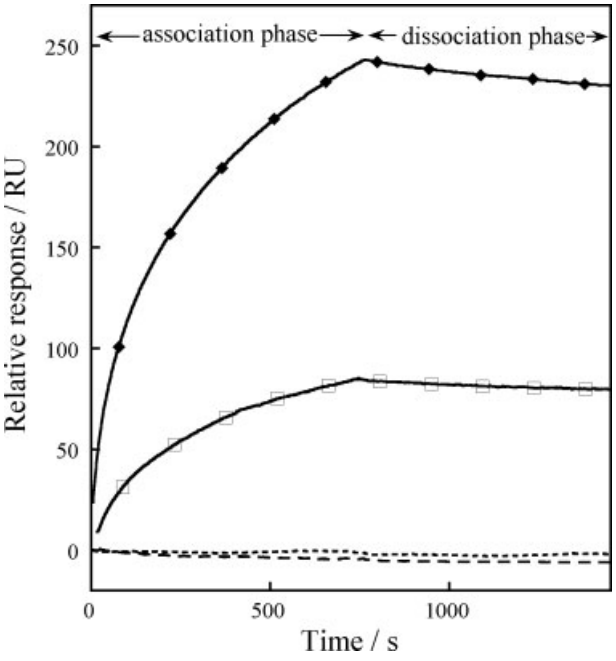

Figure 4. Surface plasmon resonance data. Synaptosomal associated protein of 25 kDa (SNAP-25) was immobilized *via* surface amines in a carboxylated dextran-matrix. Protein solution was injected over the surface between 0 and 750 s, after which buffer was injected at a constant flow rate. The increase in signal goes towards a plateau value during the association phase data for secretagogin in the presence (-◆-) and absence (-□-) of calcium. This indicates saturating binding and the relatively slow decay of the signal during the dissociation phase reflects a low dissociation rate constant for the complex between SNAP-25 and secretagogin. The data for the two control proteins, calbindin D28k (- -) and calmodulin (···) do not indicate any interaction with SNAP-25 with high enough affinity to be observed by this technique.

[71]. For multidomain proteins, where domain motion is a prerequisite for the binding of a partner, amine coupling may perturb or completely abolish binding, and in these cases coupling *via* a single cysteine or a single biotin may solve the problem. Biotinylation may be controlled to sub-stoichiometric levels so that few protein molecules will have more than one biotin moeity. An important application of immobilization for SPR studies involves the capture of membrane proteins in liposomes to study interactions with, for examples, G-protein coupled receptors [67, 72],

SPR has proven very useful in studies of antibody-antigen interactions [73] and for verifying protein–protein interactions, for example, between plasminogen and insulin-like growth factor II [74], between CD47 and macrophage membrane signal regulatory protein [75] and between secretagogin and SNAP-25 [60], and the method is also powerful in combination with protein engineering for localization of the binding sites [76]. The SPR technique has traditionally been used in low-throughput studies, but is now extended towards imaging and high-throughput studies [77–79], and towards the use of SPR to excite fluorophores close to the surface [80].

A potential limitation of the SPR method is that the immobilized protein may be inactivated. Also, the interaction is studied close to a surface, thus the measured parameters may not reflect those in solution. In addition, careful control experiments are needed to exclude binding to the surface *per se* in the absence of immobilized protein.

9.4 Spectroscopy studies

Spectroscopic techniques can be used to verify protein–protein interactions in solution. The requirement is that the complex exhibits a difference relative to the free components in any spectroscopy parameter, for example, the fluorescence intensity, wave-length maximum or polarization, fluorescence resonance energy transfer efficiency, circular dichroism, or NMR chemical shift or intensity. If a change occurs by any method, this provides a verification of an interaction. For measurement of the equilibrium dissociation constant, K_D, of the complex it is also required that the technique offers a significant S/N at a concentration that is not too far from the value of K_D. Typically, one of the components is held at constant concentration and the other one is titrated in steps and the binding parameters are estimated by fitting to the obtained data [81, 82]. One advantage with spectroscopic techniques is that the interacting partners can be labelled by, for example, NMR-active isotopes or by fluorescent probes and the same interaction can be studied both *in vitro* using purified components, in the background of other molecules in a solution that more closely reflects the *in vivo* situation, or inside cells [83]. Expression of one or two interacting partners as fusion proteins with different derivatives of green fluorescent protein is one means to allow verification of the interaction in living cells [84, 85].

10 Perspectives

Due to the fantastic development in the field of proteomics, a number of powerful methods have been developed recently for the detections of protein–protein interactions. Some of these methods did not exist at all a few years ago. Some are based on old principles, but with improved performance (*e.g.*, to use two tags instead of one in TAP). In some cases an 'old' method (*e.g.*, affinity chromatography) is used in combination with a 'new' method (*e.g.*, quantitative proteomics). Many of the newer methods require access not only to expensive equipment, but also to knowledge on how to conduct and interpret experiments. Furthermore, many of the new methods are generating a vast amount of data which has to be sorted and statistically worked at. Due to this, it is beyond the capabilities of many labs to use state–of-the-art methods for the detection of protein–protein interactions. However, we have in the last few years witnessed a global trend towards the establishments of resource centres, which makes the technology more accessible. In the future, more and better kits, *e.g.*, for quantitative proteomics, will probably arrive on the market. This will make it easier for an inexperienced researcher to prepare samples in his/her own laboratory before final evaluation together with a collaborator or a resource centre. With more powerful software and more user friendly methods, many of the methods which are presently only accessible to a few research groups, might become routine assays in the future.

We thank Dr. Maria Berggård Silow for helping with the figures and for critical comments on the manuscript.

11 References

[1] Jeong, H., Mason, S. P., Barabasi, A. L., Oltvai, Z. N., *Nature* 2001, *411*, 41–42.

[2] Yu, H., Greenbaum, D., Xin, Lu, H., Zhu, X., Gerstein, M., *Trends Genet* 2004, *20*, 227–231.

[3] Wuchty, S., *Proteomics* 2002, *2*, 1715–1723.

[4] Hahn, M. W., Kern, A. D., *Mol. Biol. Evol.* 2005, *22*, 803–806.

[5] van den Heuvel, R. H., Heck, A. J., *Curr. Opin. Chem. Biol.* 2004, *8*, 519–526.

[6] Benesch, J. L., Robinson, C. V., *Curr. Opin. Struct. Biol.* 2006, *16*, 245–251.

[7] Sharon, M., Robinson, C. V., *Annu. Rev. Biochem.* 2007, *76*, 167–193.

[8] Synowsky, S. A., van den Heuvel, R. H., Mohammed, S., Pijnappel, P. W., Heck, A. J., *Mol. Cell. Proteomics* 2006, *5*, 1581–1592.

[9] Uetz, P., Giot, L., Cagney, G., Mansfield, T. A. *et al.*, *Nature* 2000, *403*, 623–627.

[10] Ito, T., Chiba, T., Ozawa, R., Yoshida, M. *et al.*, *Proc. Natl. Acad. Sci. USA* 2001, *98*, 4569–4574.

[11] Giot, L., Bader, J. S., Brouwer, C., Chaudhuri, A. *et al.*, *Science* 2003, *302*, 1727–1736.

[12] Li, S., Armstrong, C. M., Bertin, N., Ge, H. *et al.*, *Science* 2004, *303*, 540–543.

[13] Formstecher, E., Aresta, S., Collura, V., Hamburger, A. *et al.*, *Genome Res.* 2005, *15*, 376–384.

[14] Stelzl, U., Worm, U., Lalowski, M., Haenig, C. *et al.*, *Cell* 2005, *122*, 957–968.

[15] Rual, J. F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T. *et al.*, *Nature* 2005, *437*, 1173–1178.

[16] Gavin, A. C., Bosche, M., Krause, R., Grandi, P. *et al.*, *Nature* 2002, *415*, 141–147.

[17] Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D. *et al.*, *Nature* 2002, *415*, 180–183.

[18] Ewing, R. M., Chu, P., Elisma, F., Li, H. *et al.*, *Mol. Syst. Biol.* 2007, *3*, 89.

[19] Krogan, N. J., Cagney, G., Yu, H., Zhong, G. *et al.*, *Nature* 2006, *440*, 637–643.

[20] von Mering, C., Krause, R., Snel, B., Cornell, M. *et al.*, *Nature* 2002, *417*, 399–403.

[21] Lu, L., Arakaki, A. K., Lu, H., Skolnick, J., *Genome Res.* 2003, *13*, 1146–1154.

[22] Aloy, P., Russell, R. B., *Nat. Biotechnol.* 2004, *22*, 1317–1321.

[23] Hart, G. T., Ramani, A. K., Marcotte, E. M., *Genome Biol.* 2006, *7*, 120.

[24] Suthram, S., Shlomi, T., Ruppin, E., Sharan, R., Ideker, T., *BMC Bioinformatics* 2006, *7*, 360.

[25] Fields, S., Song, O., *Nature* 1989, *340*, 245–246.

[26] Bray, D., Lay, S., *Proc. Natl. Acad. Sci. USA* 1997, *94*, 13493–13498.

[27] Drewes, G., Bouwmeester, T., *Curr. Opin. Cell Biol.* 2003, *15*, 199–205.

[28] Bouwmeester, T., Bauch, A., Ruffner, H., Angrand, P. O. *et al.*, *Nat. Cell Biol.* 2004, *6*, 97–105.

[29] Gingras, A. C., Aebersold, R., Raught, B., *J. Physiol.* 2005, *563*, 11–21.

[30] Farazi, T. A., Waksman, G., Gordon, J. I., *J. Biol. Chem.* 2001, *276*, 39501–39504.

[31] Schirle, M., Heurtier, M. A., Kuster, B., *Mol. Cell. Proteomics* 2003, *2*, 1297–1305.

[32] Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S., *Electrophoresis* 1999, *20*, 3551–3567.

[33] Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M. *et al.*, *Nat. Biotechnol.* 1999, *17*, 1030–1032.

[34] Bauch, A., Superti-Furga, G., *Immunol. Rev.* 2006, *210*, 187–207.

[35] Burckstummer, T., Bennett, K. L., Preradovic, A., Schutze, G. *et al.*, *Nat. Methods* 2006, *3*, 1013–1019.

[36] Angrand, P. O., Segura, I., Volkel, P., Ghidelli, S. *et al.*, *Mol. Cell. Proteomics* 2006, *5*, 2211–2227.

[37] Terpe, K., *Appl. Microbiol. Biotechnol.* 2003, *60*, 523–533.

[38] Junttila, M. R., Saarinen, S., Schmidt, T., Kast, J., Westermarck, J., *Proteomics* 2005, *5*, 1199–1203.

[39] Skerra, A., Schmidt, T. G., *Methods Enzymol.* 2000, *326*, 271–304.

[40] Schmidt, T. G., Koepke, J., Frank, R., Skerra, A., *J. Mol. Biol.* 1996, *255*, 753–766.

[41] Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B. *et al.*, *Mol. Cell. Proteomics* 2002, *1*, 376–386.

[42] Blagoev, B., Mann, M., *Methods* 2006, *40*, 243–250.

[43] Mann, M., *Nat. Rev. Mol. Cell. Biol.* 2006, *7*, 952–958.

[44] Schulze, W. X., Mann, M., *J. Biol. Chem.* 2004, *279*, 10756–10764.

[45] Ranish, J. A., Yi, E. C., Leslie, D. M., Purvine, S. O. *et al.*, *Nat. Genet.* 2003, *33*, 349–355.

[46] Melcher, K., *Curr. Protein Pept. Sci.* 2004, *5*, 287–296.

[47] Guerrero, C., Tagwerker, C., Kaiser, P., Huang, L., *Mol. Cell. Proteomics* 2006, *5*, 366–378.

[48] Walhout, A. J., Sordella, R., Lu, X., Hartley, J. L. *et al.*, *Science* 2000, *287*, 116–122.

[49] Estojak, J., Brent, R., Golemis, E. A., *Mol. Cell. Biol.* 1995, *15*, 5820–5829.

[50] Deane, C. M., Salwinski, L., Xenarios, I., Eisenberg, D., *Mol. Cell. Proteomics* 2002, *1*, 349–356.

[51] Lee, J. W., Lee, S. K., *Methods Mol. Biol.* 2004, *261*, 327–336.

[52] Dang, C. V., Barrett, J., Villa-Garcia, M., Resar, L. M. *et al.*, *Mol. Cell. Biol.* 1991, *11*, 954–962.

[53] Vasavada, H. A., Ganguly, S., Germino, F. J., Wang, Z. X., Weissman, S. M., *Proc. Natl. Acad. Sci. USA* 1991, *88*, 10686–10690.

[54] Aronheim, A., Zandi, E., Hennemann, H., Elledge, S. J., Karin, M., *Mol. Cell. Biol.* 1997, *17*, 3094–3102.

[55] Broder, Y. C., Katz, S., Aronheim, A., *Curr. Biol.* 1998, *8*, 1121–1124.

[56] Stagljar, I., Korostensky, C., Johnsson, N., te Heesen, S., *Proc. Natl. Acad. Sci. USA* 1998, *95*, 5187–5192.

[57] Fetchko, M., Stagljar, I., *Methods* 2004, *32*, 349–362.

[58] Hoffmann, R., Valencia, A., *Bioinformatics* 2005, *21*, ii252–ii258.

[59] Hermjakob, H., Montecchi-Palazzi, L., Lewington, C., Mudali, S. *et al.*, *Nucleic Acids Res.* 2004, *32*, D452–D455.

[60] Rogstam, A., Linse, S., Lindqvist, A., James, P. *et al.*, *Biochem. J.* 2007, *401*, 353–363.

[61] Sorensen, J. B., Nagy, G., Varoqueaux, F., Nehring, R. B. *et al.*, *Cell* 2003, *114*, 75–86.

[62] Miyashita, T., *Methods Mol. Biol.* 2004, *261*, 399–410.

[63] Masters, S. C., *Methods Mol. Biol.* 2004, *261*, 337–350.

[64] Szabo, A., Stolz, L., Granzow, R., *Curr. Opin. Struct. Biol.* 1995, *5*, 699–705.

[65] Huber, W., Mueller, F., *Curr. Pharm. Des.* 2006, *12*, 3999–4021.

[66] Katsamba, P. S., Park, S., Laird-Offringa, I. A., *Methods* 2002, *26*, 95–104.

[67] Navratilova, I., Dioszegi, M., Myszka, D. G., *Anal. Biochem.* 2006, *355*, 132–139.

[68] Cedervall, T., Lynch, I., Lindman, S., Berggård, T. *et al.*, *Proc. Natl. Acad. Sci. USA* 2007, *104*, 2050–2055.

[69] O'Shannessy, D. J., Brigham-Burke, M., Soneson, K. K., Hensley, P., Brooks, I., *Anal. Biochem.* 1993, *212*, 457–468.

[70] O'Shannessy, D. J., Brigham-Burke, M., Soneson, K. K., Hensley, P., Brooks, I., *Methods Enzymol.* 1994, *240*, 323–349.

[71] Dell'Orco, D., Xue, W. F., Thulin, E., Linse, S., *Biophys. J.* 2005, *88*, 1991–2002.

[72] Stenlund, P., Babcock, G. J., Sodroski, J., Myszka, D. G., *Anal. Biochem.* 2003, *316*, 243–250.

[73] Malmborg, A. C., Borrebaeck, C. A., *J. Immunol. Methods* 1995, *183*, 7–13.

[74] Oesterreicher, S., Blum, W. F., Schmidt, B., Braulke, T., Kubler, B., *J. Biol. Chem.* 2005, *280*, 9994–10000.

[75] Vernon-Wilson, E. F., Kee, W. J., Willis, A. C., Barclay, A. N. *et al.*, *Eur. J. Immunol.* 2000, *30*, 2130–2137.

[76] He, X., Shen, L., Malmborg, A. C., Smith, K. J. *et al.*, *Biochemistry* 1997, *36*, 3745–3754.

[77] Boozer, C., Kim, G., Cong, S., Guan, H., Londergan, T., *Curr. Opin. Biotechnol.* 2006, *17*, 400–405.

[78] Yuk, J. S., Kim, H. S., Jung, J. W., Jung, S. H. *et al.*, *Biosens. Bioelectron.* 2006, *21*, 1521–1528.

[79] Rich, R. L., Myszka, D. G., *Anal. Biochem.* 2007, *361*, 1–6.

[80] Ekgasit, S., Thammacharoen, C., Yu, F., Knoll, W., *Anal. Chem.* 2004, *76*, 2210–2219.

[81] Persson, E., Selander, M., Linse, S., Drakenberg, T. *et al.*, *J. Biol. Chem.* 1989, *264*, 16897–16904.

[82] Andre, I., Linse, S., *Anal. Biochem.* 2002, *305*, 195–205.

[83] Hink, M. A., Bisselin, T., Visser, A. J., *Plant Mol. Biol.* 2002, *50*, 871–883.

[84] Zaccolo, M., *Circ. Res.* 2004, *94*, 866–873.

[85] Pfleger, K. D., Eidne, K. A., *Nat. Methods* 2006, *3*, 165–174.

[86] Blagoev, B., Kratchmarova, I., Ong, S. E., Nielsen, M. *et al.*, *Nat. Biotechnol.* 2003, *21*, 315–318.

