A Fusion of Disciplines: Chemical Approaches to Exploit Fusion Proteins for Functional Genomics

Nils Johnsson*[a] and Kai Johnsson*[b]

KEYWORDS:

functional genomics · fusion proteins · protein labeling · proteomics

Introduction

Understanding the functions and interactions of proteins requires studying their localization, interactions, and activities in vivo and in vitro. Confronted with the large numbers of proteins resulting from the sequencing of entire genomes, cell biologists and protein chemists are struggling to invent more efficient and generally applicable methods of studying proteins. By far the most successful strategy to meet this challenge has been the use of fusion proteins. In this approach the protein of interest is genetically engineered to contain an additional sequence at either the N or C terminus. This so-called tag equips the resulting fusion with a unique property that can be exploited to study certain activities of the protein. In this review we will first give a short overview of the most commonly used tags before focusing on a new class of tags that are not limited to their genetically encoded function but rather serve as general acceptors for synthetic molecules.

Tags of fusion proteins currently have two main applications: they are used in purification schemes and as tools to explore the basic cellular properties of the protein. Examples of tags used in purification schemes are the polyhistidine tag recognized by immobilized metals, glutathione *S*-transferase recognized by immobilized glutathione, and short epitope tags recognized by antibodies.[1–3] Improved purification is sometimes achieved by combining two tags in one fusion protein and utilizing their sequential binding to the two corresponding affinity matrices.[4]

Most applications for the use of fusion proteins to examine biological functions in vivo involve the use of autofluorescent proteins, with the green fluorescent protein (GFP) being the most prominent example.[5] Due to their autofluorescence the fusion proteins can be observed in live cells without further fixation and labeling procedures. Sensitive imaging techniques allow the behavior of the proteins to be recorded in real time, thereby giving new insights into the dynamic distribution and localization of proteins in the cell.[6] More sophisticated strategies of employing fusion proteins as functional probes require the presence of two tags that are linked to two different proteins in the cell. Here, a detectable activity of the two tags is dormant until the two different fusion proteins either co-localize or directly bind to each other. These strategies allow the monitoring of interactions and proximities between proteins; examples include the two-hybrid method, split-protein sensors, and two different fluorescent proteins that allow for fluorescence resonance transfer measurements.[5, 7–10] When two appropriate tags are linked to the N and C termini of the same protein, the interaction between the two tags and the resulting detectable activity depend on the spatial distance between the two tags. The activity of the tags is therefore directly influenced by the conformation of the sandwiched polypeptide and can be used to monitor conformational alterations in this protein.[11–13] Tags that have been used in this approach include the modules of the split-ubiquitin sensor, fragments of fluorescent proteins, and entire fluorescent proteins. As conformational changes of proteins are very often at the center of cellular activity it is important that new assays with high conformational sensitivity will be developed.

With the increasing application of fusion proteins the inherent limitations of this approach tend to be overlooked. In general, the main objection towards this approach is the possibility that the tag might interfere with the function of the investigated protein and that the fusion protein does therefore not reflect the behavior of its unmodified kin. This effect can range from subtle hindrance to complete inactivation. Unfortunately, it is not always trivial to estimate the damage that is inflicted by

[a] *N. Johnsson*
Institute of Toxicology and Genetics
Forschungszentrum Karlsruhe
76021 Karlsruhe (Germany)
Fax: (+49) 7247-823354
E-mail: Nils.Johnsson@itg.fzk.de

[b] *K. Johnsson*
Institute of Molecular and Biological Chemistry
Institute of Biomolecular Sciences
École Polytechnique Fédérale de Lausanne (EPFL)
1015 Lausanne (Switzerland)
Fax: (+41) 21-693-9365
E-mail: Kai.Johnsson@epfl.ch

 DOI: 10.1002/cbic.200200603

attaching the tag to the protein of interest. A straightforward test is to replace the native copy in the cell by the fusion protein. If the protein is essential or if lack of the protein has a measurable phenotype, the influence of the tag on the activity of the protein can be quantified. This works very well in organisms that allow rapid gene replacements or gene knock-outs, like the yeast *Saccharomyces cerevisiae*. The strategies that are pursued in those organisms more resilient to reverse genetics very much depend on what is already known about the protein. However, recent advances in the depletion of cellular activity of a protein by transforming the cells with small interfering RNAs should allow the development of functionality test for fusion proteins in higher eukaryotes.[14, 15] On the other hand, many applications do not require full activity of the protein. Measuring a protein interaction by the two-hybrid method is often incompatible with the natural localization and function of the protein. Still, the insights that are gained with this technique are very often also valid for the unmodified copies.

A second objection concerns the often unphysiological concentrations of the fusion proteins compared to their native copies. In most applications the fusions proteins are expressed ectopically from relatively strong promoters, and overexpressed proteins have a tendency to mislocalize, to capture binding partners, and thereby to influence the regulation or kinetics of the processes they participate in. Keeping the intracellular concentration of the fusion protein as close as possible to that of the wild type is therefore always recommended, although it is not obligatory for all applications.

Despite these objections an impressive proof of the general acceptance of fusion proteins can be found in the efforts to fuse all open reading frames (ORFs) of a given organism to an appropriate tag and to exploit the properties of the tag to investigate certain aspects of the biological functions of the corresponding fusion-protein library. Examples include efforts to construct genome-wide protein interaction maps by using the two-hybrid system, to gain an inventory of all cellular protein complexes by using affinity tags, to observe the intracellular localization of all proteins in the cell with GFP fusions, and to display the entire proteome of an organism as a protein microarray by using a polyhistidine tag and nickel-coated glass slides.[16–22] Most of these comprehensive approaches were restricted to budding yeast because the number of proteins therein (approximately 6000) made it a limited, although still very demanding, exercise to obtain and probe all fusion constructs. As impressive as these efforts were, the genome-wide application of fusion proteins dramatically revealed a shortcoming of all the currently available tags: their limitation to a single, genetically encoded function which can be exploited for functional studies. As a consequence, proteins have to be linked to multiple tags to illuminate their different facets, which is tolerable for the study of individual proteins but painful for genome-wide approaches. Before applying the fusion-protein approach on the genome-wide scale to other, more ORF-rich organisms than yeast, it is therefore worthwhile considering a new and promising class of tags. The new approaches are based on a tag-mediated labeling of fusion proteins with synthetic molecules that transfer a unique and specific property to the fusion protein. Provided that the mechanism of transfer is both broadly applicable, allowing the attachment of chemically diverse compounds that might function as specific probes for a variety of biological questions, and highly specific, allowing for labeling to take place in vivo, a single fusion protein might suffice for studying a variety of different properties of the protein of interest. In the following section we will give an overview of the current trends in this new chemical approach to probe protein function.

Tag-Mediated In Vivo Labeling of Fusion Proteins with Small Molecules

The labeling of proteins with small molecules for functional studies, such as spectroscopic probes or cross-linkers, is one of the cornerstones of protein chemistry. However, the lack of specificity in the underlying chemistry used for traditional protein labeling, which usually targets one or several nucleophilic residues of the protein of interest, makes its application in the living cell or in complex protein mixtures impossible. One approach to address the shortcomings of traditional labeling strategies relies on the specific incorporation of unnatural amino acids in vivo based on suppressor tRNA technology.[23] The enormous potential of using unnatural amino acids for in vivo studies is, however, not yet fully realized due to the enormous technological obstacles that have to be overcome to specifically incorporate the amino acids into proteins in living cells. The use of tags to mediate the labeling of the corresponding fusion proteins with small molecules represents a complementary approach to provide proteins with new functionalities. Currently used approaches for the labeling of fusion proteins, in vitro or in vivo, with small molecules or ligands that carry the desired functionality are based on four different kinds of tags (Figure 1):

A) intein-based labeling of proteins with small molecules,
B) tags that are specifically modified with a small molecule by a third protein,
C) tags that reversibly and specifically bind to a small molecule,
D) tags that irreversibly and specifically bind to a small molecule through covalent-bond formation.

Intein-based labeling of proteins with small molecules

Intein-based labeling of proteins has become an important technique in protein engineering over the last years.[24, 25] In its natural function the intein excises itself in a self-splicing reaction by ligating its N-terminal fusion to its C-terminal fusion. The splicing mechanism in general involves the transient formation of thioesters between cysteine side chains of the intein and of the N terminus of the C-terminal fusion. By using engineered inteins it is possible to isolate the protein of interest as a C-terminal thioester. By applying the native chemical ligation methodology developed by Kent et al.,[27] this thioester can then be reacted with derivatized cysteines to yield the protein of interest with an appropriate label attached to its C terminus.[26–28] This so-called expressed protein ligation is very elegant as the tag removes itself in the process of the labeling reaction and thus avoids any possible interference with the function of the

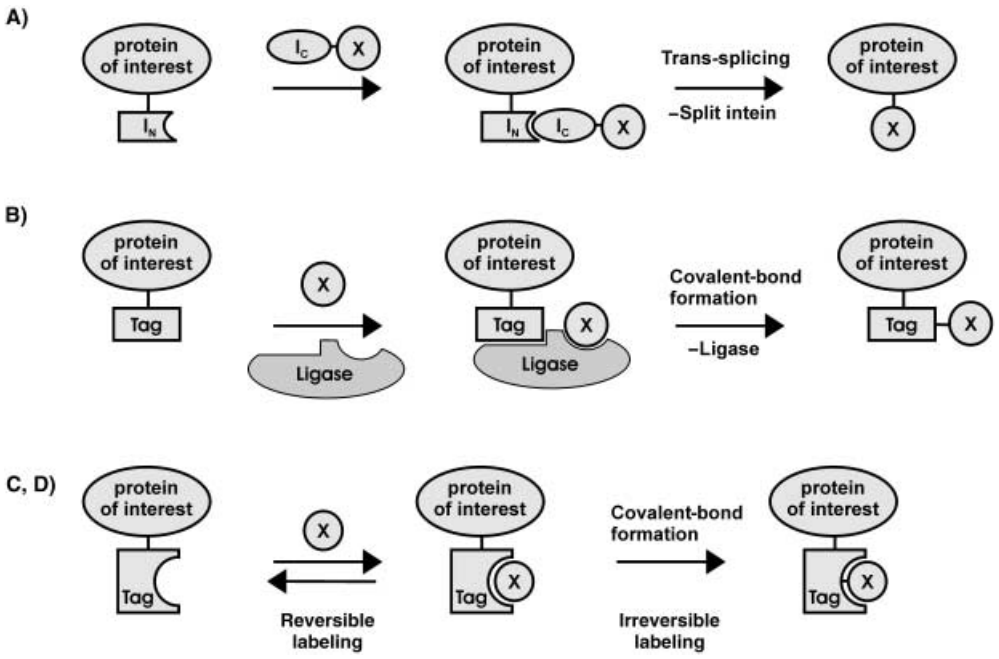


Figure 1. *Schematic representation of the different approaches for tag-dependent labeling of fusion proteins with small molecules in vivo: A) Split intein-based labeling of proteins. The transient conjugation of I_C-X to the protein transduction domain necessary for cellular uptake of I_C-X is omitted; B) tags that are specifically modified by another protein, that is, a ligase, which attaches the small molecule with the desired functionality to the tag; C, D) tags that possess a high affinity towards the small molecule with the desired functionality. The interaction of the tag and the ligand might lead to the formation of a covalent bond and result in irreversible labeling. X = synthetic label; I_N, I_C = N- and C-terminal fragment of the split intein.*

protein of interest. In addition, there are practically no restrictions regarding the nature of the label that can be fused to the C terminus of the protein. Examples of labels range from peptides or whole proteins to fluorophores or affinity labels such as biotin.[26, 29, 30] However, since thioesters react not only with the derivatized cysteine but also with underivatized cysteines, related thiols, and other nucleophiles this strategy can not be employed for the specific labeling of proteins in vivo. To circumvent this problem the group of Muir took advantage of trans-splicing by so-called split inteins (Figure 1 A).[31] In this reaction the individually expressed N-terminal (I_N) and C-terminal (I_C) fragments of the intein spontaneously reassemble into a functional intein. The reassembled intein performs in a second step the usual self-splicing (here: trans-splicing) reaction. To use trans-splicing for in vivo labeling, the protein of interest is expressed as an N-terminal fusion of I_N and I_C is synthesized with a C-terminal extension that carries the desired label to give I_C-X, where X represents a synthetic label. The labeled I_C-X must be subsequently introduced into the cells expressing the N-terminal fusion of I_N. This can be achieved by microinjection or, as shown by Giriat and Muir, with the help of a protein transduction domain (PTD) that was linked by a disulfide bridge to the reactive cysteine of I_C-X.[31] PTDs can travel across the membrane of many cells and deliver I_C-X, after reduction of the disulfide bond in the reducing environment of the cell, to the intracellular I_N fusion protein. The subsequent trans-splicing reaction then leads to the labeling of the protein of interest with X (Figure 1 A). The technique is elegant and general, however, further studies will have to show whether for some applications the nonreacted intracellular I_C-X might not become a source of unwanted background reactions.

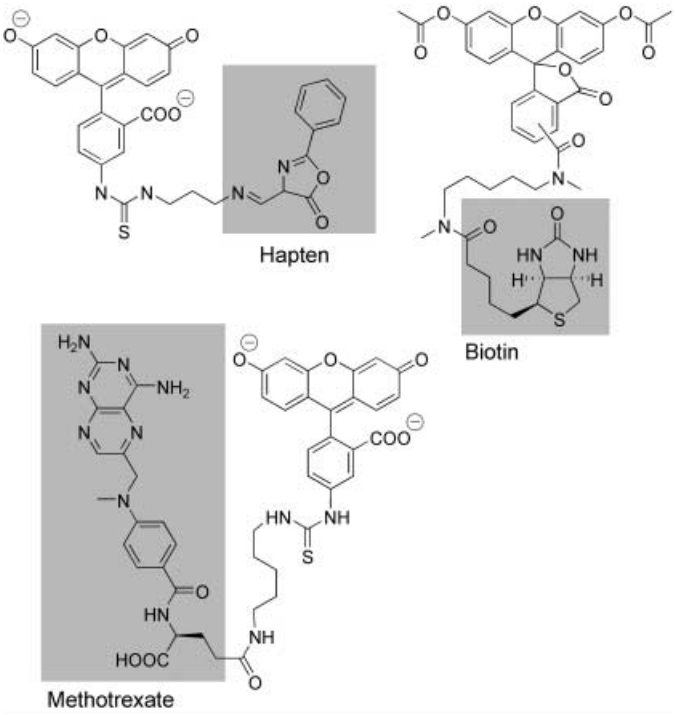
Tags that are specifically modified with a small molecule by a third protein

Most posttranslational modifications of proteins rely on the recognition of a specific sequence motif by an enzyme that performs the corresponding modification. A priori, a strategy in which the labeling of a fusion protein relies on the transfer of a small molecule with the desired functionality to the tag by an appropriate ligase should also be feasible in living cells (Figure 1 B). In order to obtain completely specific labeling, the acceptor domain, that is, the tag, must not be present in any of the other endogenous proteins and the substrate used for ligation to the tag, that is, the label, must not be accepted by any other endogenous proteins with similar ligase activity. So far the only practical example of this approach is the in vivo biotinylation of peptides, so-called biotin acceptor domains, by biotin ligases.[32–34] The biotinylated protein is then recognized by avidin or streptavidin, which bind with high affinity to biotin. While the biotinylation of fusion proteins is attractive due to the availability of various avidin-based reagents and materials, the in vivo labeling itself is restricted to biotin and thus lacks versatility. Furthermore, since biotinylation is a relatively common posttranslational modification, the biotinylated fusion proteins generally have to be at least partially purified before further use. A typical application of biotin acceptor domains is the immobilization of biotinylated fusion proteins on avidin- or streptavidin-coated surfaces for binding assays.[35] Clearly, better tag – ligase pairs are needed to make this approach more general and attractive. Ideally, these pairs are orthogonal to the biochemistry of those organisms commonly investigated in functional genomics and proteomics.

Tags that reversibly and specifically bind to a small molecule

Currently, most approaches for the in vivo labeling of fusion proteins rely on the specific affinity of a given tag towards a synthetic ligand carrying the desired chemical functionality (Figure 1 C, D). Examples for the noncovalent and reversible labeling of (fusion) proteins with reporter molecules include the

avidin – biotin pair, the dihydrofolate reductase – methotrexate pair, and an antibody – hapten pair (Scheme 1).[36–40]



Scheme 1. *Structures of the hapten – fluorescein, biotin – fluorescein, and methotrexate – fluorescein dimers used for the labeling of antibody, avidin, and dihydrofolate reductase fusion proteins, respectively. The part of the dimer recognized by the tag is highlighted. In the case of the biotin – fluorescein dimer, the fluorescein is synthesized as the membrane-permeable diacetate which is rapidly hydrolyzed in vivo.*

In this approach, in contrast to the in vivo biotinylation discussed above, avidin functions as the genetically encoded tag and derivatized biotin acts as the label. The number of possible labels is nearly unlimited as the derivatization of biotin with different functional groups does not interfere with its binding to avidin. The avidin – biotin pair has been used to measure the pH value in the mammalian organelles of the secretory pathway: Avidin fused to an appropriate targeting sequence was expressed in mammalian cells and the cells were subsequently incubated with a membrane-permeable biotin – fluorescein dimer (Scheme 1).[36, 41, 42] The pH dependence of the spectral properties of fluorescein then allowed the pH regulation in these organelles to be studied. However, expression of avidin or streptavidin fusion proteins in the cytosol tends to be toxic for the cell and biotinylated probes have to compete with the natural cofactor for binding to the tag, two features that significantly limit the scope of the approach.[36] Furthermore, both avidin and streptavidin are tetramers and will therefore directly influence the oligomeric state of the attached protein.

An alternative to avidin – biotin is the dihydrofolate reductase (DHFR) – methotrexate pair.[37–39] DHFR fusion proteins bind with high affinity to the cell-permeable methotrexate or its fluorescent derivative (Scheme 1). A first example employed this technique to localize a DHFR fusion protein with a methotrexate – fluorescein dimer and fluorescence microscopy.[37] In such experiments, care has to be taken to wash away excess fluorescent ligand and the measurement has to be performed rapidly after the washing step in order to prevent loss of the label. This points towards a principal problem encountered with tags that bind a label reversibly: If excess label has to be washed away, the fusion protein will, depending on the strength of the binding, also be slowly stripped of its label. This effect significantly restricts the applications of these kinds of tags.

An important application of the noncovalent binding of a ligand to the tag of a fusion protein is the chemically induced dimerization of proteins.[43] Protein activity in biological processes is often triggered by dimerization and the control of this dimerization by small molecules offers intriguing possibilities for the study and manipulation of these processes. Chemical inducers of dimerization (CIDs) are pairs of ligands that are synthetically fused to form dimers or heterodimers. These bivalent molecules can then be used to dimerize the corresponding fusion proteins where the corresponding ligand-binding protein plays the role of the tag (Figure 2). The first example of such a synthetic CID is FK1012 (Figure 2), which is prepared by dimerization of the natural product FK506.[44] FK1012 can be used to selectively dimerize fusion proteins containing the FK506 binding protein. Up to now, CIDs have been used to study signal-transduction mechanisms, to control gene transcription, to induce apoptosis, and to monitor enzymatic reactions.[45–48] In contrast to applications where the

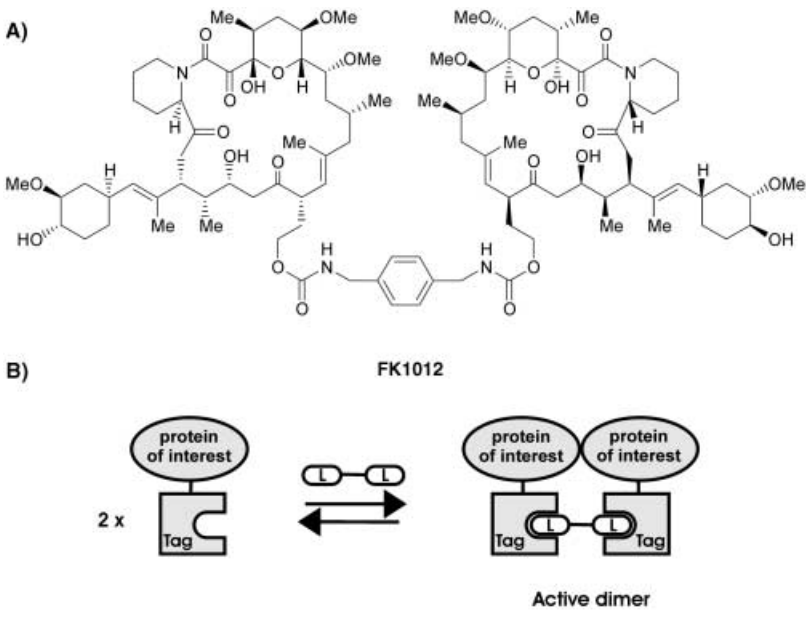
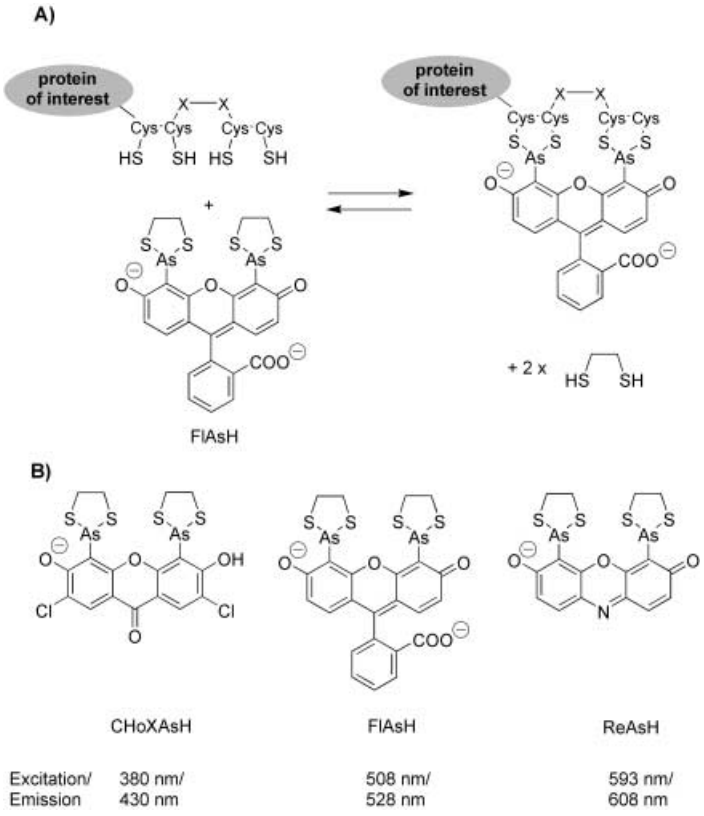


Figure 2. *A) Structure of the chemical inducer of dimerization (CID) FK1012; B) general mechanism of action of CIDs. L = ligand.*

small molecule is a spectroscopic probe, excess CID by itself does not yield a detectable signal and therefore does not have to be washed away.

The first truly multifunctional tag for the reversible and specific labeling of fusion proteins with synthetic molecules is based on the formation of stable complexes between biarsenical compounds such as the fluorescein derivative FlAsH and peptides containing tetracysteine motifs (Scheme 2).[49–52] The



Scheme 2. *Biarsenical dyes for the labeling of tetracysteine motifs: A) Complex formation between FlAsH and the tetracysteine tag of a fusion protein; B) biarsenical dyes used for the in vivo labeling and their excitation and emission maxima. X = any amino acid.*

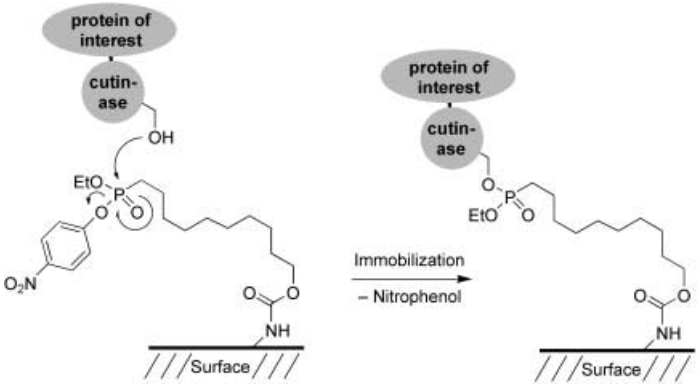
As^{III}-disubstituted tricyclic ring systems allow for a bivalent interaction with the tetracysteine motif: Each of the two As^{III} atoms binds cooperatively to a pair of vicinal cysteines. To ensure specific binding in in vivo applications, the biarsenical compounds have to be added in the presence of 1,2-dithiols such as 1,2-ethanedithiol (EDT). EDT can outcompete endogenous proteins possessing pairs of cysteines or the cellular cofactor lipoamide. Importantly, the formation of the complex between the biarsenical compound and the tetracysteine tag can be reversed by flushing the cells with millimolar concentrations of EDT. So far, applications of this technology have focused on the fluorescence labeling of fusion proteins displaying the tetracysteine motif with blue (CHoXAsH), green (FlAsH), or red (ReAsH) fluorescent biarsenical dyes (Scheme 2).[49–51, 53, 54] In addition, an analogue of nile red with a biarsenical moiety was synthesized.[52] This fluorescent dye possesses a high environmental sensitivity and has been used to study conformational changes of a protein.[52] Important for these applications is the fact that the fluorescence of the dye is efficiently quenched until it is bound to the tetracysteine motif, thereby significantly reducing the background signal originating from nonbound or unspecifically bound dye. The availability of differently colored biarsenical dyes allows the consecutive labeling of fusion proteins in pulse–chase experiments; this offers an exciting opportunity to differentiate old from new copies of a fusion protein. For example, the mechanism of assembly, transport, and dynamics of multiprotein complexes can be studied in such experiments, as demonstrated for the assembly and dynamics of gap-junction plaques.[50] Gap junctions are intercellular channels that are formed from two half-channels, called connexons, of two adjacent cells. Six connexins assemble in the membrane to form a connexon. Gap-junction plaques are tightly packed arrays of such intercellular channels and are known to be very dynamic structures. By expressing a connexin43–tetracysteine fusion protein (Cx43-TC) and subsequently labeling the protein with FlAsH and ReAsH, it was shown that gap-junction plaques are assembled from the outer edge and removed from the central core.[50] Importantly, the tetracysteine tag also allowed verification of these findings with high-resolution electron microscopy imaging, as ReAsH-labeled Cx43-TC can be used for the localized photoconversion of diaminobenzidine into an electron-dense precipitate.[50] It is noteworthy that the images obtained in the electron microscopy measurements with the genetically encoded tetracysteine tag were superior to images obtained with immunogold-labeled antibodies. In addition to the in vivo and in vitro labeling with spectroscopic probes, the tetracysteine tag can also be used to purify the corresponding fusion proteins with the aid of immobilized biarsenicals compounds: The tetracysteine fusion proteins are bound to immobilized biarsenicals and then subsequently eluted with millimolar concentrations of dithiols.[51] From a practical point of view it is also noteworthy that the tetracysteine tag can not only be fused to the N or C terminus of a protein but also be incorporated, due to its small size, in appropriate loops or on outer surfaces of α helices.

What are the drawbacks and limitations of the biarsenical compound–tetracysteine system? Most importantly, specific in vivo labeling is only possible when micromolar concentrations of dithiols are added as agonists and even then the background signal in fluorescence labeling experiments has been reported to be relatively high.[51, 54] It appears that the expression level of the protein of interest, the cell type, the careful control of the concentrations of dithiol, and the addition of nonfluorescent dyes that can suppress nonspecific hydrophobic binding of the biarsenicals are important for the signal-to-noise ratio.[51, 55] The cysteines of the tag also have to be in their reduced form which will make applications in organelles with an oxidizing environment difficult.[51] The simultaneous and specific labeling of two or more fusion proteins possessing tetracysteine tags with different specificities would also open up exciting possibilities but is, at present, not feasible and will most likely require the use of tags based on a completely different chemistry.

Irreversible labeling of fusion proteins

Compared to the approaches discussed above, the irreversible labeling of fusion proteins through tag-mediated covalent-bond formation between the tag and a small molecule offers certain advantages. Particularly important for in vivo applications is the fact that excess label can be readily washed away. One possible approach to achieve such covalent labeling of fusion proteins in living cells or complex mixtures relies on active-site-directed irreversible inhibitors and suicide inhibitors, that is, the irreversible modification of an enzyme through the interaction with a substrate analogue (Figure 1 D). Such irreversible inhibitors have a long tradition in enzymology and there exists a large variety of enzyme – inhibitor pairs.[56] The application of this approach towards the specific labeling of fusion proteins requires the expression of the enzyme or protein as a tag fused to the protein of interest and the synthesis of the corresponding inhibitor fused to an appropriate label. Provided that the label does not interfere with the reaction between the enzyme and the inhibitor, the label will be covalently coupled to the fusion protein. While many irreversible inhibitors for enzymes are known, care has to be taken when choosing an enzyme – inhibitor pair for the labeling of fusion proteins. In addition to the specificity of the enzyme – inhibitor interaction, one also has to take into account the effect the expression of the enzyme can have on the host. For example, the many known protease – inhibitor pairs are probably not good candidates as the proteases can hydrolyze other proteins before being inactivated through labeling.[57] An example of the successful application of an enzyme – inhibitor pair is provided by the selective immobilization of fusion proteins of the esterase cutinase on surfaces displaying alkyl 4-nitrophenyl phosphonates (Scheme 3).[58] Activated phosphonates form covalent adducts with the reactive serine of serine hydrolases, and the specificity of the reaction has been used to selectively immobilize cutinase fusion proteins out of periplasmic extracts of *Escherichia coli* overexpressing the fusion protein. The immobilized cutinase fusion proteins have then been used in functional assays. While this enzyme – inhibitor pair could be used for the labeling of cutinase fusion proteins with a variety of small molecules, a specific labeling in vivo or in more complex mixtures might not be possible as phosphonate inhibitors of this type are known to react with a variety of other serine hydrolases.[59, 60]

A very low intrinsic reactivity of an irreversible inhibitor towards the proteome of a given cell is clearly a prerequisite for its use in the specific and covalent labeling of its corresponding protein tag. One inhibitor – tag pair that fulfills this criterion is based on the unusual mechanism of the DNA repair protein O^6-alkylguanine-DNA alkyltransferase (AGT; Scheme 4).[61] The cellular role of AGT is to irreversibly transfer the alkyl group from alkylated O^6-alkylguanine-DNA to one of its reactive cysteine residues, so that efficient and rapid self-modification leads to the repair of the alkylated DNA.[62] Importantly, the human AGT (hAGT) not only repairs alkylated guanine in DNA but also reacts with the cell-permeable nucleobase O^6-benzylguanine (BG).[62] It was recently shown that fusion proteins of hAGT can be specifically labeled with BG derivatives substituted at the 4-position of the benzyl ring in vivo and in vitro.[61, 63] The rate of the reaction of hAGT with substituted BG derivatives does not significantly depend on the nature of the substituent at the 4-position of the benzyl ring. The specificity of the reaction can be best evaluated in the fluorescence labeling of hAGT fusion proteins in mammalian cells by BGAF. BGAF contains the membrane-permeable diacetate of fluorescein, which is rapidly hydrolyzed inside the cell to the corresponding fluorescein derivative BGFL (Scheme 4). In the experiment, an hAGT fusion protein with a nuclear localization sequence is transiently expressed in AGT-deficient CHO cells. Cell cultures are incubated for short periods of time (5 min) with low concentrations of BGAF ($< 5\ \mu M$) and then washed repeatedly to remove access label. Examination of the cells by fluorescence microscopy shows that only cells transiently expressing the hAGT fusion protein display a fluorescence signal in the nucleus (Figure 3). In contrast, no



Scheme 3. *Immobilization of cutinase fusion proteins on surfaces displaying alkyl 4-nitrophenyl phosphonates.*

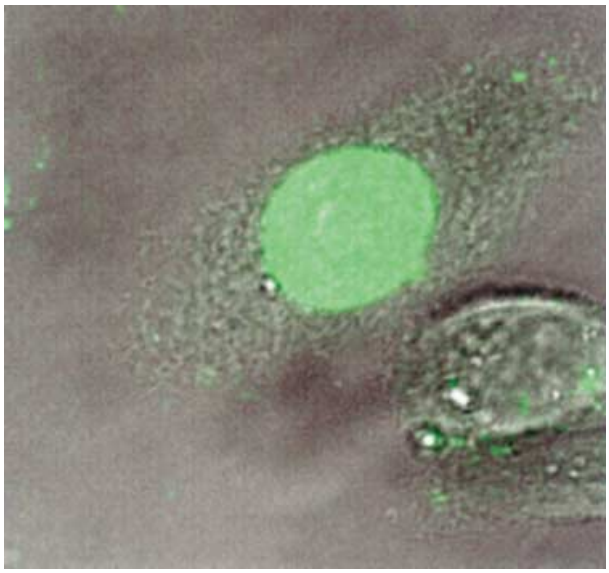
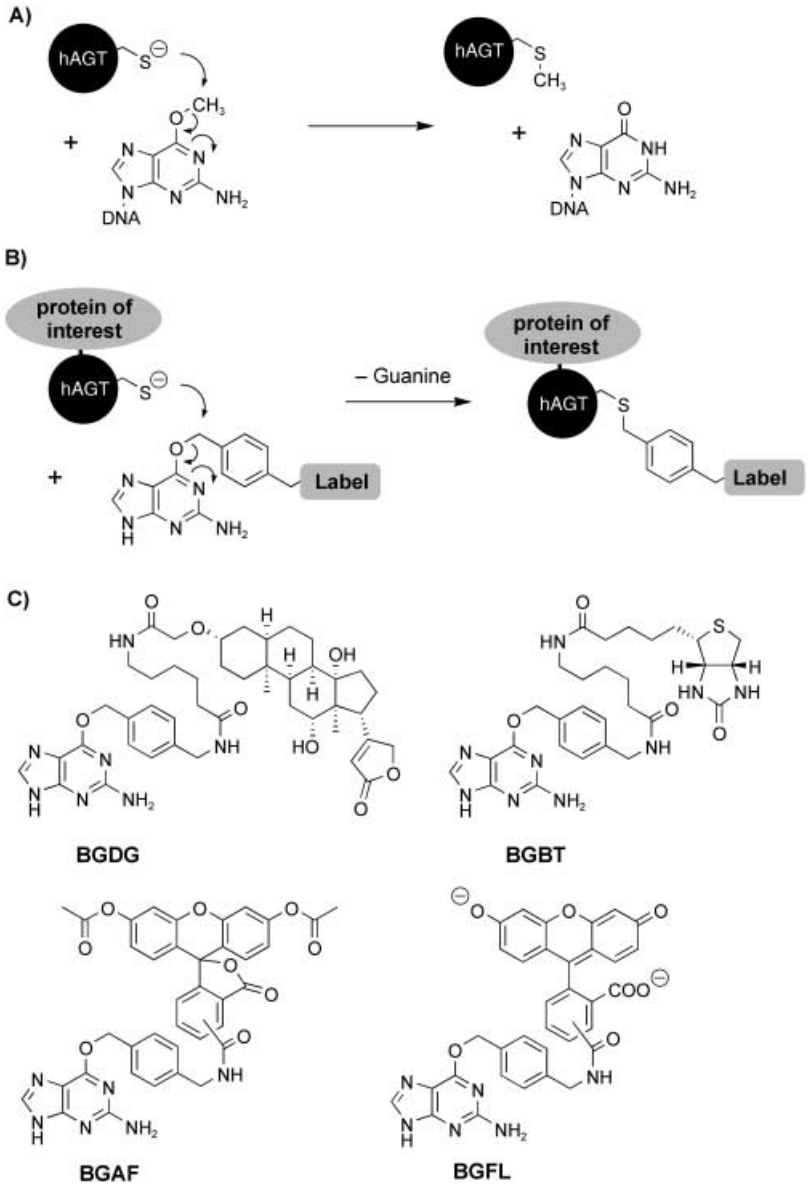
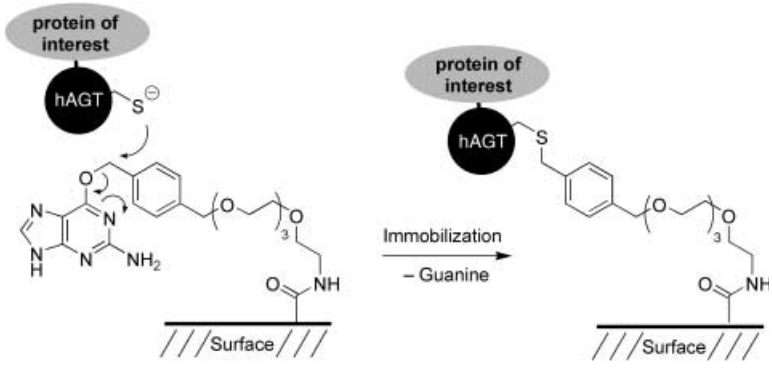


Figure 3. Confocal micrograph of fluorescence-labeled, nuclear-targeted hAGT in AGT-deficient CHO cells. An overlay of the transmission and fluorescence channel is shown (excitation at 488 nm). The image was recorded 30 min after pulse labeling with BGAF (5 μM) for 5 min and 3 washes with phosphate buffered saline.



Scheme 4. *Covalent labeling of hAGT fusion proteins: A) General mechanism for DNA repair of O^6-methylguanine-DNA by hAGT; B) general mechanism for the labeling of hAGT fusion proteins using O^6-benzylguanine (BG) derivatives; C) some BG derivatives used for the labeling of hAGT fusion proteins. BGAF contains the membrane-permeable diacetate of fluorescein, which is rapidly hydrolyzed inside cells by esterases to yield BGFL.*



Scheme 5. *General scheme for the selective immobilization of hAGT fusion proteins on surfaces displaying BG derivatives.*

fluorescence signal is observed in cells where the hAGT fusion was quenched with nonfluorescent BG prior to the labeling with BGAF.

hAGT can be fused to either the N or C terminus of the protein of interest.[61] It is a monomeric protein of 207 amino acids, which can be truncated to 177 residues without significantly affecting the activity against BG derivatives.[62] In vivo labeling is possible in bacterial, yeast, and mammalian cells, and the efficiency of the labeling depends on the membrane permeability of the corresponding BG derivative. The relatively narrow substrate specificities of *E. coli* and yeast AGTs, which do not readily react with BG, allow for a selective labeling of hAGT fusion proteins in these hosts. However, labeling in mammalian cells requires the use of AGT-deficient cell lines. Despite the presence of a reactive cysteine which can render the protein sensitive to oxidative conditions, first experiments indicate that hAGT fusion proteins can be labeled after secretion into the periplasm of *E. coli* or when displayed on the surface of mammalian cells.[64] The reaction between hAGT and BG derivatives is relatively rapid, with the second-order rate constant for the reaction of a recently isolated highly reactive mutant of hAGT with BGBT being 8000 $sec^{-1}\,M^{-1}$.[64] Equally important is the ease of preparation of the BG derivatives, which allow the attachment of a seemingly unlimited number of possible labels to hAGT fusion proteins. Highlighting this versatility is the use of immobilized BG derivatives for a selective, covalent immobilization of hAGT fusion proteins out of *E. coli* cell extracts, an approach that could become important for the generation of protein microarrays (Scheme 5).[65] The main drawbacks of the technology are the size of the tag, which is only slightly below that of GFP, and the requirement to use AGT-deficient mammalian cell lines. However, engineering the substrate specificity and reactivity of hAGT by directed evolution might allow the latter problem to be solved.[64]

Conclusions

Recent developments in the labeling of fusion proteins with synthetic molecules in the living cell or in complex mixtures have opened up new ways for studying protein function. These new tags allow for the attachment of a wide variety of different labels under a broad range of conditions and are predicted to become important for functional genomics, proteomics, and in particular, in the genome-wide preparation of fusion proteins. The tetracysteine tag, which reversibly binds to biarsenical compounds, and the hAGT tag, which

is irreversibly alkylated by benzylguanine derivatives, are prototypes of this approach. Both provide the respective fusion protein with functionalities that can not be genetically encoded and thereby complement the properties of the traditional fusion proteins. Despite this progress, there is still a need for new and alternative labeling schemes of fusion proteins. Very high on the list of desirable techniques is the simultaneous and specific labeling of different fusion proteins within a single cell to measure the spatial and temporal organization of entire pathways. Ideally, these tags would have a minimal influence on the behavior of the fusion protein and could be used in different hosts and organelles. It would then become the ingenuity of chemists and biologists to design and synthesize novel compounds for labeling that will limit the application of this class of fusion proteins.

The authors thank Maik Kindermann, Petra Tafelmeyer, Susanne Gendreizig, and Antje Keppler for critical reading of the manuscript and Antje Keppler and Horst Pick for the preparation of the cover picture.

[1] J. A. Bornhorst, J. J. Falke, *Methods Enzymol.* **2000**, *326*, 245 – 254.
[2] D. B. Smith, *Methods Enzymol.* **2000**, *326*, 254 – 270.
[3] C. E. Fritze, T. R. Anderson, *Methods Enzymol.* **2000**, *327*, 3 – 16.
[4] G. Rigaut, A. Shevchenko, B. Rutz, M. Wilm, M. Mann, B. Seraphin, *Nat. Biotechnol.* **1999**, *17*, 1030 – 1032.
[5] R. Y. Tsien, *Annu. Rev. Biochem.* **1998**, *67*, 509 – 544.
[6] F. S. Wouters, P. J. Verveer, P. I. Bastiaens, *Trends Cell Biol.* **2001**, *11*, 203 – 211.
[7] S. Fields, O. Song, *Nature* **1989**, *340*, 245 – 246.
[8] N. Johnsson, A. Varshavsky, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 10340 – 10344.
[9] S. W. Michnick, *Curr. Opin. Struct. Biol.* **2001**, *11*, 472 – 477.
[10] A. Miyawaki, *Dev. Cell* **2003**, *4*, 295 – 305.
[11] A. Miyawaki, J. Llopis, R. Heim, J. M. McCaffery, J. A. Adams, M. Ikura, R. Y. Tsien, *Nature* **1997**, *388*, 882 – 887.
[12] X. Raquet, J. H. Eckert, S. Muller, N. Johnsson, *J. Mol. Biol.* **2001**, *305*, 927 – 938.
[13] G. Dues, S. Muller, N. Johnsson, *FEBS Lett.* **2001**, *505*, 75 – 80.
[14] S. M. Elbashir, J. Harborth, K. Weber, T. Tuschl, *Methods* **2002**, *26*, 199 – 213.
[15] D. Forler, T. Kocher, M. Rode, M. Gentzel, E. Izaurralde, M. Wilm, *Nat. Biotechnol.* **2003**, *21*, 89 – 92.
[16] M. R. Martzen, S. M. McCraith, S. L. Spinelli, F. M. Torres, S. Fields, E. J. Grayhack, E. M. Phizicky, *Science* **1999**, *286*, 1153 – 1155.
[17] H. Zhu, M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R. A. Dean, M. Gerstein, M. Snyder, *Science* **2001**, *293*, 2101 – 2105.
[18] A. C. Gavin, M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A. M. Michon, C. M. Cruciat, M. Remor, C. Hofert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M. A. Heurtier, R. R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer, G. Superti-Furga, *Nature* **2002**, *415*, 141 – 147.
[19] P. Uetz, L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, J. M. Rothberg, *Nature* **2000**, *403*, 623 – 627.
[20] T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, Y. Sakaki, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4569 – 4574.
[21] D. Q. Ding, Y. Tomita, A. Yamamoto, Y. Chikashige, T. Haraguchi, Y. Hiraoka, *Genes Cells* **2000**, *5*, 169 – 190.
[22] X. Morin, R. Daneman, M. Zavortink, W. Chia, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 15050 – 15055.
[23] Z. Zhang, L. Wang, A. Brock, P. G. Schultz, *Angew. Chem.* **2002**, *114*, 2964 – 2966; *Angew. Chem. Int. Ed.* **2002**, *41*, 2840 – 2842.
[24] C. J. Noren, J. Wang, F. B. Perler, *Angew. Chem.* **2002**, *112*, 458 – 476; *Angew. Chem. Int. Ed.* **2000**, *39*, 450-466.
[25] M. Q. Xu, T. C. Evans, Jr., *Methods* **2001**, *24*, 257 – 277.
[26] T. W. Muir, *Annu. Rev. Biochem.* **2003**, *72*, 175 – 207.
[27] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **1994**, *266*, 776 – 779.
[28] T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705-6710.
[29] T. J. Tolbert, C. H. Wong, *J. Am. Chem. Soc.* **2000**, *122*, 5421 – 5428.
[30] M. L. Lesaicherre, R. Y. Lue, G. Y. Chen, Q. Zhu, S. Q. Yao, *J. Am. Chem. Soc.* **2002**, *124*, 8768-8769.
[31] I. Giriat, T. W. Muir, *J. Am. Chem. Soc.* **2003**, *125*, 7180 – 7181.
[32] J. E. Cronan, Jr., *J. Biol. Chem.* **1990**, *265*, 10327 – 10333.
[33] J. E. Cronan, Jr., K. E. Reed, *Methods Enzymol.* **2000**, *326*, 440 – 458.
[34] M. G. Cull, P. J. Schatz, *Methods Enzymol.* **2000**, *326*, 430 – 440.
[35] P. A. Smith, B. C. Tripp, E. A. DiBlasio-Smith, Z. Lu, E. R. LaVallie, J. M. McCoy, *Nucleic Acids Res.* **1998**, *26*, 1414 – 1420.
[36] M. M. Wu, J. Llopis, S. Adams, J. M. McCaffery, M. S. Kulomaa, T. E. Machen, H. P. Moore, R. Y. Tsien, *Chem. Biol.* **2000**, *7*, 197 – 209.
[37] D. I. Israel, R. J. Kaufman, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 4290 – 4294.
[38] I. Remy, S. W. Michnick, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 5394 – 5399.
[39] I. Remy, S. W. Michnick, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 7678 – 7683.
[40] J. Farinas, A. S. Verkman, *J. Biol. Chem.* **1999**, *274*, 7603 – 7606.
[41] M. M. Wu, J. Llopis, S. R. Adams, J. M. McCaffery, K. Teter, M. S. Kulomaa, T. E. Machen, H. P. Moore, R. Y. Tsien, *Methods Enzymol.* **2000**, *327*, 546 – 564.
[42] M. M. Wu, M. Grabe, S. Adams, R. Y. Tsien, H. P. Moore, T. E. Machen, *J. Biol. Chem.* **2001**, *276*, 33027 – 33035.
[43] G. R. Crabtree, S. L. Schreiber, *Trends Biochem. Sci.* **1996**, *21*, 418 – 422.
[44] D. M. Spencer, T. J. Wandless, S. L. Schreiber, G. R. Crabtree, *Science* **1993**, *262*, 1019 – 1024.
[45] J. D. Klemm, S. L. Schreiber, G. R. Crabtree, *Annu. Rev. Immunol.* **1998**, *16*, 569 – 592.
[46] R. Pollock, T. Clackson, *Curr. Opin. Biotechnol.* **2002**, *13*, 459 – 467.
[47] P. J. Belshaw, D. M. Spencer, G. R. Crabtree, S. L. Schreiber, *Chem. Biol.* **1996**, *3*, 731 – 738.
[48] K. Baker, C. Bleczinski, H. Lin, G. Salazar-Jimenez, D. Sengupta, S. Krane, V. W. Cornish, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 16537 – 16542.
[49] B. A. Griffin, S. R. Adams, R. Y. Tsien, *Science* **1998**, *281*, 269 – 272.
[50] G. Gaietta, T. J. Deerinck, S. R. Adams, J. Bouwer, O. Tour, D. W. Laird, G. E. Sosinsky, R. Y. Tsien, M. H. Ellisman, *Science* **2002**, *296*, 503 – 507.
[51] S. R. Adams, R. E. Campbell, L. A. Gross, B. R. Martin, G. K. Walkup, Y. Yao, J. Llopis, R. Y. Tsien, *J. Am. Chem. Soc.* **2002**, *124*, 6063 – 6076.
[52] J. Nakanishi, T. Nakajima, M. Sato, T. Ozawa, K. Tohda, Y. Umezawa, *Anal. Chem.* **2001**, *73*, 2920 – 2928.
[53] M. C. Rice, K. Czymmek, E. B. Kmiec, *Nat. Biotechnol.* **2001**, *19*, 321 – 326.
[54] K. Stroffekova, C. Proenza, K. G. Beam, *Pflugers Arch.* **2001**, *442*, 859 – 866.
[55] B. A. Griffin, S. R. Adams, J. Jones, R. Y. Tsien, *Methods Enzymol.* **2000**, *327*, 565 – 578.
[56] R. B. Silverman, *The organic chemistry of drug design and drug action*, Academic Press, Inc., San Diego, **1992**, p. 178.
[57] J. C. Powers, J. L. Asgian, O. D. Ekici, K. James, *Chem. Rev.* **2002**, *102*, 4639 – 4750.
[58] C. D. Hodneland, Y. S. Lee, D. H. Min, M. Mrksich, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 5048 – 5052.
[59] E. L. Becker, *Biochim. Biophys. Acta* **1967**, *147*, 289 – 296.
[60] F. Bjorkling, A. Dahl, S. Patkar, M. Zundel, *Bioorg. Med. Chem.* **1994**, *2*, 697 – 705.
[61] A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel, K. Johnsson, *Nat. Biotechnol.* **2003**, *21*, 86 – 89.
[62] A. E. Pegg, *Mutat. Res.* **2000**, *462*, 83 – 100.
[63] R. Damoiseaux, A. Keppler, K. Johnsson, *ChemBioChem* **2001**, *2*, 285 – 287.
[64] A. Juillerat, A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel, K. Johnsson, *Chem. Biol.* **2003**, in press.
[65] M. Kindermann, N. George, N. Johnsson, K. Johnsson, *J. Am. Chem. Soc.* **2003**, *125*, 7810 – 7811.

Received: March 25, 2003 [M603]

