Critical Reviews in Biochemistry and Molecular Biology, 41:315–327, 2006

ISSN: 1040-9238 print / 1549-7798 online
DOI: 10.1080/10409230600872872

informa
healthcare

Proteome-Scale Analysis of Biochemical Activity

Eric M. Phizicky and Elizabeth J. Grayhack
Department of Biochemistry and Biophysics, University of Rochester School of Medicine, Rochester, NY, USA

ABSTRACT During the last 10 years, there has been a large increase in the number of genome sequences available for study, altering the way that the biology of organisms is studied. In particular, scientific attention has increasingly focused on the proteome, and specifically on the role of all the proteins encoded by the genome. We focus here on several aspects of this problem. We describe several technologies in widespread use to clone genes on a genome-wide scale, and to express and purify the proteins encoded by these genes. We also describe a number of methods that have been developed to analyze various biochemical properties of the proteins, with attention to the methodology and the limitations of the approaches, followed by a look at possible developments in the next decade.

KEYWORDS functional genomics, biochemical genomics, protein microarrays, high throughput cloning, enzyme genomics

INTRODUCTION

It has been a little over 10 years since the first genome sequence was published; since then, genome sequencing has accelerated to such an extent that as of this writing (February, 2006) there are 331 complete genomes listed at the NCBI (http://www.ncbi.nlm.nih.gov/). The availability of these sequences has changed the way that biological science is investigated, as more and more attention has been focused on a holistic view of the organism, including expression of its genes; interactions among its genes; the identity, number, and variety of its proteins; the location of the proteins within the cell; and biochemical analysis of the proteins.

In this review, we describe recent advances in the biochemical analysis of the proteome. Accompanying the acquisition of genome sequences during the past 10 years there has been a revolution in methodology enabling rapid biochemical analysis of large numbers of proteins. Almost every important step in this process has been accelerated. Genes are rapidly cloned using modern methods of ligation-less cloning. Proteins are readily expressed from the cloned genes and are easily and efficiently purified using appropriate affinity reagents. Any of several formats are readily used to query proteins in parallel for particular biochemical properties. These experiments have already resulted in the assignment of a large number of biochemical activities and properties to proteins. Furthermore, new approaches are being developed to

Address correspondence to Eric Phizicky, Department of Biochemistry and Biophysics, University of Rochester School of Medicine, 601 Elmwood Avenue, Rochester, NY 14642. E-mail: eric_phizicky@urmc.rochester.edu

extend these methods to address increasingly complex biochemical questions.

CONSTRUCTION OF GENOMIC ORF COLLECTIONS

The foundation for biochemical analysis of the proteome is the availability of numerous genomic scale collections of strains expressing affinity tagged proteins, all of which depend on collections of cloned genes or on genetically manipulated strains that are constructed by high-efficiency methods (Hunt, 2005; Marsischky and LaBaer, 2004; Phizicky *et al.*, 2003). High-throughput cloning is readily achieved by methods that do not require ligation, including ligation-independent cloning (Aslanidis and de Jong, 1990; Aslanidis *et al.*, 1994), *in vitro* recombination of *att* sites mediated by Int and Xis from phage λ (Gateway^{TM}) (Hartley *et al.*, 2000; Walhout *et al.*, 2000) or of *lox* sites mediated by Cre recombinase of phage P1 (Creator^{TM}) (Liu, *et al.*, 1998), or *in vivo* recombination by gap-repair mediated transformation in yeast (Muhlrad *et al.*, 1992). Cloning has become so rapid using these types of methodologies that a single full-time researcher can easily clone a plate of 96 target ORFs in just a few days, and a genomic number of ORFs within a matter of months. Furthermore, use of high fidelity polymerase ensures error rates as low as 0.8×10^{-4} (Gelperin *et al.*, 2005). These efforts have resulted in a number of large collections of plasmids and strains that can be used for rapid parallel biochemical analysis of expressed and purified proteins.

Four genomic collections have been made that express yeast ORFs fused to purification tags. Two early genomic collections of yeast strains were designed for large scale preparation of ORFs tagged at their N-terminus with GST (GST-ORFs), after expression of the GST-ORFs under P_{CUP1} (Martzen *et al.*, 1999) or P_{GAL} control (Zhu *et al.*, 2001). Each of these was constructed by co-transformation of gapped plasmids and polymerase chain reaction (PCR) products directly into yeast. Two other genomic yeast collections that have been designed for purification and biochemical analysis express C-terminal ORF-fusion proteins. In one of these collections, the ORFs are expressed from plasmids under P_{GAL} control (Gelperin *et al.*, 2005). This movable ORF (MORF) collection was constructed by Gateway-mediated recombination into expression plasmids, followed by extensive sequence-verification (~1100 base pairs of perfect sequence), and then transformation of verified clones into yeast, isolation of single colonies, and examination of strains for expression of the fusion protein. The other yeast C-terminal ORF-fusion collection expresses ORFs from their own natural promoter, in their normal chromosomal location, which are tagged with the tandem affinity purification tag (TAP) (Ghaemmaghami *et al.*, 2003). Since this collection is produced by homologous recombination of the TAP-tag at the precise C-terminus of the natural yeast protein, the ORF sequence is unperturbed. This collection is extensively analyzed and quantified for expression of proteins.

A number of other large collections of clones/strains exists for other organisms. Several collections have been constructed with Gateway technology, including a collection of *Pseudomonas aeruginosa* ORF clones (LaBaer *et al.*, 2004), the *C. elegans* ORFeome collection (Reboul *et al.*, 2003; Wei *et al.*, 2005), the human ORFeome collection (Rual *et al.*, 2004), and the *Brucella melitensis* ORFeome (Dricot *et al.*, 2004). In addition, there are numerous large scale collections of ORFs cloned by structural biology consortia (estimated total ~50,000 ORF expression clones as of February, 2006), mostly cloned by Gateway or ligation-independent cloning methods (Dieckman *et al.*, 2002)(see http://www.mcsg.anl.gov/).

EXTENSIVE USE OF AFFINITY PURIFICATION METHODS

The presence of these collections of cloned ORFs allows the facile purification of large numbers of proteins on a parallel scale, because of the affinity tags to which the ORFs are fused (Braun and LaBaer, 2003). The purification of the proteins is highly effective because the affinity tags are efficient, specific, and well suited to batch treatment, using resins that are commercially available, including glutathione agarose for GST, IgG Sepharose for the ZZ domain of protein A, immobilized metal ions for His_6 tags, neutravidin for Strep2 tags, and Flag antibody for the flag tag. Use of a single affinity tag often yields protein of extraordinarily high quality. For example, the flag tag has been used for mass spectrometry analysis of complexes from over-expressed yeast proteins (Ho *et al.*, 2002). Although purification of His_6-tagged proteins by immobilized metal ion affinity chromatography (IMAC) is not quite as efficient as that achieved with some other tags, for

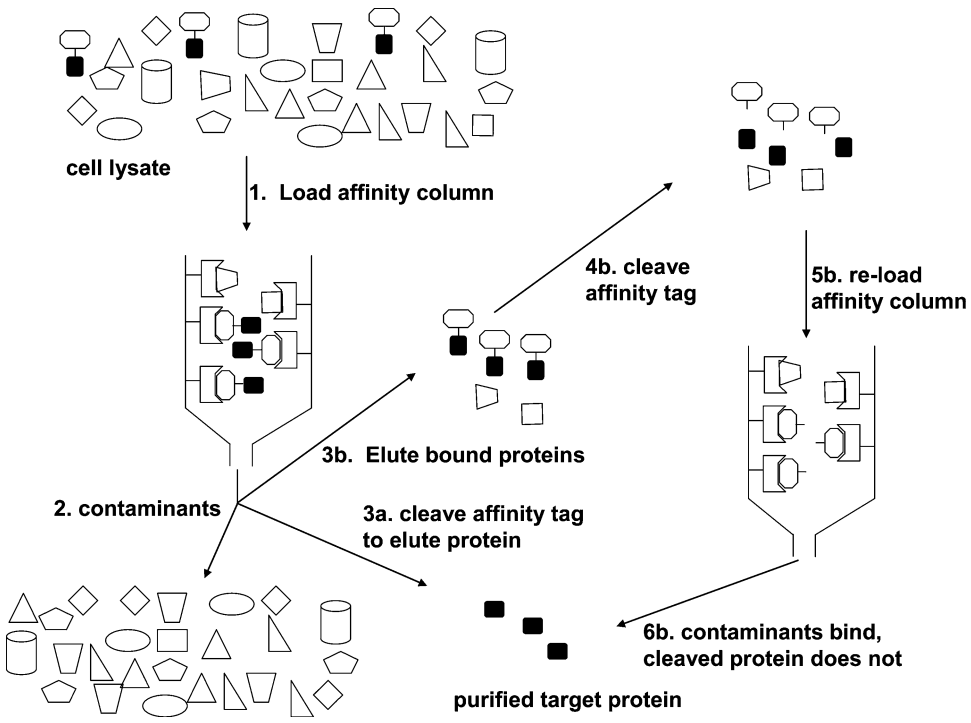

Figure 1 **Purification of proteins with a cleavable affinity tag. Cells expressing an affinity tagged target protein (with a *filled rectangle*) are lysed, and then applied to an affinity column (*step 1*). After contaminants are washed away (*step 2*), the target protein is purified either by proteolytic cleavage to release the target protein (*step 3a*), or by standard elution with salt or a small molecule competitor (*step 3b*), followed by cleavage of the tag (*4b*), and re-application of the protein to the same affinity column (*5b–6b*).**

moderately expressed proteins it is still often sufficient to yield preparations in which the tagged polypeptide is predominant, both in *E. coli* and in yeast (Gelperin *et al.*, 2005). Indeed, use of a subsequent gel filtration column to remove minor contaminants often yields His_6-tagged protein suitable for structural analysis, for proteins highly expressed in *E. coli*.

Use of a cleavable affinity tag enhances purification even further because it effectively results in two purification steps. Binding of the tagged protein to the affinity resin results in selective binding, and proteolytic cleavage of the tag uses a different selective step to elute bound protein, based on recognition of the cleavage site by the protease (Figure 1). This combination of steps can yield proteins of very high purity. We and others have used a related approach for purification of His-tagged proteins for structural analysis (Kim *et al.*, 2004; Sanishvili *et al.*, 2004). This method involves binding to IMAC columns such as Talon, followed by imidazole elution of the tagged protein, protease cleavage to remove the tag from the protein, and re-chromatography on the same IMAC resin, because the same contaminants bind the IMAC resin the second time, but the target protein does not because it no longer has the His_6 tag (Figure 1). Since the protease itself has a His_6-tag, it is also removed during this step. This is generally followed by a gel filtration column to remove remaining trace contaminants. Using this type of procedure, two full-time researchers can easily purify 8 proteins per week, each yielding 10 mg or more of purified protein. This type of double-purification of cleavable tags can be used in principle with any tag and any sufficiently specific protease.

One of the most sophisticated approaches for fusion protein purification uses two different affinity tags, together with a highly specific protease cleavage site. The original of this type of tag is the TAP-tag, which is composed of the ZZ domain of protein A, a tobacco etch virus (TEV) protease site, and a calmodulin binding peptide (CBP) (Rigaut *et al.*, 1999). Because each of the tags and the protease are so specific, use of this tag effectively amounts to a triple-purification, and therefore yields nearly pure preparations, even for proteins expressed at low or very low levels. Indeed, with tagged protein expressed at physiological levels, the resulting purified protein is often suitable for mass spectrometry analysis of co-purifying proteins (Gavin *et al.*, 2002). This tag, and variants of it (Ma *et al.*, 2005; Rubio *et al.*, 2005), have been used for purification of numerous proteins from numerous organisms.

A number of different proteases are in common use. The TEV protease used in the TAP-tag is highly specific and in wide use, although it is somewhat sensitive to detergents (Mohanty *et al.*, 2003). Rhinovirus 3C protease is also highly specific, is highly active at low temperatures, and retains activity in the presence of some detergents. Other proteases used to remove tags include thrombin, Factor Xa, and enterokinase, all of which are somewhat less specific but still widely used (Jenny *et al.*, 2003). One particularly powerful approach employs N-terminal SUMO-ORF fusion proteins. The small SUMO polypeptide not only acts as a chaperone to promote solubility, but can be completely removed by the Ulp1 protease, leaving native protein initiating at its natural codon (Chance *et al.*, 2002).

One major limitation in genomic scale purification of proteins is the variable levels of expression and solubility observed for foreign proteins expressed in bacteria. Although high levels of protein expression and solubility have been observed for some heterologous systems (Christendat *et al.*, 2000), substantial levels of soluble protein are often only observed in a small fraction of cloned ORFs (Service, 2002), including 15% of *C. elegans* ORFs (Luan *et al.*, 2004), and *Leishmania* ORFs (Figure 2). Several solutions to the solubility problem have been attempted, including expression of proteins as fusions of the maltose binding protein or NusA protein (Kapust and Waugh, 1999; Nallamsetty and Waugh, 2006), induction of expression under low temperature conditions to improve solubility, the use of *E. coli* strains that express ORFs at different induction levels and under different oxidation conditions, and the use of *E. coli* strains expressing rare tRNA genes to improve expression of proteins with poor codon usage (Baca and Hol, 2000). However, no general solution has emerged for this problem. One possible approach is cell-free synthesis, since this method has been used for expression of soluble proteins at high levels from wheat germ (Madin *et al.*, 2000; Tyler *et al.*, 2005; Vinarov *et al.*, 2004) and from *E. coli* (Kigawa *et al.*, 1999; Kim *et al.*, 1996; Yokoyama, 2003).

Expression of proteins in homologous systems is clearly a better option, based on the observation that between 87% and 93% of yeast fusion proteins are functional when expressed in yeast (Ghaemmaghami *et al.*, 2003; Huh *et al.*, 2003). However, even using cloned yeast ORFs in yeast there are limitations due to variations in expression levels (Gelperin *et al.*, 2005).

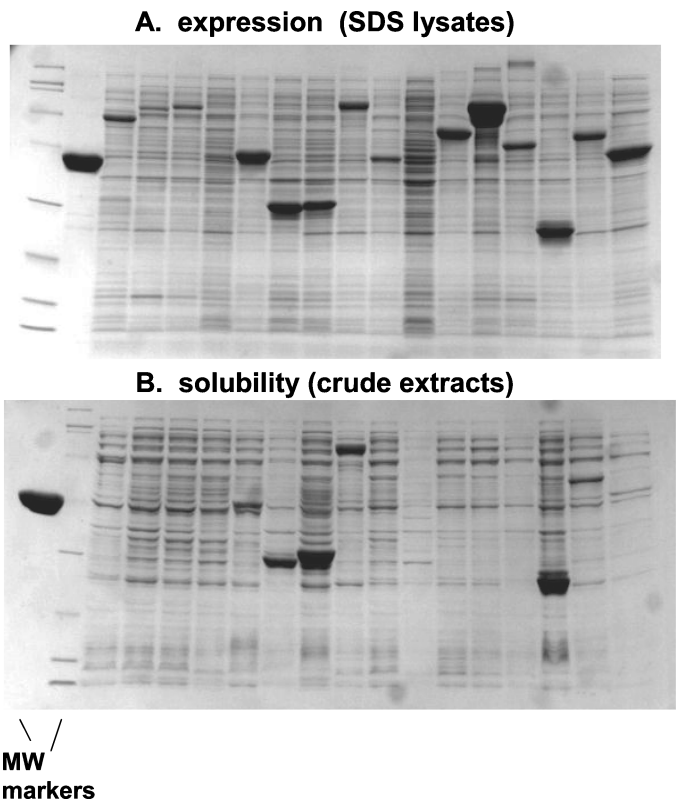

Figure 2 **Expression of foreign proteins in *E. coli*. 16 ORF targets from *Leishmania major* were cloned into an expression plasmid, and then analyzed for expression and solubility of the target proteins in *E. coli*. Although expression is observed in SDS lysates for 14 of the 16 proteins, expression levels vary. Furthermore, only 6 of these proteins are soluble, as measured by presence of the protein in the supernatant after sonication of cells and centrifugation to remove debris, and only two of the clones produce high levels of soluble protein.**

ANALYSIS OF BIOCHEMICAL ACTIVITY

Concomitant with the increased accessibility of genomic collections of clones and purified proteins, a number of sophisticated methods have been developed to biochemically analyze proteins in parallel. In general, these methods differ from each other in both the types of assays that are used to detect activity and the environment of the proteins during the assay. Assays for enzymatic activity have been carried out in pools of purified proteins with specific assays to identify the protein responsible for a particular molecular function, and with individual proteins using generic substrates to detect broad classes of activity, whereas assays for binding frequently employ microarrays and fluorescent ligands. In any of these formats, the relatively high concentrations of the purified proteins facilitates detection of activity, even under non-optimal conditions.

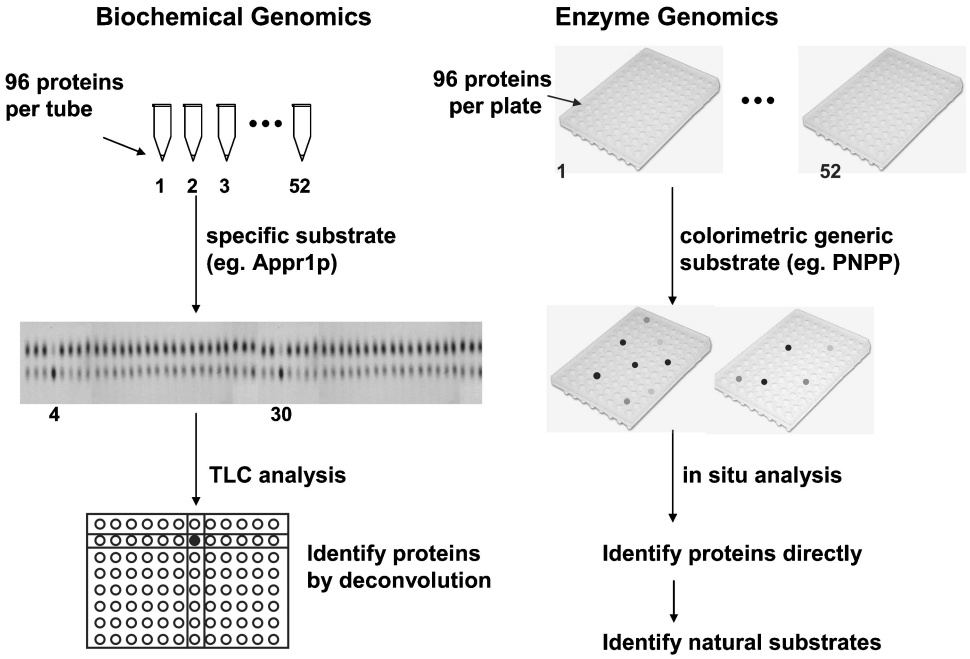

Figure 3 Comparison of biochemical genomics and enzyme genomics approaches for assignment of proteins to biochemical activities. In the biochemical genomics approach, pools of proteins are screened with specific substrates (such as ADP-ribose 1″-2″ cyclic phosphate, Appr1p) to identify a pool with a particular activity, and then active pools are deconvoluted to identify the protein. Because relatively few samples are processed in screening pools, assays can involve lower throughput methods such as gel electrophoresis or thin layer chromatography. In the enzyme genomics approach a generic substrate (such as paranitrophenylphosphate, PNPP) is used to screen a large array of proteins in situ with a color assay, resulting in direct identification of the proteins. Subsequently, a more specific natural substrate is identified by analysis of a battery of compounds.

Analysis of Enzymatic Activity

Enzymatic activity is remarkably easy to detect by parallel analysis of highly purified proteins, for several reasons. First, and foremost, unlike the detection of binding, or of proteins as substrates of modifying enzymes, detection of enzymatic activity allows the possibility that multiple turnovers of activity can be detected from a single protein molecule. Thus, in effect, the multiple molecules of product that are produced act as an amplifier of the presence of the protein. Second, because the proteins are highly purified, these enzymatic assays can be performed for extended periods of time without destruction of the substrate or the product to obtain the maximum number of turnovers of the enzyme. Third, since many enzymatic activities are catalyzed by single subunit proteins, or have a single subunit that has residual activity in the absence of other stimulatory subunits, these activities are amenable to assay from collections of purified proteins.

Biochemical Genomics Approach

In the biochemical genomics approach, enzymatic activity of proteins is analyzed in liquid solution, typically in pools of expressed and purified proteins (Figure 3). Since many types of enzymatic activity are easily observed in crude extracts, where the protein comprises 0.01% of the protein or less, detection of the activity in a pool containing only 100 different proteins is very easy. The major use of this approach has been with pools of purified yeast ORF fusion proteins prepared from a genomic library of yeast strains expressing GST-ORF fusion proteins (Martzen *et al.*, 1999). The library is comprised of 64 microtiter plates containing 96 yeast strains, each of which bears a plasmid expressing a particular ORF as a GST-ORF fusion protein under P_{CUP1} control, which was constructed by co-transformation of a gapped plasmid and a collection of previously made PCR products directly into yeast. Typically, each pool of purified GST-ORF fusion proteins is derived from the 96 strains present in one microtiter plate. Activity is assayed from the 64 pools of purified proteins, and active pools are deconvoluted to identify the strain and ORF that co-purifies with the particular activity, by preparation and analysis of subpools of GST-ORF fusion proteins.

This procedure has proven to be a powerful method to assay the whole yeast proteome to rapidly link

enzymatic activity and gene function. It has been used to identify several enzymes that modify RNAs and proteins (Alexandrov *et al.*, 2002; Gu *et al.*, 2003; Jackman *et al.*, 2003; Ma *et al.*, 2003, 2005; Martzen *et al.*, 1999; Park *et al.*, 2006; Xing *et al.*, 2002) and enzymes that act in the metabolism of small molecules (Bieganowski and Brenner, 2004; Martzen *et al.*, 1999; Shull *et al.*, 2005). These enzymatic activities comprise a wide variety of different types of reaction, including representatives of 5 of the 6 known classes of enzymes. Indeed, this approach is completely general for any enzymatic activity that can be assayed.

Recently, we have extended this approach by construction and use of the MORF collection of yeast strains (Gelperin *et al.*, 2005). This collection comprises 5,854 strains, each expressing a C-terminal ORF-His_6-HA-3C-ZZ fusion protein under tightly regulated control from the P_{GAL}-promoter. In principle, this library is more complete than the GST-ORF libraries because the ORFs derive from recent annotations (as of August 2002) rather than the original annotations used for the first set of PCR primers (Hudson *et al.*, 1997), which have changed substantially in the interim. Moreover, the C-terminal ORF-fusion proteins are particularly useful for those proteins that are not functional as N-terminal fusions, including many that require a functional N-terminus for entry into the secretory system. Proteins purified from this library are also easily assayed for enzymatic activity (Gelperin *et al.*, 2005).

Extraordinarily high sensitivity can be achieved with this approach, for two reasons. First, activities that are vanishingly small can be detected because the purified proteins in the pools are present at reasonably high concentrations (~20 nM each on average) and can be assayed for extended periods of time without inactivation of the enzyme or substrate and without loss of the product due to degradation. Indeed, we often continue to observe product formation for as long as 18 hours with purified pools, whereas activities in crude extracts (~10 mg/mL, 30 nM each protein) often stop within minutes. Under these conditions, we can easily observe formation of as little as 10^{-17} moles product from a substrate with a single ^{32}P-label, which corresponds to 1/2000 of a turnover of the enzyme during this span of time. The slow loss of activity that is observed during extended incubations might occur because the enzyme is slowly unfolded, or degraded by protease contaminants. It can also occur because of the slow loss of substrate due to its own instability or degradation during the incubation. For example, in the search for specific tRNA modifying activities, we have observed the slow action of contaminating RNases in some pools of purified proteins, which gradually destroys the integrity of the tRNA substrate, preventing its recognition by the modifying enzyme. Second, the purification is so efficient that there is almost no noise in the assays for activity. Although there are some remaining contaminating activities from one-step purification of the pools of GST-ORF and MORF fusion proteins, we estimate that more than 99.9% of the contaminants are removed by the purification, based on assay of phosphatase activity and alcohol dehydrogenase activity in crude extracts and purified pools. This efficiency of purification permits extended incubations of pools without detection of false positives from pools that should not have the catalytic activity. In practice, one can easily observe virtually any activity in the purified pools that is present in the lysate, provided that the ORF-fusion protein is present in the library, is functional as a fusion, and purifies with required necessary components.

This sensitivity is important for detection of activity for four distinct reasons: First, and most important, it allows detection of activity of proteins that are not optimally active or stable under buffer conditions in which they were purified, stored, or assayed. Thus, although the advantage of massive parallel purification of tagged proteins is that they are purified under one set of buffer conditions, these same buffer conditions can result in reduced activity of proteins with special requirements, such as particular metal ions, pH, or salts that have not been included in the purification. Second, the sensitivity allows detection of proteins that do not necessarily have all their cofactors. This occurred with the identification of the pool containing tRNA dihydrouridine synthase activity, because we did not know during our initial screen that the activity was stimulated by FAD in addition to NADPH/$NADH_2$ (Xing *et al.*, 2002). Third, the sensitivity allows detection of activities of proteins that are part of complexes. Although only one member of the complex might be tagged in a particular pool, activity will still be present in the pool as long as the other components co-purify, much as for a co-immunoprecipitation experiment. Since the amount of activity from the complex is limited by the amount of co-purifying components, there is no enrichment of activity derived from overproduction

of the tagged component; in this case, enrichment of activity in the corresponding ORF pool derives solely from the affinity purification. Indeed, we detected both members of the complex required for formation of m^7G_{46} in tRNA in this way (Alexandrov *et al.*, 2002). Fourth, the sensitivity allows new insight into catalytic potential, because it can result in detection of activity from proteins that might normally act on different substrates *in vivo*. Extended incubation of pools of GST-ORF fusion proteins resulted in the detection of two proteins that could catalyze removal of the terminal phosphate from ADP-ribose-1″-phosphate; one of these, Poa1, is highly specific for its substrate; and the other, Hal2, normally acts on a different set of phosphate-containing molecules (Shull *et al.*, 2005).

Enzyme Genomics

A major advance in this discipline is the approach taken by Yakunin and coworkers (2004) to assign enzymatic functions to genes from the large numbers of purified proteins of unknown function that have been acquired from structural genomics initiatives (Kuznetsova *et al.*, 2005). This approach depends on the hypothesis that many enzymes will react with a variety of different substrates to catalyze the same chemical reaction. Thus, in this enzyme genomics approach (Figure 3), purified proteins are assayed for activity using generalized screens to broadly detect enzymatic classes of activity, and then positives are re-screened with more specific assays to determine the best substrate for the enzyme (Kuznetsova *et al.*, 2005). Since oxidoreductases and hydrolases comprise a large fraction of the known enzymes, many of the assays are focused on these activities. One generic substrate used to great effect in this system is paranitrophenylphosphate (PNPP), which is generally active with members of the hydrolase family, is stable for extended periods of time at typical incubation temperatures, and produces a product, paranitrophenol, that absorbs strongly (molar extinction coefficient, 18300 at 410 nm). Because the product can be detected in a standard plate reader, assays for biochemical activity can easily be run in high-throughput mode. Furthermore, since in general 1 μg of purified protein is assayed for 2 to 4 hours, this is sufficient for detection of activity from many phosphatases, including protein phosphatases, which have a high Km for PNPP, and low overall activity with this substrate. Other generic enzymatic screens examined by this group include dehydrogenases, proteases, esterases, and oxidases, all likewise assayed using colorimetric assays performed in high throughput mode.

Using these substrates, 600 proteins were examined for activity, yielding 36 assignments of enzymatic activity, including 2 dehydrogenases, 4 phosphodiesterases, 13 esterases, and 17 phosphatases. Although many of these proteins had conserved sequence motifs indicating the particular class of activity to which they belonged, at least one protein with no observable motif had enzymatic activity, another had been mis-assigned as a phosphoglucomutase instead of a phosphatase, and a third was shown to have a second type of activity. This latter protein, tRNA nucleotidyl transferase, normally adds the CCA to the 3′ end of tRNAs, a polymerase reaction that occurs in the absence of template (Weiner, 2004). Surprisingly, Yakunin and colleagues. (2004) found that this CCA-adding enzyme also has significant 2′-3′ cyclic phosphodiesterase, 2′ nucleotidase and phosphatase activity in its HD domain, consistent with an active role in repairing ends of tRNA damaged by base hydrolysis. Furthermore, for many of the enzymes that were, as predicted, in a certain enzyme class, subsequent examination of different substrates led to assignment of specific biochemical activities that had not been predicted. For example, three enzymes with significant 5′ nucleotidase activity were found from three proteins in different protein families, and each had somewhat different specificities (Proudfoot *et al.*, 2004). Finally, since 15 of the phosphatases had been annotated as putative members of the haloacid dehalogenase (HAD) family, these results provided strong evidence that this family has phosphatase activity, consistent with its previous assignment as part of the superfamily including phosphatases, phosphonatases, and phosphoglucomutases (Koonin and Tatusov, 1994).

The potential for this type of approach is enormous, as increasing numbers of ORFs are cloned and encoded proteins purified and available for analysis. One can imagine development of this methodology in several ways: First, a number of more sensitive assays might be employed using fluorescent substrates; these would require correspondingly less protein. Second, a variety of new assays are likely to be developed and used to expand the repertoire of enzymatic activities that can be addressed with this approach. These might be based on spectrophotometric detection of product, as used until now, or on any other type of detection system that can be conducted in single microtiter plate wells or

by subsequent simple liquid handling steps that can be executed robotically. However, such assays are unlikely to involve extensive subsequent manipulation steps to separate reactants and products, such as electrophoresis or chromatography because of the time and effort required for such analysis, and the advantage of pooling proteins for these activities. Third, since so little protein is required for detection of activity (1 μg or less for standard experiments), proteins can readily be prepared in a high throughput mode for massive parallel analysis of enzymatic activity from expression clones that yield of the order of 1 μg protein per mL or more.

Analysis of Ligand Binding Activity

Genomic approaches have been used with great effect to identify proteins that bind different types of ligands. The biochemical genomics approach was used to identify two previously identified proteins and one new protein that bound to the UAS of *SUC2* DNA (Hazbun and Fields, 2002) as well as 15 proteins that bound tRNA (Wilkinson, Grayhack and Phizicky, unpublished data).

An increasingly powerful method of identifying proteins that bind ligands employs protein microarrays, in which genomic collections of proteins are tethered on glass slides, in microwells or other similar surfaces, and then probed with appropriately labeled ligands (MacBeath and Schreiber, 2000; Zhu *et al.*, 2000, 2001). Using this approach, Zhu and coworkers (2001) detected a number of different binding activities with various biotinylated probes and Cy3-labeled streptavidin. Calmodulin binding was detected for all 6 previously known calmodulin binding proteins that were present in a genomic collection of purified proteins, as well as for 33 newly identified proteins. Similarly, a number of phospholipid binding proteins were detected by use of liposomes containing a biotinylated lipid mixed with various unlabeled phosphoinositides and phosphatidyl choline. Protein microarrays have also been used to detect more than 200 yeast proteins that bind Cy3-labeled single-stranded and double-stranded DNA, about half of which have roles consistent with DNA binding. Moreover, 3 proteins not previously known to bind DNA were found associated with specific DNA sequences *in vivo*, including Arg5,6 a mitochondrial enzyme involved in arginine biosynthesis (Hall *et al.*, 2004). In addition, proteins that bind small molecules have been detected with this method. Thus, analysis of the binding targets of small molecule inhibitors of rapamycin (SMIRs) revealed 30 proteins that bind SMIR4 and one of these, encoded by YBR077w, proved to be responsible for the SMIR4 effect in vivo, based on altered rapamycin sensitivity and altered SMIR4 sensitivity of the corresponding mutant (Huang *et al.*, 2004).

Detection of ligand binding by microarrays is surprisingly sensitive and specific, based on a survey of binding of 59 leucine zipper peptides that were arrayed on glass slides. When these peptides were probed with the same set of peptides labeled with Cy3, binding was detected for pairs with binding constants ranging from 50 nM to 3 μM, and with high specificity (Newman and Keating, 2003). From 3481 pairs of peptides examined, comprising all possible combinations of the peptides, interactions were detected with reasonable signal to noise ratio in 14% of the pairs, and 90% of the interactions were detected in reciprocal fashion. Moreover, signal intensity correlated well with known binding constants for several peptide pairs examined, and a very high frequency of known interactions were detected in two evaluations of the data examined in detail, with almost no false negatives.

The major limit to the use of protein microarrays to detect ligand interactions is that imposed by the physical properties of the interaction, including the binding constant, the amount of immobilized protein on the microarray, and the effective off rate of the interaction in the majority of cases where the bound ligand has to be washed away. Other limits include availablility of appropriate fluorescent or labeled derivatives, availability of adequate amounts of purified protein, lack of knowledge of appropriate binding conditions, and retention of activity during immobilization of the protein.

Analysis of Protein-Protein Interactions

A special case of protein-ligand interactions is the definition of protein-protein interactions. Because of the central importance of protein complexes in all aspects of macromolecular organization, a number of different studies have been devoted to this problem, including several genomic two-hybrid screens (Ito *et al.*, 2001; Uetz *et al.*, 2000), which are not discussed further here, and several affinity purification schemes (Gavin *et al.*, 2002; Ho *et al.*, 2002). The most comprehensive

study to date is the recent genomic definition of 491 yeast protein complexes, obtained by purification of 1,993 proteins from 3206 attempted purifications of yeast strains bearing chromosomally TAP-tagged ORFs (Gavin *et al.*, 2006). Since the proteins are present at their native levels in chromosomally tagged ORFs, the complexes that are detected have a high likelihood of being correct, just as in a co-IP experiment. Consistent with this, the study resulted in the detection of 73% of known complexes annotated in MIPS or annotated anecdotally, and duplicate purifications resulted in 69% of proteins identified in both purifications. Additionally, some false positives could be convincingly eliminated based on their co-purification with numerous complexes. Also consistent with correct assignment of complexes is the correlation of complexes with co-expression, co-localization, protein abundance, similarity of annotated function, and occurrence of orthologous pairs. By contrast, the failure to detect complexes can be ascribed to a number of factors, other than failure of the tags to produce functional protein. Rarer protein complexes are more difficult to detect by these methods, as demonstrated by their underrepresentation in the set of complexes (Gavin *et al.*, 2006). In addition, complexes are difficult to detect if they are regulated in any way, such as by developmental stage, growth cycle, cell cycle, or nutrient status, or if they are inherently transient in nature. Furthermore, complexes are difficult to detect if they are weak, or if they involve sub-stoichiometric amounts of proteins. Weak interactions can be immediately lost during the dilution necessary to lyse the cells, and wash the columns, and non-stoichiometric purifications can easily be mis-annotated as background.

IDENTIFICATION OF MODIFIED PROTEINS AND SUBSTRATE PROTEINS OF MODIFICATION ENZYMES

Proteins undergo a huge number of different modifications, including phosphorylation, ubiquitination, sumoylation, methylation, acetylation, and palmitoylation, as well as many others. Since proteins have such a large number of modifications, and since activity is frequently modulated by these modifications, a large amount of research is devoted to unraveling the identity of the modifications, the identity of the proteins that are modified, the location of the modifications on the proteins, and the enzymes that catalyze the modifications. The development of genomic libraries of strains expressing tagged proteins has led to new and powerful methods for identification of modified proteins and for identification of substrates of protein modification enzymes.

Identification of Proteins with Specific Modifications

One method of detecting modifications employs Western analysis of small pools of immunopurified proteins from the collection of chromosomally TAP-tagged yeast strains (Figure 4), as was done to identify proteins covalently modified with SUMO (Wykoff and O'Shea, 2005). In this study, the original TAP-tagged collection of strains was modified to include an *ulp1*^{ts} allele that prevents degradation of SUMO-modified proteins, and to express high levels of SUMO protein tagged with an HA epitope, which is not present in the TAP-tag. Then, the TAP-tagged proteins were immunopurified with IgG, and proteins were probed for the presence of the SUMO HA tag. In this way, 4,034 strains were analyzed, yielding 13 proteins covalently modified with SUMO, 5 of which were new. The success of this approach underscores the value of individual query of proteins on a high throughput scale, as an

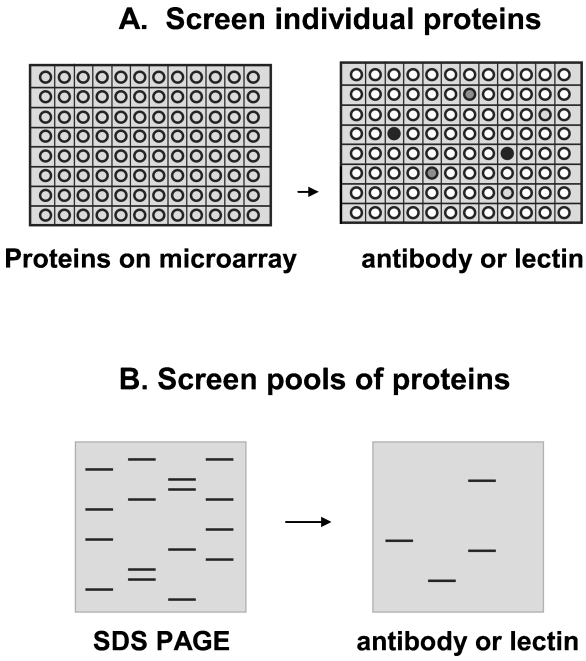

Figure 4 **Detection of proteins with modifications. Proteins with post-translational modifications can be detected directly by analysis of protein microarrays with appropriate antibody, lectin, *etc.*, or by screens of pools of purified proteins with similar reagents, and subsequent deconvolution.**

alternative to mass spectrometry analysis of the purified portion of the proteome containing the modification (Panse *et al.*, 2004; Wohlschlegel *et al.*, 2004; Zhou *et al.*, 2004), although the individual query approach is clearly more time consuming.

A second method for detection of modified proteins features protein microarray analysis of individually purified proteins (Figure 4), as was done to identify proteins covalently modified by glycosylation (Gelperin *et al.*, 2005). In this experiment, 5,573 proteins from the MORF collection were purified individually and applied to protein microarrays, and then probed with antibody against whole cells with exposed α 1–6 mannose linkages. Of 509 putative glycosylated proteins detected, 55 of 136 that were previously known were identified based on stringent cut-off values, and a further 109 of 344 proteins tested were confirmed to be glycosylated because their mobility on Westerns was sensitive to treatment with Endo H and PNGase F, which remove N glycans.

Large Scale Screening of Purified Proteins with Specific Modifying Enzymes

The most direct method for identification of protein substrates of modifying enzymes is the screening of purified proteins with the modification enzymes. This approach was recently used with a large set of candidate proteins to identify substrate proteins of the yeast Rsp5 ubiquitin ligase. In this assay, biotinylated ubiquitin is reacted with GST-ORF fusion proteins in the presence of Rsp5 and cofactors, and ubiquitinylated proteins are detected by a luminescence assay based on proximity of two beads: a chemically modified streptavidin bead binding the biotin moiety of the ubiquitin, and a chemically modified anti-GST bead binding the GST-ORF fusion protein (Kus *et al.*, 2005). Of 130 randomly chosen GST-ORF fusion proteins and 58 GST-ORF fusion proteins containing the PY motif implicated as a target of the Rsp5 WW domain, a number of potential new substrate proteins were detected, at least 7 of which were confirmed, based on genetic interactions.

Protein microarrays have also recently been used to identify substrates of modifying enzymes, enabling a huge gain in throughput. By direct examination of protein microarrays containing ~4400 purified GST-ORF fusion proteins, Snyder and coworkers identified 1325 proteins that were substrates of one or more of 87 yeast protein kinases examined, comprising a total of 4200 protein kinase-protein substrate pairs (Ptacek *et al.*, 2005). Of these pairs, 1384 occur between kinases and substrates in the same compartment, and 768 involve kinases and substrates in the same functional categories. Although several of the identified substrates were confirmed by examination of appropriate mutants, for the large majority it was not possible to attempt such confirmation; thus the frequency of false positives is not yet known.

Large-Scale Screening of Proteins Incubated with Specific Modifying Enzymes in Extracts

Screening of cell lysates of tagged strains for potential substrate proteins of modification enzymes is the closest approximation to direct examination of substrates in their natural milieu. This approach was first used by Morgan and coworkers to identify substrates of the cyclin-dependent kinase Cdc28 in complex with its cyclin Clb2 (Ubersax *et al.*, 2003). Cell lysates of strains expressing individual GST-ORF fusion proteins were incubated with an engineered variant of the Cdc28-Clb2 kinase complex that could use N_6-benzyl ATP (unlike the vast majority of other protein kinases), followed by purification of the GST-ORF fusion protein, and analysis of incorporated phosphate from the γ-^{32}P-labeled N_6-benzyl ATP. Using this assay, 181 proteins exhibited significant phosphorylation among 695 ORFs tested, and at least 12 of 35 targets examined were confirmed by in vivo analysis. Subsequent analysis demonstrated that at least 36 of 150 substrates were more specific for Cdc28-Clb5 (Loog and Morgan, 2005).

A variation of this approach involves examination of lysates containing TAP-tagged proteins expressed at their normal cellular levels, as was recently done by O'Shea and coworkers to identify substrates of Pho85-Pcl1 (Dephoure *et al.*, 2005). As described above, Pho85 was first engineered so that it could use N_6-benzyl ATP. Then pools of lysates were incubated with Pho85-Pcl1 and γ-^{32}P-labeled N_6-benzyl ATP, and TAP-tagged proteins were immunopurified, resolved by SDS-PAGE, and examined for phosphorylation. From 4250 strains examined, 55 positive pools were obtained, resulting in the identification of 34 TAP-tagged source strains as potential kinase substrates, 24 of which were shown directly to contain the

TAP-tag because the immunopurified phosphorylated protein was reduced in size after treatment with TEV protease to remove the tag. Presumably the other 10 proteins represent co-purifying polypeptides. Many of the identified substrates were judged to be authentic based on their specificity in vitro for Pho85-Pcl1 but not Pho85-Pho80, as well as their known localization and known function.

Since virtually any kinase can be modified to use N_6-benzyl ATP (Bishop *et al.*, 2001), and since examination of activity of each of two kinases in lysates yields a large fraction of valid or likely valid substrates (Dephoure *et al.*, 2005; Loog and Morgan, 2005; Ubersax *et al.*, 2003), this method is likely to be completely generalizable to the examination of substrates of kinases.

There are several limitations of this approach for identification of substrates of modifying proteins. First, detection of the modifications requires that the proteins examined are not already completely modified. Thus, this method may work best for modifications such as regulatory modifications that are not normally present at full levels. For those modifications that are complete or nearly complete, one might be able to maximize signal by growth of cells in conditions that prevent the modification or with mutations that limit or remove the modification. Alternatively, one might be able to treat the samples with enzymes or chemicals to effectively remove the modification before screening. Second, each modification to be detected requires a convenient detection system. In the case of kinases, this is relatively simple because phosphorylated substrates are available at high specific activity. For other modifications, the detection of modified substrate proteins with radioactivity might be more difficult, but could still be accomplished with increased amount of sample extract. Other detection systems employing antibodies or mass spectrometry can also be envisioned.

CONCLUSIONS AND FUTURE PROSPECTS

It is clear from the discussion above that the last decade has produced major advances in our ability to clone large numbers of genes in parallel, to express and purify proteins in parallel at high levels, and to examine biochemical activity in multiple different ways. The next decade holds nearly as much promise. It seems likely that there will be significant advances in large-scale preparation of proteins by cell free protein synthesis, by extension of methods to synthesize proteins directly on a protein microarray (Ramachandran *et al.*, 2004), and by the development of efficient homologous expression systems. In addition, assessment of biochemical activity is likely to be advanced by new methods to detect different enzymatic activities using sensitive fluorescent substrates, development of enzymatic assays compatible with microarrays, and the continued development of methods for examining activities of complexes of proteins or metabolic pathways (Jung and Stephanopoulos, 2004).

ACKNOWLEDGMENTS

We thank M. Wilkinson, Y. Kon, L. Kotelawala, I. Chernyakov, E. Quartley, J. Babulski, M. White, M. Dumont and J. Jackman for many fruitful discussions. This work was supported by Grants GM52347 and HG02311 to EMP and grant GM74899 to G. De Titta.

REFERENCES

Alexandrov, A., Martzen, M.R., and Phizicky, E.M. 2002. Two proteins that form a complex are required for 7-methylguanosine modification of yeast tRNA. *RNA* 8:1253–1266.

Aslanidis, C. and de Jong, P.J. 1990. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res* 18:6069–6074.

Aslanidis, C., de Jong, P.J., and Schmitz, G. 1994. Minimal length requirement of the single-stranded tails for ligation- independent cloning (LIC) of PCR products. *PCR Methods Appl* 4:172–177.

Baca, A.M. and Hol, W.G. 2000. Overcoming codon bias: a method for high-level overexpression of *Plasmodium* and other AT-rich parasite genes in *Escherichia coli*. *Int J Parasitol* 30:113–118.

Bieganowski, P. and Brenner, C. 2004. Discoveries of nicotinamide riboside as a nutrient and conserved NRK genes establish a Preiss-Handler independent route to NAD^+ in fungi and humans. *Cell* 117:495–502.

Bishop, A.C., Buzko, O., and Shokat, K.M. 2001. Magic bullets for protein kinases. *Trends Cell Biol* 11:167–172.

Braun, P. and LaBaer, J. 2003. High throughput protein production for functional proteomics. *Trends Biotechnol* 21:383–388.

Chance, M.R., Bresnick, A.R., Burley, S.K., Jiang, J.S., Lima, C.D., Sali, A., Almo, S.C., Bonanno, J.B., Buglino, J.A., Boulton, S., *et al.* 2002. Structural genomics: a pipeline for providing structures for the biologist. *Protein Sci* 11:723–738.

Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Savchenko, A., Cort, J.R., Booth, V., Mackereth, C.D., Saridakis, V., Ekiel, I., *et al.* 2000. Structural proteomics of an archaeon. *Nat Struct Biol* 7:903–909.

Dephoure, N., Howson, R.W., Blethrow, J.D., Shokat, K.M., and O'Shea, E.K. 2005. Combining chemical genetics and proteomics to identify protein kinase substrates. *Proc Natl Acad Sci U S A* 102:17940–17945.

Dieckman, L., Gu, M., Stols, L., Donnelly, M.I., and Collart, F.R. 2002. High throughput methods for gene cloning and expression. *Protein Expr Purif* 25:1–7.

Dricot, A., Rual, J.F., Lamesch, P., Bertin, N., Dupuy, D., Hao, T., Lambert, C., Hallez, R., Delroisse, J.M., Vandenhaute, J., *et al.* 2004. Generation of the *Brucella melitensis* ORFeome version 1.1. *Genome Res* 14:2201–2206.

Gavin, A., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L., Bastuch, S., Dumpelfeld, B., *et al.* 2006. Proteome

survey reveals modularity of the yeast cell machinery. *Nature* (in press).

Gavin, A.C., Bosche, 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.

Gelperin, D.M., White, M.A., Wilkinson, M.L., Kon, Y., Kung, L.A., Wise, K.J., Lopez-Hoyo, N., Jiang, L., Piccirillo, S., Yu, H., *et al.* 2005. Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes Dev* 19:2816–2826.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. 2003. Global analysis of protein expression in yeast. *Nature* 425:737–741.

Gu, W., Jackman, J.E., Lohan, A.J., Gray, M.W., and Phizicky, E.M. 2003. $tRNA^{His}$ maturation: an essential yeast protein catalyzes addition of a guanine nucleotide to the 5′ end of $tRNA^{His}$. *Genes Dev* 17:2889–2901.

Hall, D.A., Zhu, H., Zhu, X., Royce, T., Gerstein, M., and Snyder, M. 2004. Regulation of gene expression by a metabolic enzyme. *Science* 306:482–484.

Hartley, J.L., Temple, G.F., and Brasch, M.A. 2000. DNA cloning using in vitro site-specific recombination. *Genome Res* 10:1788–1795.

Hazbun, T.R. and Fields, S. 2002. A genome-wide screen for site-specific DNA-binding proteins. *Mol Cell Proteomics* 1:538–543.

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.

Huang, J., Zhu, H., Haggarty, S.J., Spring, D.R., Hwang, H., Jin, F., Snyder, M., and Schreiber, S.L. 2004. Finding new components of the target of rapamycin (TOR) signaling network through chemical genetics and proteome chips. *Proc Natl Acad Sci U S A* 101:16594–16599.

Hudson, J.R. Jr., Dawson, E.P., Rushing, K.L., Jackson, C.H., Lockshon, D., Conover, D., Lanciault, C., Harris, J.R., Simmons, S.J., Rothstein, R., *et al.* 1997. The complete set of predicted genes from *Saccharomyces cerevisiae* in a readily usable form. *Genome Res* 7:1169–1173.

Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'Shea, E.K. 2003. Global analysis of protein localization in budding yeast. *Nature* 425:686–691.

Hunt, I. 2005. From gene to protein: a review of new and enabling technologies for multi-parallel protein expression. *Protein Expr Purif* 40:1–22.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98:4569–4574.

Jackman, J.E., Montange, R.K., Malik, H.S., and Phizicky, E.M. 2003. Identification of the yeast gene encoding the tRNA m^1G methyltransferase responsible for modification at position 9. *RNA* 9:574–585.

Jenny, R.J., Mann, K.G., and Lundblad, R.L. 2003. A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. *Protein Expr Purif* 31:1–11.

Jung, G.Y. and Stephanopoulos, G. 2004. A functional protein chip for pathway optimization and in vitro metabolic engineering. *Science* 304:428–431.

Kapust, R.B. and Waugh, D.S. 1999. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci* 8:1668–1674.

Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T., and Yokoyama, S. 1999. Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett* 442:15–19.

Kim, D.M., Kigawa, T., Choi, C.Y., and Yokoyama, S. 1996. A highly efficient cell-free protein synthesis system from *Escherichia coli*. *Eur J Biochem* 239:881–886.

Kim, Y., Dementieva, I., Zhou, M., Wu, R., Lezondra, L., Quartey, P., Joachimiak, G., Korolev, O., Li, H., and Joachimiak, A. 2004. Automation of protein purification for structural genomics. *J Struct Funct Genomics* 5:111–118.

Koonin, E.V. and Tatusov, R.L. 1994. Computer analysis of bacterial haloacid dehalogenases defines a large superfamily of hydrolases with diverse specificity. Application of an iterative approach to database search. *J Mol Biol* 244:125–132.

Kus, B., Gajadhar, A., Stanger, K., Cho, R., Sun, W., Rouleau, N., Lee, T., Chan, D., Wolting, C., Edwards, A., *et al.* 2005. A high throughput screen to identify substrates for the ubiquitin ligase Rsp5. *J Biol Chem* 280:29470–29478.

Kuznetsova, E., Proudfoot, M., Sanders, S.A., Reinking, J., Savchenko, A., Arrowsmith, C.H., Edwards, A.M., and Yakunin, A.F. 2005. Enzyme genomics: Application of general enzymatic screens to discover new enzymes. *FEMS Microbiol Rev* 29:263–279.

LaBaer, J., Qiu, Q., Anumanthan, A., Mar, W., Zuo, D., Murthy, T.V., Taycher, H., Halleck, A., Hainsworth, E., Lory, S., *et al.* 2004. The *Pseudomonas aeruginosa* PA01 gene collection. *Genome Res* 14:2190–2200.

Liu, Q., Li, M.Z., Leibham, D., Cortez, D., and Elledge, S.J. 1998. The univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes. *Curr Biol* 8:1300–1309.

Loog, M. and Morgan, D.O. 2005. Cyclin specificity in the phosphorylation of cyclin-dependent kinase substrates. *Nature* 434:104–108.

Luan, C.H., Qiu, S., Finley, J.B., Carson, M., Gray, R.J., Huang, W., Johnson, D., Tsao, J., Reboul, J., Vaglio, P., *et al.* 2004. High-throughput expression of *C. elegans* proteins. *Genome Res* 14:2102–2110.

Ma, X., Yang, C., Alexandrov, A., Grayhack, E.J., Behm-Ansmant, I., and Yu, Y.T. 2005. Pseudouridylation of yeast U2 snRNA is catalyzed by either an RNA-guided or RNA-independent mechanism. *EMBO J* 24:2403–2413.

Ma, X., Zhao, X., and Yu, Y.T. 2003. Pseudouridylation (Psi) of U2 snRNA in *S. cerevisiae* is catalyzed by an RNA-independent mechanism. *EMBO J* 22:1889–1897.

MacBeath, G. and Schreiber, S.L. 2000. Printing proteins as microarrays for high-throughput function determination. *Science* 289:1760–1763.

Madin, K., Sawasaki, T., Ogasawara, T., and Endo, Y. 2000. A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes. *Proc Natl Acad Sci U S A* 97:559–564.

Marsischky, G. and LaBaer, J. 2004. Many paths to many clones: a comparative look at high-throughput cloning methods. *Genome Res* 14:2020–2028.

Martzen, M.R., McCraith, S.M., Spinelli, S.L., Torres, F.M., Fields, S., Grayhack, E.J., and Phizicky, E.M. 1999. A biochemical genomics approach for identifying genes by the activity of their products. *Science* 286:1153–1155.

Mohanty, A.K., Simmons, C.R., and Wiener, M.C. 2003. Inhibition of tobacco etch virus protease activity by detergents. *Protein Expr Purif* 27:109–114.

Muhlrad, D., Hunter, R., and Parker, R. 1992. A rapid method for localized mutagenesis of yeast genes. *Yeast* 8:79–82.

Nallamsetty, S. and Waugh, D.S. 2006. Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expr Purif* 45:175–182.

Newman, J.R. and Keating, A.E. 2003. Comprehensive identification of human bZIP interactions with coiled-coil arrays. *Science* 300:2097–2101.

Panse, V.G., Hardeland, U., Werner, T., Kuster, B., and Hurt, E. 2004. A proteome-wide approach identifies sumoylated substrate proteins in yeast. *J Biol Chem* 279:41346–41351.

Park, J.H., Aravind, L., Wolff, E.C., Kaevel, J., Kim, Y.S., and Park, M.H. 2006. Molecular cloning, expression, and structural prediction of deoxyhypusine hydroxylase: a HEAT-repeat-containing metalloenzyme. *Proc Natl Acad Sci U S A* 103:51–56.

Phizicky, E., Bastiaens, P.I., Zhu, H., Snyder, M., and Fields, S. 2003. Protein analysis on a proteomic scale. *Nature* 422:208–215.

Proudfoot, M., Kuznetsova, E., Brown, G., Rao, N.N., Kitagawa, M., Mori, H., Savchenko, A., and Yakunin, A.F. 2004. General enzymatic screens identify three new nucleotidases in *Escherichia coli*. Biochemical characterization of SurE, YfbR, and YjjG. *J Biol Chem* 279:54687–54694.

Ptacek, J., Devgan, G., Michaud, G., Zhu, H., Zhu, X., Fasolo, J., Guo, H., Jona, G., Breitkreutz, A., Sopko, R., *et al.* 2005. Global analysis of protein phosphorylation in yeast. *Nature* 438:679–684.

Ramachandran, N., Hainsworth, E., Bhullar, B., Eisenstein, S., Rosen, B., Lau, A.Y., Walter, J.C., and LaBaer, J. 2004. Self-assembling protein microarrays. *Science* 305:86–90.

Reboul, J., Vaglio, P., Rual, J.F., Lamesch, P., Martinez, M., Armstrong, C.M., Li, S., Jacotot, L., Bertin, N., Janky, R., *et al.* 2003. *C. elegans* ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. *Nat Genet* 34:35–41.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. 1999. A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* 17:1030–1032.

Rual, J.F., Hirozane-Kishikawa, T., Hao, T., Bertin, N., Li, S., Dricot, A., Li, N., Rosenberg, J., Lamesch, P., Vidalain, P.O., *et al.* 2004. Human ORFeome version 1.1: a platform for reverse proteomics. *Genome Res* 14:2128–2135.

Rubio, V., Shen, Y., Saijo, Y., Liu, Y., Gusmaroli, G., Dinesh-Kumar, S.P., and Deng, X.W. 2005. An alternative tandem affinity purification strategy applied to *Arabidopsis* protein complex isolation. *Plant J* 41:767–778.

Sanishvili, R., Wu, R., Kim, D.E., Watson, J.D., Collart, F., and Joachimiak, A. 2004. Crystal structure of *Bacillus subtilis* YckF: structural and functional evolution. *J Struct Biol* 148:98–109.

Service, R.F. 2002. Structural genomics. Tapping DNA for structures produces a trickle. *Science* 298:948–950.

Shull, N.P., Spinelli, S.L., and Phizicky, E.M. 2005. A highly specific phosphatase that acts on ADP-ribose 1''-phosphate, a metabolite of tRNA splicing in *Saccharomyces cerevisiae*. *Nucleic Acids Res* 33:650–660.

Tyler, R.C., Aceti, D.J., Bingman, C.A., Cornilescu, C.C., Fox, B.G., Frederick, R.O., Jeon, W.B., Lee, M.S., Newman, C.S., Peterson, F.C., *et al.* 2005. Comparison of cell-based and cell-free protocols for producing target proteins from the *Arabidopsis thaliana* genome for structural studies. *Proteins* 59:633–643.

Ubersax, J.A., Woodbury, E.L., Quang, P.N., Paraz, M., Blethrow, J.D., Shah, K., Shokat, K.M., and Morgan, D.O. 2003. Targets of the cyclin-dependent kinase Cdk1. *Nature* 425:859–864.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., *et al.* 2000. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403:623–627.

Vinarov, D.A., Lytle, B.L., Peterson, F.C., Tyler, E.M., Volkman, B.F., and Markley, J.L. 2004. Cell-free protein production and labeling protocol for NMR-based structural proteomics. *Nat Methods* 1:149–153.

Walhout, A.J., Temple, G.F., Brasch, M.A., Hartley, J.L., Lorson, M.A., van den Heuvel, S., and Vidal, M. 2000. GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol* 328:575–592.

Wei, C., Lamesch, P., Arumugam, M., Rosenberg, J., Hu, P., Vidal, M., and Brent, M.R. 2005. Closing in on the *C. elegans* ORFeome by cloning TWINSCAN predictions. *Genome Res* 15:577–582.

Weiner, A.M. 2004. tRNA maturation: RNA polymerization without a nucleic acid template. *Curr Biol* 14:R883–885.

Wohlschlegel, J.A., Johnson, E.S., Reed, S.I., and Yates, J.R., 3rd. 2004. Global analysis of protein sumoylation in *Saccharomyces cerevisiae*. *J Biol Chem* 279:45662–45668.

Wykoff, D.D. and O'Shea, E.K. 2005. Identification of sumoylated proteins by systematic immunoprecipitation of the budding yeast proteome. *Mol Cell Proteomics* 4:73–83.

Xing, F., Martzen, M.R., and Phizicky, E.M. 2002. A conserved family of *Saccharomyces cerevisiae* synthases effects dihydrouridine modification of tRNA. *RNA* 8:370–381.

Yakunin, A.F., Proudfoot, M., Kuznetsova, E., Savchenko, A., Brown, G., Arrowsmith, C.H., and Edwards, A.M. 2004. The HD domain of the *Escherichia coli* tRNA nucleotidyltransferase has 2',3'—cyclic phosphodiesterase, 2' —nucleotidase, and phosphatase activities. *J Biol Chem* 279:36819–36827.

Yokoyama, S. 2003. Protein expression systems for structural genomics and proteomics. *Curr Opin Chem Biol* 7:39–43.

Zhou, W., Ryan, J.J., and Zhou, H. 2004. Global analyses of sumoylated proteins in *Saccharomyces cerevisiae*. Induction of protein sumoylation by cellular stresses. *J Biol Chem* 279:32262–32268.

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., *et al.* 2001. Global analysis of protein activities using proteome chips. *Science* 293:2101–2105.

Zhu, H., Klemic, J.F., Chang, S., Bertone, P., Casamayor, A., Klemic, K.G., Smith, D., Gerstein, M., Reed, M.A., and Snyder, M. 2000. Analysis of yeast protein kinases using protein chips. *Nat Genet* 26:283–289.

Editor: Marvin Wickens