

Proteomics in living cells

Stephen W. Michnick

The study of protein dynamics *in vivo* is central to understanding the chemical machinery that makes up the process of life: in determining how, when, where and under what conditions proteins are created, form complexes, are modified and disintegrate. Until recently, facile methods to study these processes in intact living cells did not exist. This review describes promising experimental strategies that have been devised to visualize the dynamic nature of proteins and protein complex formation occurring in living cells. The approaches described are based largely on genetically encoded fusions to proteins of interest combined with labelling strategies. The feasibility of employing such strategies in genome-wide biochemical pathway mapping efforts is discussed.

S.W. Michnick
Département de Biochimie
Université de Montréal
C.P. 6128
Montréal, Québec
Canada H3C 3J7
e-mail: stephen.michnick@umontreal.ca

▼ We are witnessing a fundamental change in the way that biomedical research is carried out. In the recent past, it was enough to say that a normal or pathological cellular process was understood if the components of that process could be isolated and their behaviour studied out of the context of the intact cell. This has always presented a paradox: does what we see *ex cellulo* really represent what is happening in a living cell? All of this is changing, thanks to rapid developments and improvements in probe reagents, optical technologies and software available for studying proteins in intact cells or whole organisms. The term 'proteomics' can imply studying either the whole or a large subset of proteins in the cell. However, it can equally imply the study of protein dynamics. By studying protein dynamics I mean determining when, where and under what circumstances proteins are expressed, are covalently modified and subsequently interact with other molecules. From a practical point of view, proteomics implies studying proteins on a large scale; that is, studying many proteins under many conditions. Although we are beginning to witness some considerable success in studying protein dynamics on a large scale using *in vitro* techniques, notably those based on mass-spectroscopic approaches, we are far from

achieving the same levels and detail of analysis with cell-based assays. In this review, I will focus on the recent literature about technologies that are not necessarily ready for prime time but which may be useful for large-scale analyses of protein dynamics in intact cells or have the technical hallmarks of simple, scalable assays.

The hallmarks of scalable cell-based assays

In thinking about this review, I asked myself what were the criteria for deciding what were good existing or potential cellular proteomics methods, based on four basic elements: directness of the method of measurement, portability, simplicity and scalability. Bearing in mind that our goal is to ask what, where, when and in what form and in the company of what proteins a particular protein functions, we can list the following criteria:

- The state of a protein is detected directly, not through secondary events such as transcription activation.
- Genes are expressed in the relevant cellular context, reflecting the native state of the protein with the correct post-translational modifications.
- Changes in protein dynamics induced by changes in the cell state, such as the cell cycle or response to hormones, growth factors or unnatural perturbations such as small molecule inhibitors or small interfering RNAs (siRNAs), can be detected, providing target validation by linking proteins to specific pathways. For example, using insulin-sensitive cells, changes in protein dynamics, such as interactions with other proteins stimulated by insulin, can be detected and quantitated.
- Any method should provide information on at least one key aspect of protein dynamics, but preferably several, including cellular location, turnover, post-translational modifications or interactions with proteins or other molecules.

As a final point, I should mention emerging technologies for performing cell-based assays on a large scale. Whatever the merits of individual assays, the goal of following the myriad of events that occur during the lifetime of a protein would require one to monitor the effects of several perturbations in cell populations as well as, if possible, in individual cells or subpopulations. Standard technologies such as flow cytometry and certainly microscopy do not lend themselves to the practical realization of this goal, in terms neither of time nor cost. However, emerging technologies based on microfluidics or solid-phase arraying of cells promise to make large-scale analysis of protein dynamics in intact living cells a practical possibility.

Genetically encoded markers of protein dynamics

The simplest and most specific approaches to date for monitoring protein dynamics in cells are based on genetically encoded reporters; that is, the fusion of a nucleotide sequence of a reporter gene to that coding for a protein of interest and the autologous expression of the fusion product in some cell of interest. The most notable examples of such reporters are the natural autofluorescent proteins (AFPs) of the green fluorescent protein family and their artificially engineered variants. The numerous examples of their utilization as protein and cell compartment markers have been extensively reviewed [1]. In spite of their great success, there are three downsides to the use of AFPs in the study of protein dynamics in cells. First, the large size (~238 amino acids), slow folding and maturation, stable structure and ability to homo-oligomerize, even for those species designed not to do so, might result in considerable changes in the dynamics of a protein to which the AFP is fused. Another limitation of AFPs is the spectral properties of these proteins. Extensive protein engineering has been done with AFPs, resulting in those with improved folding and maturation characteristics, reduced tendencies to oligomerize as well as red-shifted emission maxima (for improved signal over cellular background fluorescence and tissue penetration) and higher quantum yield. AFPs have also been engineered to be less sensitive to changes in the physical state of the intracellular milieu, such as pH [2]. Nevertheless, the production of alternatives to AFPs as genetically encoded markers for proteins has been motivated by two considerations: first, the desire to produce markers with specific and optimal spectral characteristics; second, to produce simpler and smaller markers that would be less likely to change the dynamics of the proteins to which they are fused. Several markers that possess one or more of these properties have been reported recently. Chemical fluorophores are ideal because they can be designed to have the most desirable spectral characteristics by functionalizing

several basic fluorophores. Moreover, general strategies have been developed to assure that fluorophores can enter the cell while at the same time limiting their exit. The problem then is to get them to associate in a specific way with a protein of interest. One recently demonstrated strategy is based on the clever utilization of the nucleotide-acylating enzyme *O*⁶-alkylguanine alkyltransferase (AGT) [3]. This strategy takes advantage of the enzyme's ability to transfer an alkyl group (normally a methyl group) from the *O*⁶ position of guanine to a cysteine on the surface of the protein to form a stable thioether. In the scheme presented in Figure 1a, an *O*⁶-benzyl derivative of a fluorophore or other derivative of guanine is used to label the protein. The authors demonstrate their ability to specifically label bacterially expressed proteins and nuclear localization peptides in mammalian cells. In principle, labels with different fluorophores could be synthesized to produce proteins labelled in different ways. One obvious advantage of this approach is that, in theory, any fluorophore could be specifically associated with a protein of interest and with the most desirable spectroscopic properties. There are a couple of downsides to this approach. First, mammalian cells express an AGT that can act as an acceptor of *O*⁶-benzylguanine derivatives, resulting in background labelling. Thus, experiments must be performed in cells in which the gene has been knocked out. Second, AGTs, like the AFPs, are relatively large (human AGT used in these studies has 207 amino acids), begging the question of whether they also could alter protein dynamics.

The problem of size could be solved with the genetic encoding of relatively small and linear peptides that bind covalently or noncovalently to specific probes. Roger Tsien and his group have reported a novel strategy to stably label very small peptide tags in living cells [4–7]. The principle is based on the formation of stable complexes of arsenic with two thiols. A peptide tag containing between six and 17 amino acids, including the tetracysteine motif Cys-Cys-Xaa-Xaa-Cys-Cys, will react with biarsenide derivatives of several fluorophores, forming stable complexes between the peptide and fluorophore and, in the chosen example (Figure 1b), resulting in a fluorophore of sufficiently increased quantum yield over that of the uncomplexed fluorophore to detect and localize tagged proteins in living cells. Simple cell-permeable fluorophores that fluoresce in the green (fluorescein) [6], red (resorufin [4] and Nile red [8]) and blue (3,6-dihydroxyxanthone and 2,7-dichloro derivative) [4,5] have been described and the utilization of these tags to study protein turnover and movement in the cell has been presented, most notably in an elegant study of the subcellular trafficking of the tight junction-associated protein calnexin 43 [5]. This strategy has already been

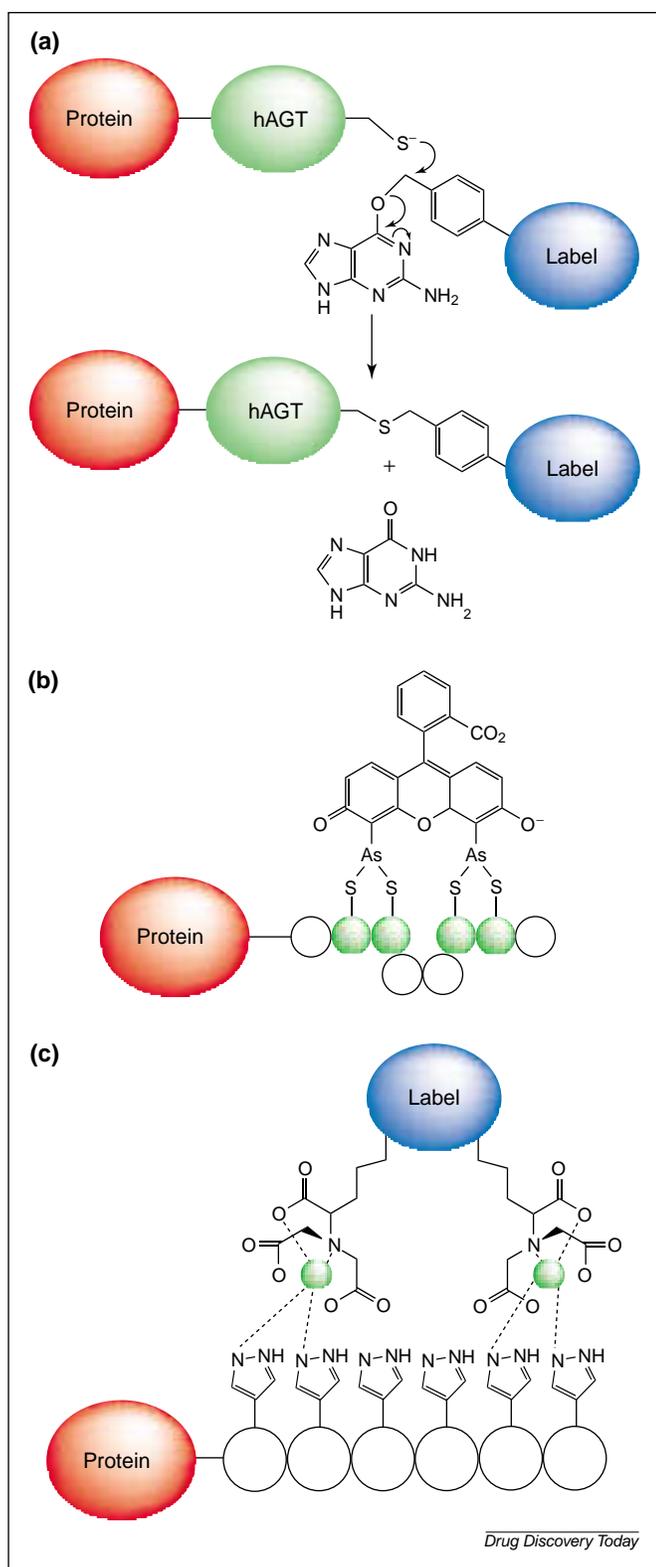


Figure 1. Genetically encoded protein fluorescent labelling strategies. The red ellipsoid represents the tagged protein. (a) Human alkylguanine alkyltransferase (hAGT) transfers the fluorescently labelled benzyl derivative of a fluorophore from the *O*⁶ position of guanine to a cysteine on the surface of the protein to form a stable thioether. (b) A peptide tag containing the tetracysteine motif Cys-Cys-Xaa-Xaa-Cys-Cys reacts with biarsenide derivatives of a fluorophore (fluorescein here) to form a stable complex between the peptide and fluorophore, resulting in a fluorophore of sufficiently increased quantum yield over that of the uncomplexed fluorophore to detect and localize tagged proteins in living cells. (c) Complexes of transition metal-nitrilotriacetic acid (Ni^{2+} :NTA) with fluorophores. Ni^{2+} :NTA can form a stable complex with a hexahistidine peptide fused to a protein. (d) Small acid-rich peptides form specific complexes with tryptophan- or tyrosine-activated luminescent lanthanides (e.g. Tb^{3+} , grey ball).

of dyes is limited only by the practical requirements of the system; the complexes form very rapidly (within seconds, compared to AFPs which require minutes to reach maximum fluorescence) and are stable for days, allowing for very long-term analysis of protein dynamics.

It would be hoped that the design of the biarsenate fluorophores and their use with small genetically encoded tags will inspire chemists to develop other *in vivo* small peptide-specific labelling strategies. Alternative peptides and labels could allow for multiplexed analysis of the dynamics of many proteins at a time, as well as potentially allowing protein co-localization and, in some cases, protein-protein interactions to be followed. Two examples have been demonstrated *in vitro* that have some promise for use *in vivo* in some form. One is based on complexes of transition metal-nitrilotriacetic acid (Ni^{2+} :NTA) with common cyanine fluorochrome (Cy3 and Cy5) chromophores. Ni^{2+} :NTA can form stable complexes with a hexahistidine peptide which can be incorporated anywhere in a peptide (Figure 1c) [9]. Another strategy is based on engineering acid-rich peptides that can form specific complexes with tryptophan- or tyrosine-activated luminescent lanthanides (Figure 1d) [10]. It is likely that these approaches are sufficiently robust to be scaled to the study of many proteins, but it remains to be seen how the dyes can be used in different cell types. The problems of differences in uptake of the fluorophores as well as devising ways to get heavy metals into cells (e.g. using artificial ionophores) and overcoming their toxicity in different cell types would need to be addressed. Labelling techniques that do not require metal coordination would be a preferred alternative.

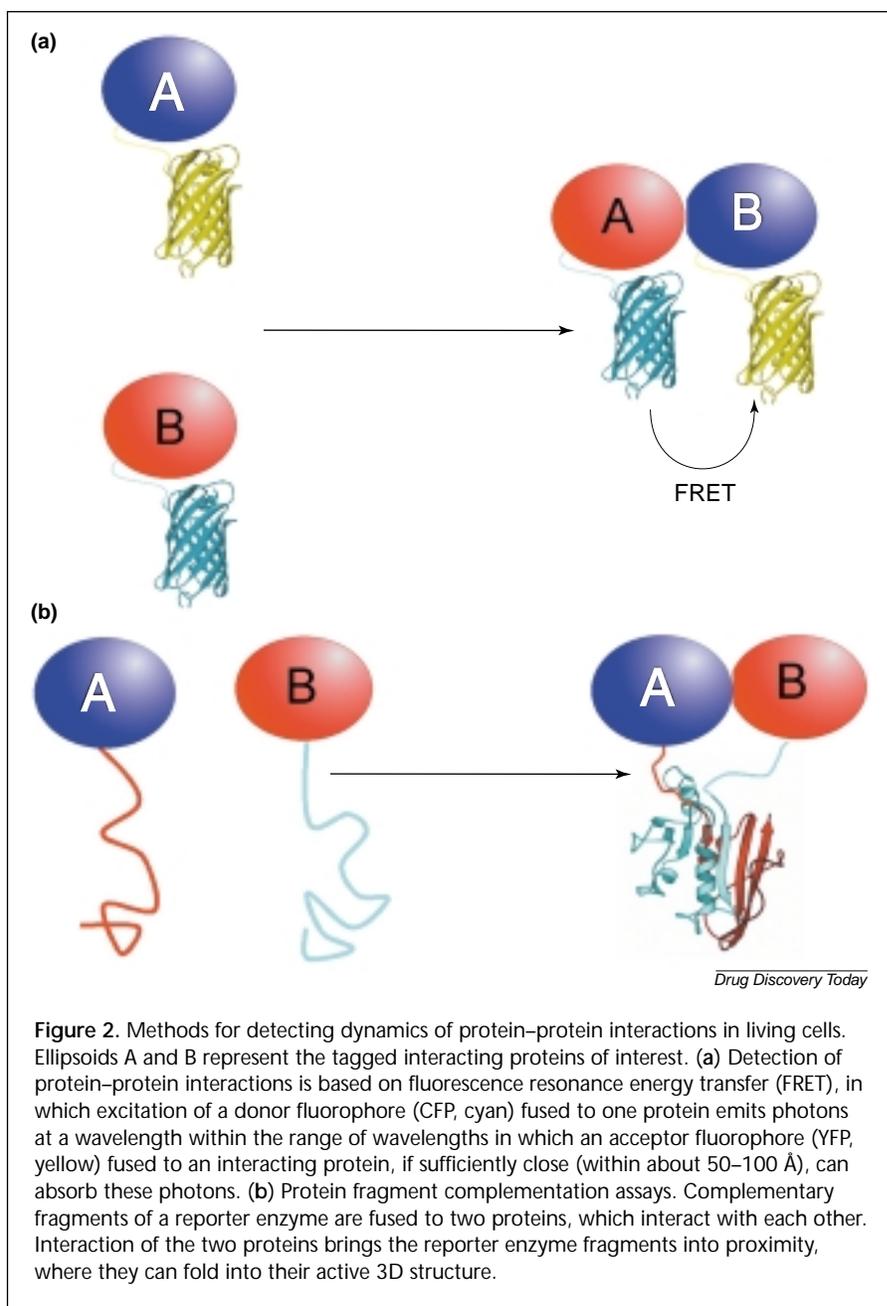
demonstrated to have significant advantages over autofluorescent proteins for the monitoring of protein dynamics. Besides the fact that the tag peptide itself is dramatically smaller than AFPs, thus preventing spurious effects on protein turnover and trafficking, the range of spectral properties

An alternative to protein or peptide labelling that holds some promise for multiplexed labelling is suggested by recent advances in the development of so-called 'quantum dots' (QDs). QDs are semiconductor nanoparticles (~10⁻⁹ m in diameter), which can be manufactured in such a way as to emit fluorescence at distinct wavelengths. If the surfaces

of individual QDs (having a unique fluorescence emission maximum) could be chemically modified to interact with specific proteins, in principle it could be possible to follow the dynamics of a series of proteins, perhaps in an individual cell at a time. Until recently, technical hurdles existed in the use of QDs *in vivo*, such as simple ways to introduce them into cells or to derivatize their surfaces chemically so that they could be used to label proteins in a specific way. However, recent examples of both cell-surface and intracellular protein labelling show that these problems are surmountable [11–13]. Even so, QDs are enormous compared with small molecule fluorophores, being comparable in size to, or larger than, proteins. By contrast, chemical fluorophores are about 100 times smaller than proteins. Nevertheless, QD conjugates with protein-specific antibodies or peptide or protein aptamers might prove powerful tools for parallel analysis of protein levels locations and modifications within individual cells or populations of cells, as has been demonstrated as a proof-of-principle with fluorescently labelled antibodies [14].

Protein–protein interactions

The interactions that proteins make with other molecules and in particular with other proteins tell us what proteins are doing in the cell; thus both the identification of protein–protein interactions and the study of their dynamics on a large scale are a key goal of proteomics. In living cells, the ways in which interactions can be directly measured are limited. As mentioned above, AFPs have been – and selectively labelled peptide tags fused to proteins could be – used to measure protein–protein interactions based on fluorescence resonance energy transfer (FRET), a weak quantum mechanical phenomenon in which excitation of a donor fluorophore emits photons at a wavelength within the range of wavelengths in which an acceptor fluorophore, if sufficiently close ($\sim 50\text{--}100\text{ \AA}$), can absorb these photons (Figure 2a) [15]. However, FRET is insensitive, has a very narrow dynamic range and is very difficult to scale because the two proteins must be expressed in the cell at



Drug Discovery Today

optimal levels to ensure that perturbation of the acceptor by donor fluorophore can be detected. It is difficult to imagine obtaining optimal levels of expression of two proteins simultaneously on a large scale. Protein interaction-assisted folding of reporter enzymes from fragments or protein fragment complementation assays (PCA) provide an alternative to FRET that overcomes this problem [16–22]. In this approach, complementary fragments of a reporter enzyme are fused to two proteins that are thought to interact with each other. Interaction of the two proteins brings the reporter enzyme fragments into proximity where they can fold into their active, three-dimensional structure (Figure 2b). The all-or-none nature of protein folding assures

that the dynamic range of response is maximal (in principle, if the two proteins do not interact before some triggering event, there is no signal in the absence of the trigger, whereas the triggered interaction results in maximal signal). Further, no careful adjustment of signal of expression levels is necessary and thus pairwise analysis of constitutive or induced interactions is possible. Nevertheless, the same caution must be exercised in interpreting PCA results as with any genetically encoded probe. In particular, care must be taken to ensure that the localization of protein complexes is not altered by attachment of reporter fragments to the two proteins of interest. Furthermore, interactions should not be interpreted as being 'binary'. In the cell, a third protein could mediate an interaction; in which case the interaction is indirect. The only way to prove that an interaction observed by PCA is binary is through direct binding assays with proteins expressed in an exogenous background. Several assays have been developed, including those based on ubiquitin, dihydrofolate reductase, β -lactamase and AFPs [16–25]. Proof-of-principle examples of the uses of these approaches to systematically study both induced and constitutive interactions have been demonstrated, and the potential of these methods to both detect novel interactions and map out biochemical pathways has been realized [25]. We should see more examples of applications of PCA to larger-scale mapping of biochemical pathways and networks in the near future.

Outstanding issues

I have no doubt that we are close to having technologies that can be applied on a large scale to the study of protein dynamics. In particular, fluorescent labelling of small, genetically encoded peptides provides a promising means of studying protein trafficking and turnover, while PCAs could prove the most general way to monitor the dynamics of protein complexes. However, we know that the proteome is more than the sum of its parts, their dynamics and interactions. How a protein functions is dependent on its state and position within a cell at a given time, and is affected by the conditions that the individual cell or cell population is under at that time. The state of a protein is often determined by specific post-translational covalent modifications. Unfortunately, there are no methods to determine exactly how a specific protein is modified in intact cells, although, in a sense, the question is meaningless at this point because we must first develop ways to determine what modifications proteins undergo during their life cycle; these problems are best addressed by mass-spectroscopic approaches [26]. Nevertheless, methods to elucidate key functional post-translational modifications can be used to monitor the activity of specific biochemical pathways. For

example, I alluded earlier to the use of antibodies that are highly specific for covalently modified peptide sequences on specific proteins to label cells in which the modification has occurred under a particular condition [14]. Fluorescent labelling of specific antibodies with chemical or QD chromophores would in principle allow the state of proteins under a wide range of conditions to be monitored, but the scale of such studies would be limited by the availability of highly specific antibodies. Alternatively, FRET-based detectors of covalent modifications of peptides that correspond to sequences in proteins known to be modified in specific ways could serve as indirect but not completely specific monitors of post-translational modifications [27,28].

In addition, a goal of all proteomics research is to determine the organization of biochemical pathways and networks. This will not be possible without a set of complementary technologies that allow us to observe changes in protein dynamics in conjunction with perturbations of pathways. Among the most obvious perturbation strategies to emerge in the last few years has been siRNAs that are finding use in knocking down the levels of expression of individual genes [29–31]. However, gene knockdowns can take a long time to have an effect on a pathway because of the relatively slow turnover of the protein knocked down, and there is always the possibility of a compensating mechanism that might hide the effects of the knockout. Direct knockout or knockdown of proteins has been achieved using fluorescence-mediated protein inactivation. In this approach, a fluorophore is covalently associated with a protein via either an AFP, one of the small genetically encoded peptide-fluorophore complexes mentioned above or a fluorophore-labelled protein-specific antibody. Excitation of the fluorophore results in emission of photons that excite molecular oxygen into a free radical triplet state which can then react with and hydrolyse peptide or other side chain groups in the vicinity of the fluorophore (over ~ 50 Å), thus destroying either the integrity or the catalytic activity of the tagged protein or those with which it is in contact. Several examples of this approach, including a proof-of-principle of its utility on a larger scale, have been reported recently [32–34]. Alternatives to siRNA or destructive protein knockdown approaches are chemical or genetically encoded inhibitors of protein function that have immediate effects on gene activity. An example of chemically encoded knockouts are those devised by Kevan Shokat to specifically inhibit one class of enzymes, protein kinases [35]; examples of genetically encoded peptide aptamer strategies have also been devised although they have not yet been demonstrated to be useful on a large scale [36,37].

Finally, the development of protein detection technologies must be married to technologies that will allow large-scale

screening of many proteins in cells that can be perturbed by siRNAs, drugs, hormones, etc., by analogy to the way that analysis of gene expression with DNA microarrays has been used to map pathways and compare normal and pathological states of cells. Given the costs of maintaining living cells and studying multiple cellular processes, this issue must be addressed soon. Several groups have demonstrated microcellular array and liquid handling on nanolitre scales that should prove to meet this challenge [38–40].

References

- Zhang, J. *et al.* (2002) Creating new fluorescent probes for cell biology. *Nat. Rev. Mol. Cell Biol.* 3, 906–918
- Heikal, A.A. *et al.* (2000) Molecular spectroscopy and dynamics of intrinsically fluorescent proteins: coral red (dsRed) and yellow (Citrine). *Proc. Natl. Acad. Sci. U. S. A.* 97, 11996–12001
- Keppler, A. *et al.* (2003) A general method for the covalent labeling of fusion proteins with small molecules *in vivo*. *Nat. Biotechnol.* 21, 86–89
- Adams, S.R. *et al.* (2002) New biarsenical ligands and tetracysteine motifs for protein labeling *in vitro* and *in vivo*: synthesis and biological applications. *J. Am. Chem. Soc.* 124, 6063–6076
- Gaietta, G. *et al.* (2002) Multicolor and electron microscopic imaging of connexin trafficking. *Science* 296, 503–507
- Griffin, B.A. *et al.* (1998) Specific covalent labeling of recombinant protein molecules inside live cells. *Science* 281, 269–272
- Griffin, B.A. *et al.* (2000) Fluorescent labeling of recombinant proteins in living cells with FLAsH. *Methods Enzymol.* 327, 565–578
- Nakanishi, J. *et al.* (2001) Imaging of conformational changes of proteins with a new environment-sensitive fluorescent probe designed for site-specific labeling of recombinant proteins in live cells. *Anal. Chem.* 73, 2920–2928
- Kapanidis, A.N. *et al.* (2001) Site-specific incorporation of fluorescent probes into protein: hexahistidine-tag-mediated fluorescent labeling with (Ni(2+):nitrilotriacetic acid (n)-fluorochrome conjugates. *J. Am. Chem. Soc.* 123, 12123–12125
- Franz, K.J. *et al.* (2003) Lanthanide-binding tags as versatile protein coexpression probes. *ChemBiochem* 4, 265–271
- Wu, X. *et al.* (2003) Immunofluorescent labeling of cancer marker Her2 and other cellular targets with semiconductor quantum dots. *Nat. Biotechnol.* 21, 41–46
- Larson, D.R. *et al.* (2003) Water-soluble quantum dots for multiphoton fluorescence imaging *in vivo*. *Science* 300, 1434–1436
- Jaiswal, J.K. *et al.* (2003) Long-term multiple color imaging of live cells using quantum dot bioconjugates. *Nat. Biotechnol.* 21, 47–51
- Perez, O.D. and Nolan, G.P. (2002) Simultaneous measurement of multiple active kinase states using polychromatic flow cytometry. *Nat. Biotechnol.* 20, 155–162
- Miyawaki, A. and Tsien, R.Y. (2000) Monitoring protein conformations and interactions by fluorescence resonance energy transfer between mutants of green fluorescent protein. *Methods Enzymol.* 327, 472–500
- Johnsson, N. and Varshavsky, A. (1994) Split ubiquitin as a sensor of protein interactions *in vivo*. *Proc. Natl. Acad. Sci. U. S. A.* 91, 10340–10344
- Pelletier, J.N. *et al.* (1998) Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proc. Natl. Acad. Sci. U. S. A.* 95, 12141–12146
- Michnick, S.W. *et al.* (2000) Detection of protein–protein interactions by protein fragment complementation strategies. *Methods Enzymol.* 328, 208–230
- Galarneau, A. *et al.* (2002) Beta-lactamase protein fragment complementation assays as *in vivo* and *in vitro* sensors of protein–protein interactions. *Nat. Biotechnol.* 20, 619–622
- Wehrman, T. *et al.* (2002) Protein–protein interactions monitored in mammalian cells via complementation of beta-lactamase enzyme fragments. *Proc. Natl. Acad. Sci. U. S. A.* 99, 3469–3474
- Spotts, J.M. *et al.* (2002) Time-lapse imaging of a dynamic phosphorylation-dependent protein–protein interaction in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15142–15147
- Remy, I. *et al.* (2002) Detection and visualization of protein interactions with protein fragment. *Methods Mol. Biol.* 185, 447–459
- Hu, C.D. *et al.* (2002) Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol. Cell* 9, 789–798
- Ghosh, I. *et al.* (2000) Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein. *J. Am. Chem. Soc.* 122, 5658–5659
- Remy, I. and Michnick, S.W. (2001) Visualization of biochemical networks in living cells. *Proc. Natl. Acad. Sci. U. S. A.* 98, 7678–7683
- Aebersold, R. and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* 422, 198–207
- Zhang, J. *et al.* (2001) Genetically encoded reporters of protein kinase A activity reveal impact of substrate tethering. *Proc. Natl. Acad. Sci. U. S. A.* 98, 14997–15002
- Ting, A.Y. *et al.* (2001) Genetically encoded fluorescent reporters of protein tyrosine kinase activities in living cells. *Proc. Natl. Acad. Sci. U. S. A.* 98, 15003–15008
- McManus, M.T. and Sharp, P.A. (2002) Gene silencing in mammals by small interfering RNAs. *Nat. Rev. Genet.* 3, 737–747
- Timmons, L. (2002) The long and short of siRNAs. *Mol. Cell* 10, 435–437
- Famulok, M. and Verma, S. (2002) *In vivo*-applied functional RNAs as tools in proteomics and genomics research. *Trends Biotechnol.* 20, 462–466
- Rajfur, Z. *et al.* (2002) Dissecting the link between stress fibres and focal adhesions by CALI with EGFP fusion proteins. *Nat. Cell Biol.* 4, 286–293
- Beck, S. *et al.* (2002) Fluorophore-assisted light inactivation: a high-throughput tool for direct target validation of proteins. *Proteomics* 2, 247–255
- Marek, K.W. and Davis, G.W. (2002) Transgenically encoded protein photoinactivation (FLAsH-FALI): acute inactivation of synaptotagmin I. *Neuron* 36, 805–813
- Bishop, A.C. *et al.* (1998) Design of allele-specific inhibitors to probe protein kinase signaling. *Curr. Biol.* 8, 257–266
- Geyer, C.R. *et al.* (1999) Mutagenesis by peptide aptamers identifies genetic network members and pathway connections. *Proc. Natl. Acad. Sci. U. S. A.* 96, 8567–8572
- Scott, C.P. *et al.* (1999) Production of cyclic peptides and proteins *in vivo*. *Proc. Natl. Acad. Sci. U. S. A.* 96, 13638–13643
- Voldman, J. *et al.* (2002) A microfabrication-based dynamic array cytometer. *Anal. Chem.* 74, 3984–3990
- Ziauddin, J. and Sabatini, D.M. (2001) Microarrays of cells expressing defined cDNAs. *Nature* 411, 107–110
- Eyal, S. and Quake, S.R. (2002) Velocity-independent microfluidic flow cytometry. *Electrophoresis* 23, 2653–2657

Access Drug Discovery Today online at:
<http://www.drugdiscoverytoday.com>