DOI: 10.1002/cbic.200800190

Split-Ubiquitin and the Split-Protein Sensors: Chessman for the Endgame

Judith Müller[b] and Nils Johnsson*[a]

Introduction

The understanding of cellular biology requires a complete, quantitative, and dynamic description of the protein interactions inside the cell. Most of the known interactions were discovered very recently through the use of high-throughput techniques. Hence knowledge about the majority of these interactions hardly exceeds an awareness of their pure existence. The reasons for the difficulty in overcoming this lack of knowledge quickly are partly technical. Citing yeast as a representative example, the majority of the known physical connections between its proteins were derived from co-precipitation studies and two hybrid screens.[1,2] Both methods investigate proteins in a non-native environment, a condition that makes the straightforward integration of these data into the cellular framework difficult. As a consequence, much weight is now put on identifying and characterizing the interactions of proteins in their natural environments.[3,4]

Casually referred to as "the endgame of protein biochemistry" this endeavor is driven by the development of new and the refinement of existing technologies.[5] Split-ubiquitin (split-Ub) is the founding member of a class of analytical tools named split-protein sensors (alternatively referred to as protein fragment complementation assay, PCA) that, based on a common principle, allows measurement of protein interactions and other features of proteins in living cells.[6] Over the years, the application of this common principle to different sensor proteins gave rise to a rich spectrum of new techniques that diverge in their experimental output and their applicability to different cell types or subcellular structures.[7–18] By focusing on split-Ub we will introduce the properties of these systems and their latest applications.

General Properties

The underlying principle of all split-protein sensors is simple and based on a feature already discovered during the early work in protein chemistry: many proteins can be artificially split into two halves that retain their ability to refold into the conformation of the uncut protein.[19] However, the step from two self-assembling fragments of a protein to a split-protein sensor requires two qualities of the assembly process. First, the refolded protein must display an easily measurable property that neither of the two halves possesses on its own. Secondly, only high concentrations of the two halves should induce a significant reassembly to the native-like protein. Expressing the split-protein fragments (SPFs) as fusions to two interacting proteins will mimic a dramatic increase in cellular concentration of the SPFs forcing them to refold into the native-like protein (Figure 1 A). The induced reassembly of the SPFs thus reflects the interaction or more generally the colocalization of the two coupled proteins within one protein complex.[6]

The performance of a split-protein sensor depends on three critical parameters of the SPFs: their cellular concentrations (c_{cell}), their effective or local concentration (c_{eff}) in the complex, and their residual affinity to each other (as measured by the dissociation constant K_{dSPF}). Values of c_{cell} and c_{eff} that allow reasonable distinctions between interacting and noninteracting proteins can be estimated for each split-protein sensor by the K_{dSPF} of its SPFs. When c_{cell} is below $0.1\,K_{dSPF}$ and c_{eff} is above $10\,K_{dSPF}$, the fraction of the reassembled SPFs goes from less than 0.1 to more than 0.9 in the complex—a difference that should unambiguously separate interacting from noninteracting proteins (Figure 1 B). The c_{cell} value of proteins can be measured and are generally in the range of 10–500 nM.[20] However, no experimental data exist neither are the c_{eff} values of SPFs in a given protein complex easily available. The values of c_{eff} will depend on individual properties of the protein complex including its stability and, perhaps even more critically, how the SPFs are aligned in relation to each other. A point of reference can be extrapolated from the thermodynamic data obtained on pairs of proteins that were artificially connected by short spacer peptides.[21] In these configurations, the c_{eff} values of the connected proteins could vary between 0.01 mM and 10 M. To illustrate the significance of the K_{dSPF} value for the type of interactions that can be monitored by the corresponding SPFs, the following example is given: Assuming that the binding partners have a c_{cell} value of 500 nM and increase the c_{eff} of the coupled SPFs to 0.3 mM, all SPFs with K_{dSPF} values between 5 and 30 μM will be appropriate to monitor these interactions. Accordingly, SPFs with a K_{dSPF} of 1 mM will already leave 75% of the complex-associated SPFs unassembled and SPFs with a K_{dSPF} of 0.5 μM (a value close to the K_{dSPF} for the fragments of

[a] N. Johnsson
Institute of Molecular Genetics and Cell Biology, Ulm University
89081 Ulm (Germany)
Fax: (+49) 731-5036302
E-mail: nils.johnsson@uni-ulm.de

[b] J. Müller
Institute of Medical Biochemistry, Cellular Biochemistry
ZMBE, University of Münster
48149 Münster (Germany)

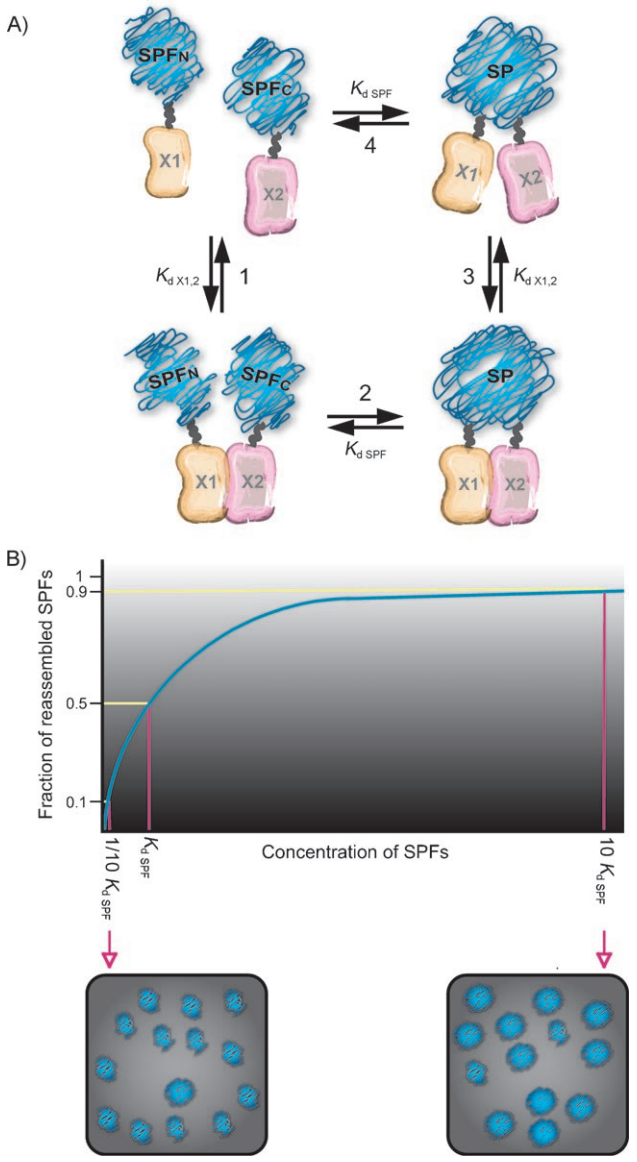

Figure 1. Split-protein sensors. A) The unfolded fragments (SPF_N and SPF_C) of a split-sensor protein (SP) are artificially linked to the proteins X1 and X2. The binding of X1 to X2 forces the SPFs into close proximity (1). Once the effective concentration of the SPFs exceeds the K_{dSPF} of their spontaneous reassembly the formation of the native-like SP occurs predominantly (2). The SP regains its enzymatic activity and signals the reassembly of SPF_N and SPF_C and thus, indirectly, the binding of X1 to X2 (Path 1–2). Path 4–3 in this scheme is negligible as long as the cellular concentrations of the SPFs are well below the K_{dSPF} of their reassembly reaction. B) The reassembly of the SPFs to their native-like split-sensor protein follows the thermodynamics of a bimolecular reaction. As a first approximation suitable SPFs should, under cellular conditions, reassemble the SP to less than 10% whereas in the X1/X2 complex more than 90% of the SPFs should form the SP. Accordingly the cellular concentrations of the SPFs should be less than $^1/_{10}$ of their K_{dSPF} and the effective concentration in the X1/X2 complex more than $10K_{dSPF}$. Split-protein sensors displaying different K_{dSPF} values will each posses their own corridor of suitable cellular and effective concentrations.

the protease barnase) will drive the unaided association of a significant 50% of the SPFs into the reassembled quasinative state.[22] Whereas the former will miss many of the protein interactions that occur, the latter will provide an intolerably high background of false positives.

This simple thermodynamic consideration applies to split sensors whose fragments are able to dissociate from each other. However, the reassociation of fragments of the green fluorescent protein (GFP) was shown to occur irreversibly, the once reassembled GFP thus serving as an interaction trap that inhibits not only the dissociation of its SPF_{GFP}s but also the disassembly of the linked proteins.[23] For these members of the split-protein sensor family a kinetic consideration of the reassembly reaction might be more appropriate. If one again assumes for the SPFs a c_{cell} value of 500 nM and a c_{eff} value of 0.3 mM, the assembly of the two SPFs when brought together in the complex will be accelerated by a factor of 360000. Although very comforting this number is not too meaningful for most split-protein sensors, as the kinetics of their reassembly cannot be easily measured. Exceptions include the split-Ub sensor where a conventional pulse chase experiment allows the formation of the SPF assembly to be followed by an irreversible change in the molecular weight of the reporter.[6] The rate of reassociation can also be determined for those interactions whose formation can be induced by external stimuli. For example, Michnick and colleagues developed an improved split-luciferase assay that made it possible to record the assembly and the disassembly of the regulatory and the catalytic subunits of protein kinase A after the addition of interaction interfering drugs. The split-luciferase emanates a luminescent signal that could be recorded within seconds after the stimuli were given.[23]

Split-protein sensors that signal their reassembly by growth/nongrowth of the transformed host cells introduce a further parameter that influences their applicability. To ensure survival they have to provide the cell with a certain minimal amount of the reassembled split-protein sensor. Only above this concentration can the growth of the cells be supported by the reconstituted split-sensor protein and the interactions measured. This concentration thus sets a further limitation on the choice of a suitable split-protein sensor. For example, a threshold concentration for the survival of the cells below $0.1\,K_{dSPF}$ of the SPF assembly reaction bears the potential for a high background of false positives and thus the value should be significantly higher than this.

The cellular space is structured into microcompartments and many proteins are organized at membranes. If both SPFs are membrane associated their movements are restricted to a common surface and their orientation to each other is already favorable for the reassembly. Thus the effective concentration of proteins in these compartments is on average probably higher than the 10–500 nM estimated for the cytosolic proteins.[24] Therefore measuring interactions of membrane proteins or freely diffusing, cytosolic proteins most likely require split-protein sensors displaying different K_{dSPF} values.[24]

Split-ubiquitin (split-Ub)

Ubiquitin is genetically divided between amino acids 34 and 35 of the 76 residue protein making ubiquitin not only the first but also the smallest split-protein sensor (split-Ub).[6] The association constant of the reversible binding of the N-terminal

SPF (N_{ub}) to the C-terminal SPF (C_{ub}) is 1.4×10^5 at pH 7.0.[25] With an estimated cellular K_{dSPF} value of 7 μM, and according to the consideration given above, split-Ub should be able to monitor the interaction of pairs of proteins whose cellular concentration is below 700 nM and which, through forming a complex with each other, increase the local concentration of the coupled SPF_{ub} to more than 0.07 mM. The molecular readout for the reassembly of the SPF_{ub} into the quasinative Ub is the recognition and cleavage by the ubiquitin specific protease (UBPs).[6] UBPs cleave any attached reporter protein from the C terminus of the folded ubiquitin but leave the singular C_{ub}-fusions intact (Figure 2).[26]

Early experiments with SPF_{ub} attached to membrane proteins of the endoplasmic reticulum revealed that the K_{dSPF} value of the wild-type ubiquitin is too low to be used as a sensor for this type of protein interaction.[24] Proteins known not to interact but to only colocalize in this compartment already gave a strong interaction signal. Obviously the effective concentration of the expressed fusion proteins was greater than the still tolerated 700 nM. To be able to adjust the split-Ub system to the inherently higher local concentrations of proteins in membranes and other subcompartments of the cell, single and double amino acid mutations were introduced into N_{ub}. By replacing the Ile13 of N_{ub} with an Ala (N_{ua}) a N_{ub}/C_{ub} combination was created that was first used to successfully map interactions among members of the ER translocation machinery of the yeast *Saccharomyces cerevisiae*.[24] Although the affinity of N_{ua} to C_{ub} was not experimentally tested, the type of replacement allows a rough estimate of its effect on the affinity to C_{ub}.[27] Reducing the Ile side chain by three methyl groups should create a cavity in the folded ubiquitin that destabilizes the structure by 2.5–5 kcal mol^{-1}. As the reassembly of the SPF_{ub} leads to a very native-like Ub structure, the stability of the SPF_{ub}-heterodimer should be affected by this mutation to a similar degree.[25] A decrease in 3 kcal mol^{-1} increases the K_{dSPF} value of the corresponding SPF_{ub} to approximately 1 mM. Accordingly, N_{ua} can still sense the contact between proteins whose c_{cell} values are around 100 μM. The strategy of introducing cavity-creating mutations into the N_{ub} was continued to create 16 further N_{ub} mutants with different affinities to C_{ub}. It is likely that this collection contains a suitable N_{ub} variant to study almost any type of protein interaction encountered in the cell.[28]

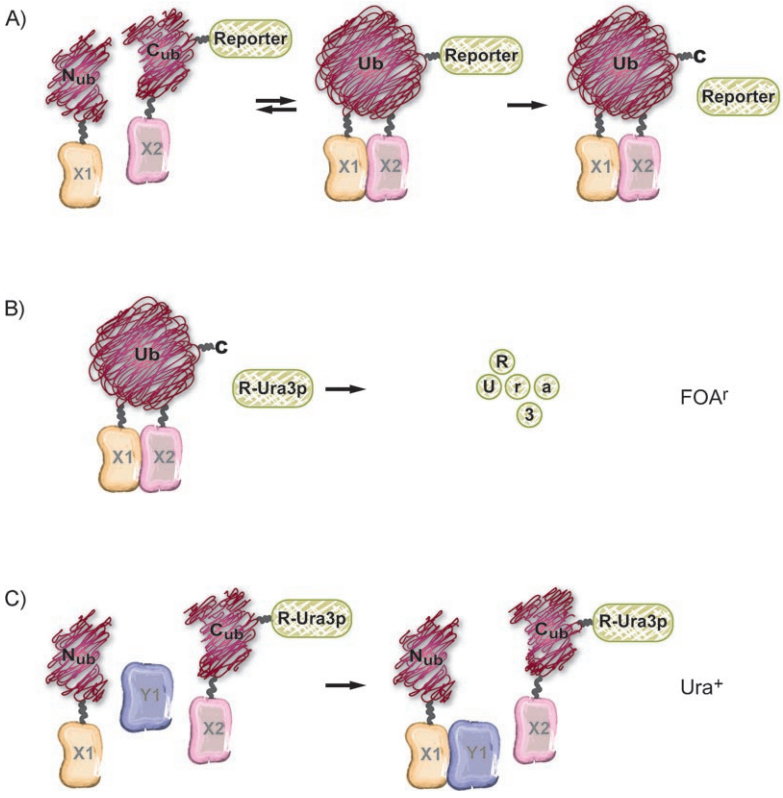

Figure 2. Split-ubiquitin. A) The SPFs of the split-ubiquitin sensor (N_{ub} and C_{ub}) upon reassembly reconstitute a native-like ubiquitin (Ub) that is recognized by the ubiquitin-specific proteases. A reporter protein attached to the C terminus of C_{ub} is cleaved. Depending on the nature of the reporter, its release can be visualized by different methodologies. B) The most universal reporter for the split-ubiquitin assay is R-Ura^{3}p. Cleavage from C_{ub} exposes the N-terminal arginine (R) that according to the N-end rule initiates the destruction of Ura^{3}p. Cells that express a pair of interacting N_{ub}- and C_{ub} fusion proteins are uracil auxotroph and 5-FOA resistant. The presence of the interaction is recorded by the growth of the cells on 5-FOA containing medium. C) The same R-Ura^{3}p reporter can signal an acquired absence of a certain protein interaction. Expression of Y1 competes with X2 for the binding to X1. Consequently less R-Ura^{3}p is cleaved from X2-C_{ub}-R-Ura^{3}p and the cells will grow on medium lacking uracil.

Monitoring the Interaction

The split-Ub system offers more than one readout option. Any protein that changes a measurable quality after its release from C_{ub} can be used as a reporter for the system. Three basic properties of a reporter are currently exploited; its subcellular localization, stability, or molecular weight (Figure 2).

Molecular weight

The most direct readout for measuring the interaction between a protein X1 coupled to N_{ub} and a protein X2 coupled to C_{ub} is the change in molecular weight of the X2-C_{ub}-R fusion. This change occurs upon reassembly with N_{ub}-X1 through proteolysis at the C terminus of C_{ub} and is accompanied by the appearance of the cleaved R.[6] R should be a small monomeric non-host protein. Antibodies directed against R will monitor the N_{ub}-X1/X2-C_{ub}-R reassociation after cell extraction, denaturing electrophoresis, and immunoblotting. Simultaneous detection of the cleaved and uncleaved fraction of the C_{ub} fusion allows quantification of the extent of N_{ub}/C_{ub} reassociation.[24,28] The sensitivity and the temporal resolution of the technique can be improved by pulse chase experiments. Measuring the kinetics of the reporter cleavage can unravel subtle changes in the association rate of protein X1 and X2 induced by cellular signals or other changes in the environment of the proteins. One minute is the highest temporal resolution that can be routinely achieved by pulse chase experiments with the split-Ub system.[6]

Subcellular localization

To use subcellular localization as a readout requires that the change in cellular localization of the reporter upon cleavage from the C_{ub} is coupled to a robust signal. An artificial transcription factor (PLV, a fusion of Protein A, LexA, and VP16) as a split-Ub reporter provided one attractive solution. Upon reassociation of N_{ub} and C_{ub}, PLV is cleaved off and diffuses into the nucleus to activate transcription from promoters containing the matching lexA sites. In manipulated yeast strains expressing *HIS3* under control of PLV, interactions between N_{ub} and C_{ub} coupled fusion proteins will enable the yeast cells to grow on medium lacking histidine.[29] As most membrane proteins are excluded from the nucleus this readout is especially suited for X2-C_{ub}-PLV fusions where X2 is a constitutive membrane protein. To apply this variation of the split-Ub system to cytosolic proteins the C_{ub} modified partner has to be artificially linked to the membrane—at the cost of depriving it of its natural habitat and artificially raising its effective concentration.[30]

The transcription of the His3p enzyme is a strong amplifier of the primary signal, the cleavage of PLV from the membrane attached C_{ub}. To suppress signals arising from random encounters between the N_{ub}- and C_{ub}-coupled fusion proteins in the same membrane, the N_{ug} mutant is exclusively used in this configuration of the assay.[31–35] N_{ug} carries an Ile/Gly replacement in position 13 of N_{ub}.[6] Although applicable to all eukaryotic cells, the PLV reporter was so far only been used in yeast cells.

A different reporter configuration was devised to measure the interaction between transporter and substrate during protein translocation across the membrane of the endoplasmic reticulum (ER) (Figure 3). Here X2 in the X2-C_{ub}-R construct is a protein that guides the fusion across the membrane of the ER whereas R is the enzyme Ura^{3}p. Ura^{3}p is instrumental in the synthesis of uracil and its presence is required in the cytosol. As an integral part of the X2-C_{ub}-Ura^{3}p construct, however, Ura^{3}p translocates into the lumen of the ER rendering the yeast cells phenotypically Ura$^-$. A N_{ub}-fusion that binds to X2-C_{ub}-Ura^{3}p before its passage through the membrane is completed will initiate the cleavage of Ura^{3}p. Consequently the cleaved Ura^{3}p accumulates in the cytosol and the cell becomes uracil prototrophic.[36] Although applied only for protein translocation across the membrane of the ER, the Ura^{3}p reporter should be equally suitable for the study of other protein import systems.

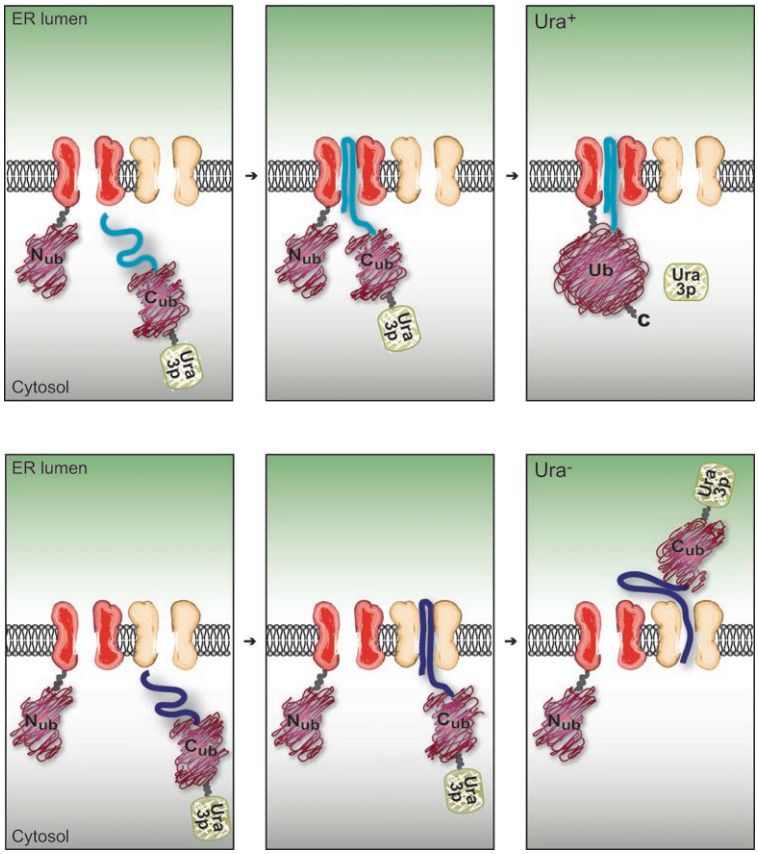

Figure 3. Counting the flow of translocation substrates through the pores of the endoplasmic reticulum (ER). A) N_{ub} is linked to the cytosolic face of a translocation pore (red) whereas C_{ub}-Ura^{3}p is coupled behind the signal sequence of an ER translocation substrate (blue). The short-lived binding of the signal sequence to the red translocation pore increases the effective concentration between N_{ub} and C_{ub}. Ura^{3}p is released into the cytosol after N_{ub}/C_{ub} reassembly has occurred. The resulting accumulation of the Ura^{3}p activity in the cytosol enables the growth of the cells on medium lacking uracil. B) C_{ub}-Ura^{3}p is coupled behind a different signal sequence (dark blue) that translocates via the yellow translocation pore across the membrane of the ER. As this pore is not modified by the addition of N_{ub}, the C_{ub}-R-Ura^{3}p moiety escapes unhindered into the lumen of the ER. The cells are phenotypically ura$^-$. Repeating the experiment with cells harboring a N_{ub} modified yellow translocation pore proves that the two signal sequences pass via different pores across the ER membrane.

Degradation

The third reporter configuration makes use of the newly exposed amino terminus of the reporter after its release from C_{ub}. According to the N-end rule pathway of protein degradation, the identity of this residue determines the half-life of the reporter protein. To construct a universal reporter for detecting interactions of membrane bound and cytosolic proteins as well as transcription factors and DNA modifying enzymes, an arginine residue followed by a short peptide including acceptor lysines for ubiquitylation was inserted at the C_{ub}-Ura^{3}p junction. After cleavage from C_{ub} the arginine (R) becomes the N-terminal amino acid of Ura^{3}p (R-Ura^{3}p) and, as a destabilizing residue according to the N-end-rule, will induce the rapid destruction of R-Ura^{3}p.[24] The R-Ura^{3}p reporter harbors the option to positively select for the presence or the absence of an interaction (Figure 2C). In cells expressing an interacting pair of N_{ub}/C_{ub} fusions, the R-Ura^{3}p moiety is rapidly degraded resulting in uracil auxotrophy. The switch from uracil auxotrophy to uracil prototrophy was used to screen for fragments of the yeast membrane protein Sec^{62}p that interfere with the interaction between its full-length version and its binding partner in the membrane of the ER, Sec^{63}p.[38] A positive selection for interact-

ing proteins can be achieved by introducing 5-fluoro-orotic acid (5-FOA) into the media. 5-FOA is converted to the toxic compound 5-fluoro-uracil (5-FU) by Ura^3p. As the interaction of the N_{ub}/C_{ub} coupled proteins leads to the rapid degradation of the R-Ura^3p, growth of the transformed yeast on 5-FOA is indicative of cells harboring interacting proteins.[24,37]

Different reporter proteins can be used in place of Ura^3p in this assay. In one example R-GFP was used to measure protein interactions by the disappearance of a fluorescent X2-C_{ub}-R-GFP fusion through coexpression of an interacting N_{ub}-fusion.[37]

The choice between numerous reporters not only provides alternative modes of observation but also permits measurement of different aspects of the reassembly reaction and adjustment of the sensitivity and kinetics of the assay to the requirements of the interaction under observation. For example, the transcriptional activation of a growth marker strongly amplifies the primary signal. In contrast it is the uncleaved X2-C_{ub}-R-Ura^3p that poisons the cell by converting 5-FOA to 5-FU, and a rather complete conversion of all C_{ub}-R-Ura^3p fusion proteins into the cut- and subsequently degraded R-Ura^3p moiety is required to render the cell unsusceptible to 5-FOA. This property of the R-Ura^3p reporter reduces the noise of the system that can be caused by the spontaneous reassembly of the SPF_{Ub}. The same property will also eliminate false interaction signals that might arise by the irregular cleavage in the sequence of the C_{ub} by proteases that recognize stretches of unfolded polypeptides. Here the reporter will be released independently of whether or not the N_{ub}/C_{ub} reassociation has occurred. The liberated, and now stable, R-Ura^3p will then mask otherwise occurring interactions by keeping the cell 5-FOA sensitive. Provided that this is a rare event, the amount of R-Ura^3p will not suffice to poison the cells, and the extra cleavage in the sequence of the C_{ub} goes unnoticed. Both outcomes are preferable to a similar release of a transcription factor as a reporter for the split-Ub system. Here the inherent amplification might transfer the irregular cleavage more easily into a false positive interaction signal.

The feature of multiple readouts is shared by the split-TEV and split-intein systems.[8,39] In the split-TEV system the interaction-induced reconstitution of the TEV protease will cleave any reporter protein containing a TEV recognition site. In the split-intein system the reconstituted intein will splice its N- and C-terminally attached peptides into a new protein. As the N- and C-terminal extensions can be freely chosen, nearly any suitable activity can be reconstituted.

Applications

Besides monitoring a particular protein interaction, the split-Ub system can also be used to analyze other aspects of protein function in the cell. By concentrating on these applications, this review aims to highlight the versatility of this and other split proteins systems.

Counting the Passage of Signal Sequences through the Different Pores in the Membrane of the ER

The proteins of the secretory system are synthesized in the cytosol and subsequently transferred across the membrane of the endoplasmic reticulum (ER) into the lumen of this compartment. For this purpose the enzymes of the targeting and translocation machinery bind to short continuous stretches at the N terminus of the protein substrates, the so-called signal sequences. As these interactions are very short lived and membrane bound, their detection poses a special experimental challenge. By linking N_{ub} to the cytosolic N terminus of the transporter and C_{ub} to a signal sequence, the interaction between transporter and substrate could be measured for the first time in vivo.[36] As a result of the transient nature of the interaction, wild-type N_{ub} providing the highest known affinity to C_{ub} had to be used.[36] The interactions were reported either by a molecular weight readout or by an increase in cytosolic Ura^3p activity (Figure 3).

For not fully understood reasons the individual signal sequences that guide proteins into the lumen of the ER show differences in length, in the degree of hydrophobicity, in the distance to the N terminus of the protein, and in the charge distribution at their N- and C-terminal ends. In addition, yeast harbors three channels of different molecular composition to transfer signal sequence-bearing proteins across the membrane of the ER. Genetic experiments and the in vitro reconstitution of protein translocation allowed prediction of the preferred channel for certain signal sequences.[40] To validate these predictions and to measure the individual route of a signal sequence under in vivo conditions, N_{ub} was coupled to components of the different tranlocation channels and C_{ub} was attached to the C termini of diverse signal sequences. This configuration permitted experimental determination of the specificity of the three channel systems towards individual signal sequences for the first time in vivo.[41] The derived interaction profiles of the different signal sequences showed a clear preference of hydrophobic signal sequences for the trimeric Sec61p channel and the alternative Ssh1p channel whereas the less hydrophobic signal sequences translocate primarily via the heptameric Sec-complex.[41]

Monitoring transient interaction during protein maturation is not an exclusive property of split-Ub. The short lived interaction between a chaperone and its cargo was captured and visualized by split-GFP in the ER- and Golgi compartment of mammalian cells.[42] In addition, the transport of proteins from the cytosol into the nucleus of cells could be monitored by split-intein using renilla luciferase as the reporter.[43]

Altered Protein Conformations

The spatial positions of the SPFs in a given protein complex will significantly influence the efficiency of their reassociation. SPFs that reside on opposite sides of a protein complex will assemble less efficiently than those that are attached on the same side. This steric sensitivity of all split-protein sensors per-

mits measurements of conformational effects or subunit arrangements in protein complexes and also permits probing conformational alterations in a single protein.[28,44–46]

Linking N_{ub} to the N and C_{ub} to the C terminus of a protein renders the spatial distance between the natural N and the C terminus, together with the rigidity of the structure, the dominant parameters determining the rate of the N_{ub}/C_{ub} reassociation of the fusion protein. Changes in the conformation of the protein that will alter this distance will also alter the rate of N_{ub}/C_{ub} reassembly (Figure 4).[28] Once coupled to a sensitive reporter the change in the intermolecular distance between the N- and C-terminal end can be used to measure conformational alterations in vivo. To compensate for the high effective concentrations of the SPF_{ub} that are connected covalently by a polypeptide, mutants of N_{ub} with strongly reduced affinities to C_{ub} are generally employed for these studies. Such mutants were obtained by simultaneously reducing the side chains of the Ile in position 3 and 13 of the N_{ub}.[24]

Quantification of the cleaved and uncleaved reporter makes the molecular weight of the reporter construct the most reliable readout to measure these conformational alterations with the split-Ub system. Examples of its use include the detection of conformational change of the gamma-subunit of the trimeric G protein after binding to the β-subunit, the detection of a conformational alteration in an allele of the yeast protein $Sec^{62}p$, and the classification of mutations in the DNA binding domain of the tumor suppressor p53.[28,47,48]

Although the reassociation between N_{ub} and C_{ub} is reversible, the irreversible cut at the C terminus of C_{ub} makes it impossible to continuously monitor fluctuations in the structure of the investigated protein. Only by using the molecular weight readout and consecutive pulse chase experiments can conformational alterations be studied over time; however, with the caveat that different cells are sampled at each time point. In contrast the reversibility of the fragment association of the split-luciferase from *Gaussia* allows the "online" monitoring of protein interactions. The same important feature might also surface in other members of the split-protein family.[10] As some of them were already applied to measure conformations of proteins, the continuous recording of the structural state of a protein may become possible as well.[44,49]

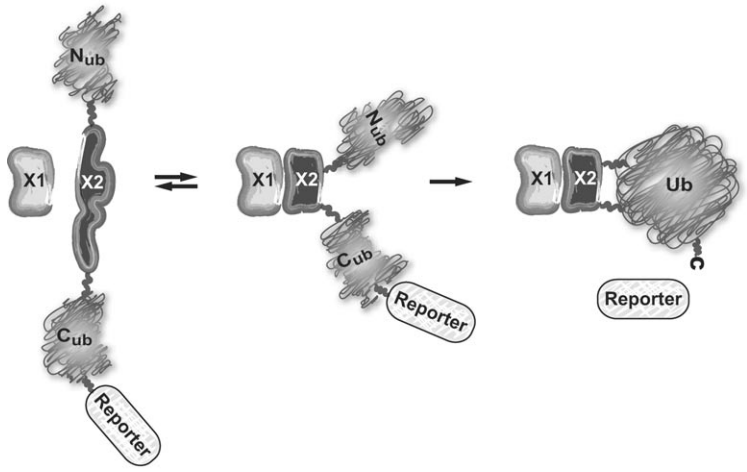

Figure 4. Split-ubiquitin as a sensor of protein conformations. N_{ub} and C_{ub}-reporter are attached to the N and C terminus of a protein X2. Upon binding to X1 the conformation of X2 is dramatically altered, shortening the distance between N_{ub} and C_{ub}. The enhanced cleavage of the reporter R reflects the conformational alteration of X2.

Quantification

The unique architectures of protein complexes preclude a strict correlation between the binding strength of the investigated complex and the intensity of the signal generated by split-Ub or other split-protein sensors. Straightforward interpretations of the quantitative differences in the split-protein assays seem therefore justified only for the analysis of structurally very similar complexes.[50–53] Even a direct comparison between the full-length protein and its fragments or domains with respect to binding a common partner is not trivial. It requires that the distances between the SPFs are very similar in the different complexes. This is probably not true in most instances. To measure the relative strength of fragments of the yeast protein $Sec^{62}p$ to its binding partner $Sec^{63}p$ in the membrane of the ER, a split-Ub based competition assay was devised.[38] Here N_{ub} was fused to the N terminus of $Sec^{62}p$ and the C_{ub}-R-Ura^3p module to the C terminus of $Sec^{63}p$. Fragments or mutants of the unmodified $Se^{62}p$ were then overexpressed in the same cell and their effects on $Sec^{63}p$ induced N_{ub}/C_{ub} reassociations were measured by the improved growth of yeast cells on medium lacking uracil (Figure 2C). As the N_{ub} and C_{ub} were attached to the full-length proteins and not to their competing fragments the distance between N_{ub} and C_{ub} remained constant, and thus the number of surviving cells permitted quantification of the relative binding strengths of the different $Sec^{62}p$ mutants to $Sec^{63}p$. After a variety of fragments had been tested, two stretches of $Sec^{62}p$ each contributing to the binding strength of the full-length Sec^{62} were identified and semiquantitatively compared.[38]

In another example, the conformations of mutants containing residue exchanges in the DNA binding domain of p53 were analyzed with the split-Ub assay in vivo. Thermodynamic in vitro measurements of the mutants' effect on the stability of the protein were compared to the signal of the split-Ub assay. It could be shown that the extent of the intramolecular N_{ub}-p53-C_{ub} reassembly depended on the type of p53 mutation and indeed roughly correlated with the mutations' quantitative effect on the thermal stability of the protein.[48]

Detecting the Binding of Proteins to Nonprotein Targets

Originally implemented to measure protein–protein interactions recent variations on the split-protein sensors shifted the focus towards the interaction of proteins with their nonprotein targets. Small synthetic molecules play an increasing role in manipulating the function of proteins in living cells. Identifying the

protein targets of compounds that interfere with certain aspects of cell physiology is laborious and would still benefit from new approaches. Besides the classical techniques based on affinity chromatography, small-molecule three-hybrid systems are increasingly used for these purposes.[54] The latter approach makes use of the transcription factor-based yeast two-hybrid system where the DNA binding domain and the transcription activating domain of a transcriptional activator are separately linked to a protein with a known affinity to a small compound X and the potential target protein. The compound of interest Y is then supplied as part of a synthetic hybrid that additionally contains compound X. The hybrid thus serves as a chemical inducer of the dimerization (CID) of the artificial transcription activator domain and the DNA binding domain. The reconstitution of the functional transcriptional activator leads to the synthesis of a reporter gene.

Dirnberger et al. reported on a variation of this scheme using split-Ub as the reporter of the CID event.[55] In a proof of principle study, N_{ub} was fused to dihydrofolate reductase (DHFR) and the C_{ub}-R-Ura^{3}p module was coupled to the rat glucocorticoid receptor (GR). Accordingly, the hybrid CIP consisted of methotrexate as the ligand for DHFR coupled to dexamethasone as the ligand for GR. Supplying the growth media with this CID allowed the yeast cells co-expressing the corresponding split-Ub fusion proteins to grow on media containing 5-FOA as a positive indicator of the binding of dexamethasone to GR.[55] The advantage of the split-Ub system over the two-hybrid approach is its flexibility with regard to the type of proteins that can be tested against small-molecule interactions including membrane-associated receptors or channels. This expanded applicability is of considerable interest as many of the potential targets for synthetic compound screens are residents of this compartment. Other split-protein sensors were also shown to be suitable for similar purposes.[16,56]

The interactions of two DNA binding proteins with their target DNA were monitored by the split-GFP system. Here, two zinc-finger DNA-binding proteins were each linked to the N- and C-terminal SPF_{GFP} respectively. Fluorescence of the reassembled GFP was only detected when DNA that contained the zinc finger recognition sites located in close proximity and with an optimal spacing to each other was provided in the reaction.[57] Although the initial studies were exclusively performed in vitro, possible in vivo applications include the documentation of genomic DNA rearrangements and the shielding of DNA by chromatin components or methylation.

Synthetic Biology

To build devices that equip the cell with new functions or serve as detectors for certain intracellular processes, synthetic biology is in demand of robust building blocks that can be used to assemble molecular "machines" or circuits in the living cell. Although not intended at the time of its invention, the split-Ub system can serve as such a module. Muir and his colleagues assembled a gadget to rapidly induce or eliminate the presence of a protein in cultured mammalian cells by the external application of a small molecule compound.[58] Their concept makes use of a short-lived protein or polypeptide stretch (Deg) that is fused to the N terminus of the FK506 binding domain of Tor1 (FRB), followed by C_{ub}, and finally the protein of interest Z1 (Deg-FRB-C_{ub}-Z1; Figure 5). As a result of the N-terminally positioned degron, the disappearance of the fusion protein is rapid, rendering the cell devoid of Z1. A fusion of the maltose binding protein (MBP) to a N_{ub} mutant of lower affinity to C_{ub} (N_{ua}) and the FK506 binding protein resulting in MBP-N_{ua}-FKB is coexpressed with Deg-FRB-C_{ub}-Z1. In this case, the amount of Deg-FRB-C_{ub}-Z1 remains low as the affinity of N_{ua} to C_{ub} is too low to capture a significant amount of C_{ub} before degradation. This situation is changed upon addition of rapamycin: Rapamycin binds to FKB, which in turn binds to FRP. This brings N_{ua} in close proximity to C_{ub} and Z1 is cleaved from the reassembled Ub. Z1 is thereby decoupled from the degron and its half-life and other biochemical characteristics are solely determined by its native properties. Thus, the appeal of this method lies in the rapid generation of the native, unmodified Z1 through the addition of a small molecule compound. The amount of Z1 can be sensitively adjusted by varying the concentration of rapamycin or ascomycin, an antagonist of rapamycin.[58]

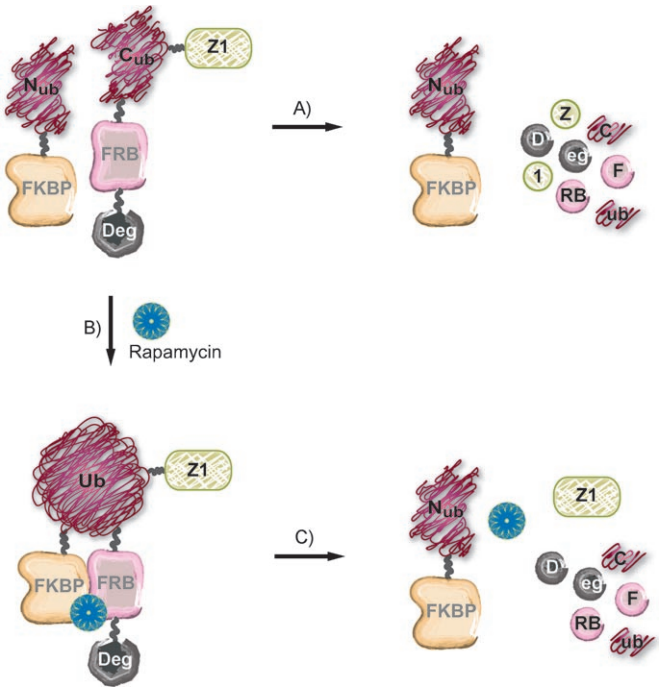

Figure 5. The SURF technology (split-ubiquitin rescue of protein function) as an example for split-Ub based synthetic biology. N_{ub}-FKBP was coexpressed with a FRB-C_{ub}-Z1 fusion protein that at its N terminus bears a degradation signal (Deg). Z1 can be any protein. A) The degron initiates the rapid degradation of the C_{ub} fusion in the cell. B) Upon addition of rapamycin a ternary complex of FKBP, FRP, and rapamycin is formed that brings N_{ub} and C_{ub} in close proximity. After N_{ub}/C_{ub} reassembly has occurred Z1 is cleaved from the C_{ub}. As a consequence, the administration of rapamycin induces the rapid appearance of Z1 in the cytosol of the cells.

Other split-protein sensors bear a similar potential for novel and unexpected applications. Split-intein is based on the inter-

action-induced reassembly of the N- and C-terminal halves of intein. The N-terminal extension of the N-terminal intein fragment and the C-terminal extension of the C-terminal intein fragment are spliced into a novel protein and consequently liberated form the reassociated intein enzyme. This reaction was exploited to fuse otherwise inactive fragments of GFP, luciferase, or a transcriptional activator to generate either autofluorescence, luminescence, or gene expression as powerful signals to follow protein interactions in cells of living animals.[11,12,59] Similar to the spit-Ub system the intrinsic flexibility of the reconstituted activity permits creation of new pathways in the cell.[60] In one example a small molecule CID drove the reassociation of the separated intein halves. The ensuing ligation freed a kinase from its inhibitory peptide.[61] Comparable to split-Ub and split-intein, split-TEV, a system based on the interaction-induced reassembly of the TEV protease, offers the option of freely selecting the reporter activity and thus harbors a similar potential yet to be exploited by synthetic biologists.[8]

A glance into the future of possible and much more sophisticated applications of split-protein sensors was provided by A. Varshavsky in a recent paper.[62] This theoretical work suggests the design of two fusion constructs each consisting of a linear array of three complementary SPFs and a sequence-specific DNA binding protein. Once the two DNA binding proteins happen to find a stretch of DNA that harbors both recognition sites in close proximity, the three split proteins refold into their native-like states. The reassembly of split-Ub and a split-Ub-like protein (two of the three complementary fragments) leads to the release of the terminally positioned split-restriction enzyme (the third split sensor) from the DNA-bound protein complex (Figure 6). As a consequence the activated restriction enzyme is then able to cleave a rare restriction site(s) engineered into the DNA molecule to be eliminated. In the context of the paper, the design of two such fusion modules is proposed to distinguish between cells that do or do not contain two homozygous DNA deletions. Cells containing the deletion are destroyed whereas the cells still harboring the respective DNA are spared. The final aim of this deletion-specific targeting strategy is the killing of cancer cells that acquired two homozygous deletions in their genome.[62] More generally, the use of different split-protein sensors in a linear fusion makes a combinatorial design of new nodes in regulatory networks plausible.

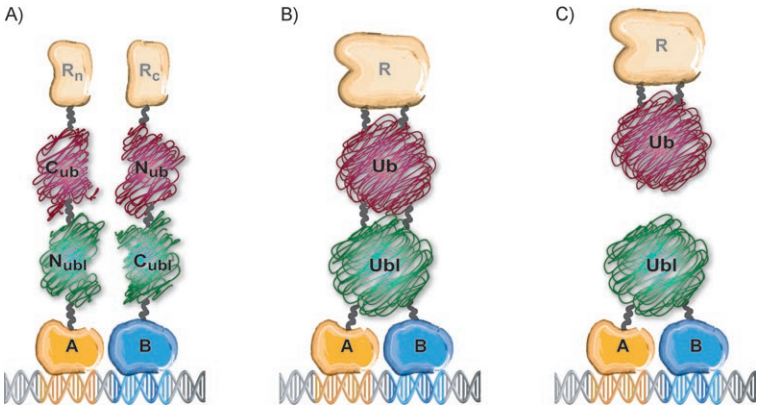

Figure 6. Targeting the absence. Two arrays of three covalently linked SPFs are expressed in the cell from a single plasmid. Both bind to two closely spaced DNA sequences via the attached DNA-binding proteins A and B. The three complementary SPFs are forced into close proximity and assemble into a linear array of the three corresponding split proteins. The first sensor is an ubiquitin-like protein (green, Ubl), the second sensor is ubiquitin (red, Ub), and the third a DNA restriction enzyme (ochre, R). The proteases that are specific for the reassembled Ub or Ubl cleave at the C terminus of C_{ub} and C_{ubl} respectively to release the reassembled R. R can leave the DNA to destroy the plasmid that contains restriction sites specific for R and that are responsible for the expression of the SPF fusion proteins as well as a toxin. Consequently cells whose genomes contain these Z1 and Z2 binding sites are spared from the action of the toxin whereas cells missing those sites will neither reassemble nor release the restriction enzyme and are therefore killed by the plasmid expressed toxin.

Systematic Interaction Studies and N_{ub}-Library Screening

With a few exceptions, the split-Ub system is predominantly used in yeast cells.[58,63] Reporter readouts that allow identification of interacting proteins by the growth of the transformed yeast cells do not only enable library screens but also the systematic and simultaneous testing of a great number of potential interaction partners. As yeast can be easily transformed with high efficiency, co-transformation of the N_{ub} and C_{ub}- carrying expression plasmids is one option. Alternatively, mating of yeast strains each carrying either a N_{ub}- or a C_{ub} fusion construct will result in diploid a/alpha cells that coexpress the two fusions. As the mating of many different yeast cells can be performed in parallel and even semiautomatically this is the preferred option for systematically testing a C_{ub} fusion protein against an array of different N_{ub}-fusion proteins. In one recent example Fields and colleagues investigated approximately 700 yeast membrane proteins for interactions among each other.[64] The readout was based on the localization of the artificial transcription factor PLV. To suppress the unspecific generation of the cleaved transcription factor the N_{ug} mutant of N_{ub} was fused to the C terminus of the proteins of interest. A machine learning algorithm was applied to help in separating true positives from potential false positives. The study thus complemented the genome-wide two hybrid screen performed in part by the same laboratory to compensate for the weakness of this method in measuring the interaction of membrane-bound proteins.[1] As is true for any genome-wide interaction study, only the number and the quality of the follow up studies that are initiated by this work can document the value of the reported interactions.

More focused split-Ub based interaction studies were performed with smaller sets of proteins, including the enzymes of the ergosterol synthesis pathway, membrane proteins of the peroxisome, channel proteins of *Arabidopsis thaliana*, the proteins of the endosome in yeast, or nuclear proteins of yeast involved in transcription.[65–69]

Transforming yeast cells expressing an X2-C_{ub}-reporter with a genomic or cDNA library harboring DNA fragments that were

randomly fused to N_{ub} allows identification of novel interaction partners of X2. This approach still has the advantage that completely unexpected interactions can be identified and even complex genomes can be screened for which systematic approaches are not feasible. Furthermore, libraries very often contain fragments of proteins that differ in their interaction behaviors from their native counterparts. These interactions will be missed in the systematic approaches that are generally based on the analysis of full-length proteins. The two reporters of the split-Ub system that enable a selection for the presence of interacting pairs of N_{ub} and C_{ub} fusions, R-Ura^{3}p and PLV, are now increasingly used for these purposes. Protocols for choosing the optimal conditions and adjustments that improve their robustness were recently published.[70,71] The classes of proteins that were successfully screened for new interaction partners with the help of the split-Ub system include transcription factors, small molecule transporters, ER-resident membrane proteins, protein kinases, and receptors of the serpentine and Erb family.[31,33,35,37,70,72]

Outlook

As the demands for sensitivity, reliability, and speed are constantly growing, the development of new methods for detecting, quantifying, and manipulating protein interactions in living cells can be expected to continue.[4] Among the many methods now available split-protein sensors have proved to be a fruitful concept. Two lines of progress are foreseeable: 1) Introducing reporter proteins with better or novel features in systems such as split-Ub, split-intein, or split-TEV and 2) The generation of completely new split-protein sensors from proteins that add novel or improved qualities to the already existing set of SPFs. To quantify the effective concentrations between two proteins in the cell, it is desirable to have more split-protein sensors whose $K_{d\,SPF}$ values are known and that, like split-Ub or split-intein, allow differentiation between the free and the assembled SPFs.

Fulfilling these requests is getting less demanding as the design of a new split-protein sensor no longer requires intuition or structural knowledge about the protein. Recently the initial trial-and-error strategy has increasingly been replaced by the systematic or random generation of pairs of fragments of a protein and the testing for their interaction induced reassociation.[10,16,73] Promising candidate proteins for split-protein sensors can also be subjected to an evolutionary strategy that will eventually yield fragments with the desired properties.[14] The many examples of new split-protein sensors that consequently emerged from these studies seem to predict that almost any protein with novel and interesting properties can indeed be converted into a novel split-protein sensor.

Acknowledgement

We would like to thank former and current lab. members for their contributions to the split-ubiquitin system and Kai Johnsson for his input and comments on the manuscript. Research in the lab. was funded by the Human Frontier Grant RGP0045/2004 and by the Deutsche Forschungsgemeinschaft A14SFB629.

Keywords: protein conformation · protein engineering · protein interactions · protein translocation · synthetic biology

[1] P. Uetz, *Nature* **2000**, *403*, 623–627 .
[2] A. C. Gavin, *Nature* **2002**, *415*, 141–147.
[3] J. Piehler, *Curr. Opin. Struct. Biol.* **2005**, *15*, 4–14.
[4] N. Johnsson, K. Johnsson, *ChemBioChem* **2003**, *4*, 803–810.
[5] A. R. Mendelsohn, R. Brent, *Science* **1999**, *284*, 1948–1950.
[6] N. Johnsson, A. Varshavsky, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 10340–10344.
[7] J. N. Pelletier, F. X. Campbell-Valois, S. W. Michnick, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12141–12146.
[8] M. C. Wehr, *Nat. Methods* **2006**, *3*, 985–993.
[9] T. Ozawa, A. Kaihara, M. Sato, K. Tachihara, Y. Umezawa, *Anal. Chem.* **2001**, *73*, 2516–2521.
[10] I. Remy, S. W. Michnick, *Nat. Methods* **2006**, *3*, 977–979.
[11] A. Kanno, T. Ozawa, Y. Umezawa, *Anal. Chem.* **2006**, *78*, 556–560.
[12] T. Ozawa, T. M. Takeuchi, A. Kaihara, M. Sato, Y. Umezawa, *Anal. Chem.* **2001**, *73*, 5866–5874.
[13] R. Paulmurugan, S. S. Gambhir, *Anal. Chem.* **2003**, *75*, 1584–1589.
[14] P. Tafelmeyer, N. Johnsson, K. Johnsson, *Chem. Biol.* **2004**, *11*, 681–689.
[15] F. Rossi, C. A. Charlton, H. M. Blau, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 8405–8410.
[16] A. Galarneau, M. Primeau, L. E. Trudeau, S. W. Michnick, *Nat. Biotechnol.* **2002**, *20*, 619–622.
[17] K. E. Luker, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 12288–12293.
[18] S. B. Kim, Y. Otani, Y. Umezawa, H. Tao, *Anal. Chem.* **2007**, *79*, 4820–4826.
[19] F. M. Richards, P. J. Vithayathil, *Brookhaven Symp. Biol.* **1960**, *13*, 115–134.
[20] S. Ghaemmaghami, *Nature* **2003**, *425*, 737–741.
[21] C. R. Robinson, R. T. Sauer, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5929–5934.
[22] J. Sancho, A. R. Fersht, *J. Mol. Biol.* **1992**, *224*, 741–747.
[23] E. Stefan, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 16916–16921.
[24] S. Wittke, N. Lewke, S. Müller, N. Johnsson, *Mol. Biol. Cell* **1999**, *10*, 2519–2530.
[25] M. Jourdan, M. S. Searle, *Biochemistry* **2000**, *39*, 12355–12364.
[26] N. Johnsson, A. Varshavsky, *Embo J.* **1994**, *13*, 2686–2698.
[27] A. E. Eriksson, *Science* **1992**, *255*, 178–183.
[28] X. Raquet, J. H. Eckert, S. Muller, N. Johnsson, *J. Mol. Biol.* **2001**, *305*, 927–938.
[29] I. Stagljar, C. Korostensky, N. Johnsson, S. te Heesen, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5187–5192.
[30] N. Möckli, *Biotechniques* **2007**, *42*, 725–730.
[31] B. Wang, *J. Biol. Chem.* **2003**, *278*, 14461–14468.
[32] B. Wang, J. Pelletier, M. J. Massaad, A. Herscovics, G. C. Shore, *Mol. Cell Biol. Res. Commun.* **2004**, *24*, 2767–2778.
[33] C. M. Paumi, *Mol. Cell* **2007**, *26*, 15–25.
[34] M. J. Massaad, A. Herscovics, *J. Cell Sci.* **2001**, *114*, 4629–4635.
[35] S. Thaminy, D. Auerbach, A. Arnoldo, I. Stagljar, *Genome Res.* **2003**, *13*, 1744–1753.
[36] M. Dünnwald, A. Varshavsky, N. Johnsson, *Mol. Biol. Cell* **1999**, *10*, 329–344.
[37] H. Laser, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 13732–13737.
[38] S. Wittke, M. Dünnwald, N. Johnsson, *Mol. Biol. Cell* **2000**, *11*, 3859–3871.
[39] T. Ozawa, Y. Umezawa, *Curr. Opin. Chem. Biol.* **2001**, *5*, 578–583.
[40] D. T. Ng, J. D. Brown, P. Walter, *J. Cell. Biol.* **1996**, *134*, 269–278.
[41] S. Wittke, M. Dunnwald, M. Albertsen, N. Johnsson, *Mol. Biol. Cell* **2002**, *13*, 2223–2232.
[42] B. Nyfeler, S. W. Michnick, H. P. Hauri, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6350–6355.
[43] S. B. Kim, T. Ozawa, S. Watanabe, Y. Umezawa, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 11542–11547.

[44] R. Paulmurugan, S. S. Gambhir, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 15883–15888.
[45] A. Yan, E. Wu, W. J. Lennarz, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 7121–7126.
[46] I. Remy, I. A. Wilson, S. W. Michnick, *Science* **1999**, *283*, 990–993.
[47] G. Dues, S. Muller, N. Johnsson, *FEBS Lett.* **2001**, *505*, 75–80.
[48] N. Johnsson, *FEBS Lett.* **2002**, *531*, 259–264.
[49] J. Jeong, *Biochem. Biophys. Res. Commun.* **2006**, *339*, 647–651.
[50] I. Remy, I. S. W. Michnick, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 5394–5399.
[51] T. S. Wehrman, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 19063–19068.
[52] X. Wang, N. Johnsson, *J. Cell Sci.* **2005**, *118*, 723–732.
[53] J. N. Pelletier, K. M. Arndt, A. Pluckthun, S. W. Michnick, *Nat. Biotechnol.* **1999**, *17*, 683–690.
[54] E. J. Licitra, J. O. Liu, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 12817–12821.
[55] D. Dirnberger, G. Unsin, S. Schlenker, C. Reichel, *ChemBioChem* **2006**, *7*, 936–942.
[56] R. Paulmurugan, S. S. Gambhir, *Cancer Res.* **2005**, *65*, 7413–7420.
[57] C. I. Stains, J. R. Porter, A. T. Ooi, D. J. Segal, I. Ghosh, *J. Am. Chem. Soc.* **2005**, *127*, 10782–10783.
[58] M. R. Pratt, E. C. Schwartz, T. W. Muir, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 11209–11214.
[59] R. Paulmurugan, Y. Umezawa, S. S. Gambhir, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 15608–15613.
[60] H. D. Mootz, T. W. Muir, *J. Am. Chem. Soc.* **2002**, *124*, 9044–9045.
[61] H. D. Mootz, E. S. Blum, T. W. Muir, *Angew. Chem.* **2004**, *116*, 5301–5304; *Angew. Chem. Int. Ed.* **2004**, *43*, 5189–5192.
[62] A. Varshavsky, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 14935–14940.
[63] E. Rojo-Niersbach, D. Morley, S. Heck, N. Lehming, *Biochem. J.* **2000**, *348*, 585–590.
[64] J. P. Miller, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12123–12128.
[65] B. S. Chew, N. Lehming, *Biochem. J.* **2007**, *406*, 265–271.
[66] J. H. Eckert, N. Johnsson, *J. Cell Sci.* **2003**, *116*, 3623–3634.
[67] P. Obrdlik, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 12242–12247.
[68] C. Mo, M. Bard, *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* **2005**, *1737*, 152–160.
[69] E. Nikko, B. Andre, *Eukaryotic Cell* **2007**, *6*, 1266–1277.
[70] D. Dirnberger, M. Messerschmid, R. Baumeister, *Nucleic Acids Res.* **2008**, *36*, e37.
[71] C. Reichel, N. Johnsson, *Methods Enzymol.* **2005**, *399*, 757–776.
[72] C. Tiedje, D. G. Holland, U. Just, T. Hofken, *J. Cell Sci.* **2007**, *120*, 3613–3624.
[73] R. Paulmurugan, S. S. Gambhir, *Anal. Chem.* **2007**, *79*, 2346–2353.

Received: March 27, 2008
Published online on August 1, 2008

