

Protein Reconstitution Methods for Visualizing Biomolecular Function in Living Cells

Takeaki OZAWA

Department of Chemistry, School of Science, The University of Tokyo, 7-3-1 Bunkyo-ku, Hongo, Tokyo, and PRESTO, Japan Science and Technology Agency, Tokyo, Japan

(Received September 19, 2008)

One of the most challenging researches in current biology and medicinal chemistry is to understand how individual cellular molecules interact together in living cells. To visualize such molecules, genetically-encoded reporters have been used widely. The most common reporters are firefly luciferase, *renilla* luciferase, green fluorescent protein (GFP) and its variants with various spectral properties. In this review, novel design of split GFP and split luciferase is described. The principle is based on reconstitution of the split-reporter fragments when they are brought together into close proximity. The reconstitution methods are used for screening organelle-localized proteins, imaging dynamics of nuclear proteins and mRNAs in living cells, and visualizing protease activities in living animals. These methods are generally applicable for imaging of complex cellular processes and evaluating chemical effects in living cells and animals.

Key words—green fluorescent protein; luciferase; mRNA; mitochondria

INTRODUCTION

Many dynamic intracellular processes inside living cells are governed over timescales ranging from milliseconds to several hours. To understand the temporal signal processes, development of optical probes for biomolecules, protein-protein interactions, enzyme activities, and protein conformational changes at molecular level is an important area of present researches. Most of our knowledge on the mechanism of living systems has been obtained by biochemical assays such as Western blotting and immunostaining. The Western blotting requires a large population of cells, which have to be used on the lysate. Cell lysis prior to the assay may result in alteration or loss of information. The immunostaining used for immunofluorescence microscopic studies with chromophore-conjugated antibodies includes a procedure of cell fixation. The fixation gives spatial information but it does not allow for dynamic studies in the same live cells and may introduce artifacts.

Recent advances in luminescent proteins and microscope technologies have provided powerful approaches that present advantages over the traditional bioassay. In particular, newly developed recombinant proteins as genetically-encoded reporters are useful tools for imaging the signaling cascades in live cells

with high temporal and spatial resolutions, thus enabling the elucidation of the mechanism of living systems. In this review, we summarize basic principle of novel design of split GFP and split luciferase based on reconstitution of their fragments, followed by their applications to biomolecular imaging and library screening.

BASIC PRINCIPLES OF SPLIT- AND CYCLIC-REPORTER RECONSTITUTION

Fluorescent proteins of GFP and its variants have proved to be useful as general tools for detecting subcellular localization of a protein, pH differences, small-molecule dynamics and protein stabilization.^{1,2} In addition, bioluminescent proteins such as firefly luciferase, *renilla* luciferase and *Gaussia* luciferase are now widely used to assess activation of promoters which up- or down-regulate transcription of their reporter genes.³ Herein, these fluorescent and bioluminescent proteins are simply termed as reporter proteins.

In contrast to the conventional use of the reporter proteins, we developed a novel principle of designing optical probes to visualize biomolecules and their molecular events in living cells (Fig. 1).^{4,5} When the reporter protein is divided into two fragments, its fluorescence or bioluminescence activity is completely lost. Next, the two fragments are connected with a protein named intein, which has an ability to connect its flanking polypeptide (Fig. 1a). If the fragments of

*e-mail: ozawa@chem.s.u-tokyo.ac.jp

本総説は、日本薬学会第128年会シンポジウムS36で発表したものを中心に記述したものである。

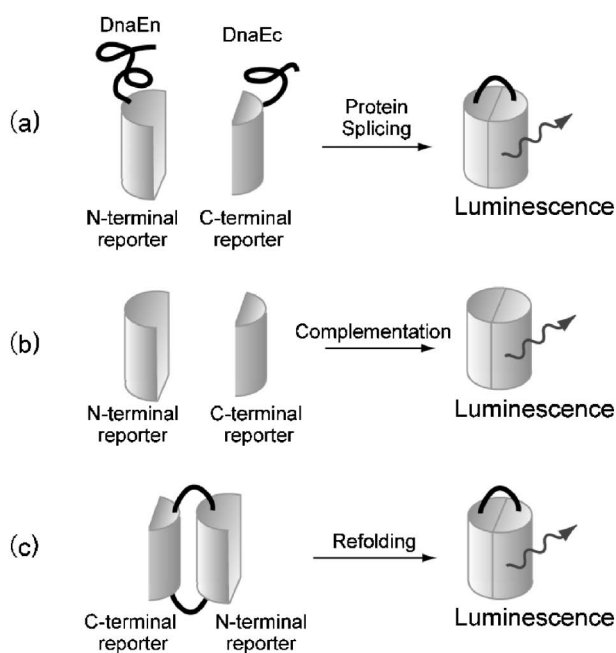


Fig. 1. Principles of Split- and Cyclic-luciferase Reconstitution

(a) Reporter protein reconstitution using protein splicing of DnaE inteins; (b) Complementation of split fragments of a reporter protein; (c) Reconstitution using refolding of a structurally-distorted protein.

the fusion protein are expressed in the cells, they generate a reconstituted reporter protein from two of its split-reporter fragments by a protein splicing reaction. A typical intein segment consists of 400–500 amino acid residues.⁶⁾ At both ends of the intein, conserved amino acid residues are inserted for efficient protein splicing reaction. While most inteins are composed of a single polypeptide chain, *DnaE* proteins from the split *dnaE* genes in the genome of *Synechocystis* sp. PCC6803 are naturally split inteins. The *Ssp* DnaE inteins can mediate a *trans*-splicing reaction with a higher efficiency when fused to foreign proteins. An important general feature of the protein splicing is a self-catalyzed excision of the intein without any exogenous cofactors or energy sources such as ATP.

Another approach to reconstructing the split fragments of a reporter protein is to use their complementation,^{7–11)} in which inteins are not required for the reconstitution (Fig. 1b). The rationally-split fragments of a reporter protein are connected with a pair of interacting proteins. Interaction of the fused proteins leads to correct folding of the reporter protein and recovering its luminescence activity. Such reconstitution strategy has been successfully used in the

reassembly of dihydrofolate reductase, β -lactamase, ubiquitin, *renilla* luciferase, firefly luciferase, GFP and its spectral variants.

Efficiency of reconstitution largely depends on the dissecting amino acids in the reporter proteins. In addition, split reporters have to be designed to meet the criteria such that each fragment does not possess fluorescence or bioluminescence activity. Moreover, fragments of the split reporters do not assemble nor refold spontaneously. Several dissecting points of GFP are found amenable to fulfill these criteria. Among the dissecting points, the most generally useful point of dissection appears to be in a surface loop near residues 157–158, spatially near the N- and C-termini, regardless of fusion topology. Strikingly, it has been suggested that a wide range of lengths of linkers, which are used for connection between interacting proteins and split reporter fragments, lead to reconstitution.¹²⁾ Precise alignment of the GFP fragments is not necessary to nucleate reassembly.

In addition to the split-fragments reconstitution, an innovative design of reconstitution method, named cyclic luciferase, has recently been developed (Fig. 1c).¹³⁾ Two fragments of DnaE inteins are fused to neighboring N- and C-terminal ends of firefly luciferase connected with a substrate sequence of protease. After translation into a single polypeptide, the N- and C-terminal ends of luciferase are ligated by protein splicing, which results in a closed circular polypeptide chain. Because the structure of the cyclic luciferase is distorted, the luciferase loses its bioluminescence activity. If the substrate sequence is digested by a protease, the luciferase changes into an active form and restores its activity. The usefulness of the cyclic luciferase has been demonstrated such as a method for quantitative sensing of caspase-3 activities in living cells upon extracellular stimuli and for noninvasive imaging of the time-dependent caspase-3 activities in living mice.

cDNA LIBRARY SCREENING USING SPLIT-EGFP RECONSTITUTION BY PROTEIN SPLICING

Localization of each protein in eukaryotic cells is tightly bound to function. Therefore, functional assays and screenings intended to characterize protein localization are very important for understanding cellular function. Now the methods for identifying proteins localized in the intracellular organelles largely

depend on the biochemical approaches.¹⁴⁾ Each organelle is fractionated by stepwise centrifugation, each protein is separated from the organelle by electrophoresis or liquid chromatography, and its amino acid composition is identified by mass spectrometry. This biochemical approach is useful for systematic identification, but it depends on the yield and purity of the intracellular organelle. Thus, the technique can be problematic for organelles that are difficult to purify.

The split GFP reconstitution strategy has overcome this drawback. In contrast to the biochemical approach, we have developed a genetic approach with the split-GFP reconstitution technique,¹⁵⁾ in which screening for mitochondrial proteins from cDNA libraries resulted in the discovery of novel genes that encode mitochondrial proteins. The basic strategy is based on reconstitution of enhanced GFP (EGFP) by protein splicing with DnaE inteins (Fig. 2). As described above, a characteristic feature of DnaEs is of natural splicing ability to ligate accompanying exteins *in trans*. A tandem fusion protein, containing a mitochondrial targeting signal (MTS) and C-terminal

fragments of DnaEs and EGFP, localizes in the mitochondrial matrix in cultured mammalian cells. cDNA libraries generated from mRNAs are genetically fused to the sequences encoding the N-terminal fragments of EGFP and DnaEs. If test proteins expressed from the cDNA libraries contain a functional MTS, the fusion products translocate into the mitochondrial matrix. The N- and C-terminal fragments of DnaEs are brought close enough to fold correctly, thereby initiating protein splicing to link the EGFP fragments with a peptide bond. It is noteworthy that only “mitochondria-positive” cells yield fluorescence signals. These fluorescent cells are screened and collected rapidly by a fluorescence-activated cell sorter (FACS). Relevant genes are subsequently identified by genetic PCR and their cDNAs are sequenced. In this work, 30 mitochondrial proteins including 3 novel proteins were identified as the mitochondrial. If the C-terminal fragments of EGFP and DnaEs are placed in the mitochondrial inter-membrane space (IMS), a series of IMS-localized proteins can be identified according to the same procedure. A short peptide of the IMS-targeting signal was identified by the split-EGFP reconstitution methods.¹⁶⁾

This genetic approach to identifying mitochondrial proteins was further extended for identifying proteins targeted to the endoplasmic reticulum.¹⁷⁾ In this case, a fusion protein of the C-terminal fragments of DnaE and EGFP is connected with an endoplasmic reticulum-targeting signal (ERTS), which localizes in the lumen of the endoplasmic reticulum. If test proteins expressed from the cDNA library contain an ERTS, the fusion products translocate into the endoplasmic reticulum, in which mature EGFP is generated. Using the same procedure as the mitochondrial case, the fluorescent cells including cDNAs are collected by FACS, and relevant genes are identified. In this work, 109 non-redundant proteins targeting to the endoplasmic reticulum were identified, of which 27 proteins were novel.

An advantage of these genetic methods over the biochemical ones is that they do not require purification of target organelles or separation of each protein. Therefore, the proteins localized in other organelles do not contaminate. Moreover, the methods provide high-throughput identification of gene products that are targeted to the mitochondria or endoplasmic reticulum.

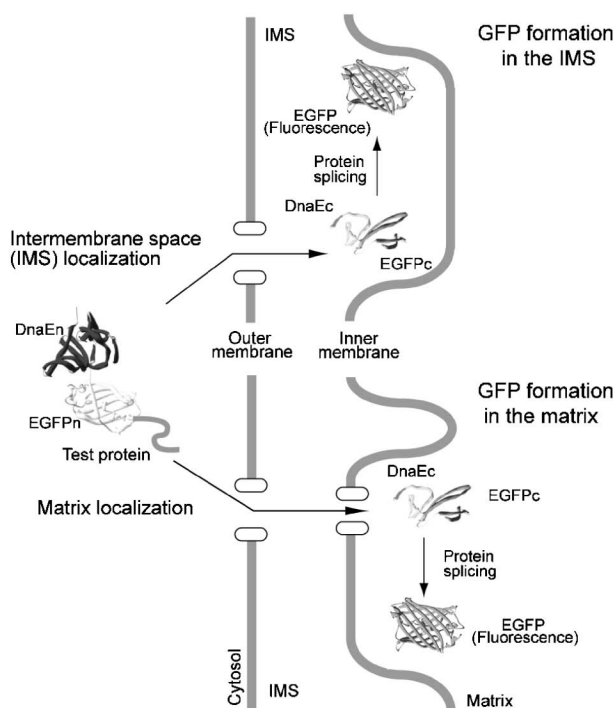


Fig. 2. Principle of Methods to Identify Mitochondrial Proteins

When a test protein moves into the IMS, only the upper cell line is fluorescent. Conversely, when a test protein moves into the mitochondrial matrix, the lower cell line shows fluorescence. DnaEn, N-terminal DnaE intein; DnaEc, C-terminal DnaE intein; IMS, mitochondrial intermembrane space.

DETECTION OF PROTEIN TRAFFICKING USING SPLIT-LUCIFERASE RECONSTITUTION BY PROTEIN SPLICING

Dynamics of proteins into and out of the nucleus plays an important role in regulating gene expression in response to extracellular signals. The movement is regulated by posttranslational modification such as ligand binding, phosphorylation and proteolysis. A technique for monitoring the protein movement across the nuclear membrane depends on the use of immunocytochemistry or optical imaging with genetically tagged GFP or its variant. These analyses are effective for imaging spatial and temporal dynamics of proteins of interest under fluorescence microscopes. However, image acquisition is slow and the obtained results are qualitative rather than quantitative. In addition, analyses of the protein localization in living animals require complex assay procedures, such as extraction of an organ and dividing it into sliced sections, which hamper temporal and quantitative analyses.

We have developed a bioluminescent reporter for detecting protein transport into the nucleus using split-*renilla* luciferase reconstitution.¹⁸⁾ *Renilla* luciferase has desirable features for a monomeric protein: small size (36 kDa), strong luminescence, and lack of necessity of ATP. In addition, its substrate, coelenterate luciferin (coelenterazine) rapidly penetrates through cell membranes, which is suitable for a live cell assay and *in vivo* imaging. The luciferase is split into N- and C-terminal fragments, and they are connected with N- and C-terminal fragments of DnaEs, respectively. The C-terminal fragment is permanently located in the nucleus, while the N-terminal fragment connected with a test protein is in the cytosol. If the test protein translocates into the nucleus, the N- and C-terminal fragments of *renilla* luciferase coexist in the nucleus, and full-length *renilla* luciferase is reconstituted by protein splicing. The usefulness of the reporter is demonstrated with a ligand-binding type of nuclear receptor, androgen receptor (AR). Upon binding to 5 α -dihydrotestosterone (DHT), AR translocated from the cytosol into the nucleus. The luminescence signals were found to increase with increasing the concentration of DHT, being strong enough to discriminate them from background luminescence. Further extending of the split *renilla* luciferase reporter was demonstrated in the

case where protein translocation is triggered by proteolysis, phosphorylation and stress.^{19–22)}

The advantage of the split *renilla* luciferase reporter is that the number of the cells once analyzed was $\sim 10^4$ cells, which was enough to precisely evaluate the extent of AR translocation into the nucleus. Therefore, this split reporter enabled precise and sensitive detection, which is of great advantage to quantitatively evaluate the extent of protein nuclear transport in a high-throughput manner.

IMAGING OF mRNA IN SINGLE LIVING CELLS USING SPLIT-EGFP COMPLEMENTATION

Methods of split-reporter complementation without intein have found applications as a biochemical tool to detect and image specific protein-protein interactions and protein-nucleic acid interactions. In addition to such applications, we have recently developed genetically-encoded fluorescent probes for characterizing localization and dynamics of an mRNA in single living cells.²³⁾ The basic principle is shown in Fig. 3. Split fragments of EGFP are connected with two sequence-specific RNA binding domain of human PUMILIO1 (Pumilio homology domain; PUM-HD). PUM-HD is composed of eight sequence repeats and recognizes a consensus sequence, UGUANAUA.^{24,25)} Each repeat acts as a molecule that specifically recognizes a single RNA base like a Watson-Crick base pairing. Importantly, specificity of PUM-HD for the RNA-base sequences can be altered by changing RNA-recognizing amino acid residues within the repeated sequence. Upon interacting RNA with the PUM-HDs, the EGFP fragments are in close proximity so that they are reconstituted into the native EGFP, recovering its fluorescence. The fluorescence signals allow for monitoring spatial and temporal analysis of mRNA localization in single living cells.

In the proof of principle, a mitochondrial mRNA (mtRNA) encoding NADH dehydrogenase subunit 6 (ND6) was selected as a target for imaging. The ND6 mRNA has two sequence motifs similar to the PUM-HD recognition sites. To match their sequences, several amino acid mutations are inserted in the PUM-HDs. In order to target the RNA probes into the mitochondrial matrix, a mitochondria-targeting signal (MTS) was connected with the amino terminus of the probes.

Using the probes, ND6 mRNA was successfully

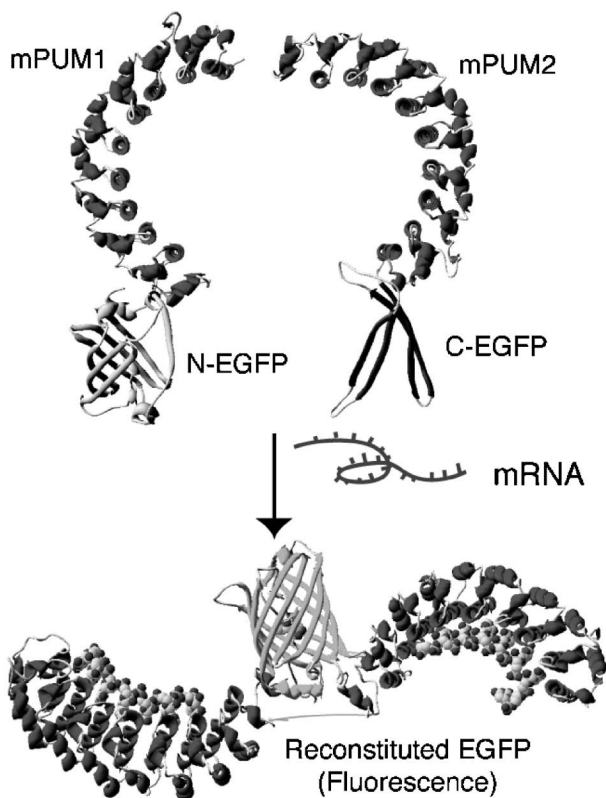


Fig. 3. Detection Methods of a Target mRNA Based on Complementation of Split EGFP Fragments

Two RNA-binding proteins of PUM-HD are engineered to recognize specific sequences on a target mRNA (mPUM1 and mPUM2). In the presence of the target mRNA, mPUM1 and mPUM2 bind to their target sequences bringing together the N- and C-terminal fragments of EGFP, resulting in functional reconstitution of EGFP.

visualized in single living cells. It was found that, in the case of HeLa cells, the complement EGFP was localized in mitochondria. In addition, several foci on the mitochondrial DNA (mtDNA) were observed. This result suggested that ND6 mRNA is concentrated on the mtDNA. The dynamics of ND6 mRNA was examined by the fluorescence recovery after photobleaching (FRAP) technique. Movement of ND6 mRNA was found restricted but oxidative stress induced the mRNA to disperse in the mitochondria and gradually decompose. Although usefulness of the probe has been demonstrated only in the case of ND6 mRNA, but, in the near future, the basic principle will provide a means to study spatial and temporal dynamics of a variety of mRNAs in living cells.

SENSING CASPASE-3 ACTIVITIES IN LIVING MICE USING CYCLIC LUCIFERASE RECONSTITUTION

One of the complicated proteolytic signaling inside

living organisms is used to decide the programmed cell death, apoptosis. Understanding the physiological proteolytic processes is of crucial importance in assessing the roles of proteases in normal states and diseases. Development of a rapid screening system to detect the protease activities and a method to non-invasively image the apoptosis is essential for the discovery of novel compounds that can be potential therapeutic chemicals. We developed a genetically-encoded cyclic luciferase with general applicability for detecting protease activities in living cells and animals.¹³⁾

The principle is shown in Fig. 4. Firefly luciferase folds into two compact domains; a large N-terminal domain (Fluc-N) and a small C-terminal one (Fluc-C).²⁶⁾ These domains are separated by a wide cleft, in which there is an active center of the luciferase. The N-terminal end of firefly luciferase is distant from the C-terminal one, but both ends of the luciferase are located on the same side behind the active center. On the basis of the structural information of the luciferase, the N- and C-terminal ends are connected with a substrate sequence of caspase-3, DEVD. By doing so, the overall structure is changed into a distorted form, which results in great attenuation of the luciferase activity. In order to connect the N-terminal end with the C-terminal one, naturally split DnaE inteins are used. In addition, a PEST sequence, which is known to accelerate degradation of the protein in 4 hours,²⁷⁾ was attached to the C-terminal ends of the fusion construct. The PEST sequence results in the degradation of only unspliced products, whereas cyclic luciferase does not possess the PEST sequence. Consequently, cyclic luciferase uniquely accumulates in the cells.

Applicability of the cyclic luciferase was demonstrated for quantitatively detecting the caspase-3 activities in living cells and animals. Cell-based analysis using the cyclic luciferase allowed for precise and quantitative measurements of caspase-3 activities, because it enabled analysis of a statistically significant number of cells in a single assay format. The response of cyclic luciferase upon caspase-3 activation was very fast, which will promise high-throughput screening and characterization of therapeutic anticancer drugs and caspase inhibitors. Moreover, the cyclic luciferase enabled *in vivo* real-time imaging of caspase-3 activities in living mice. Chemical compounds, in many cases, are metabolized or chemically modified

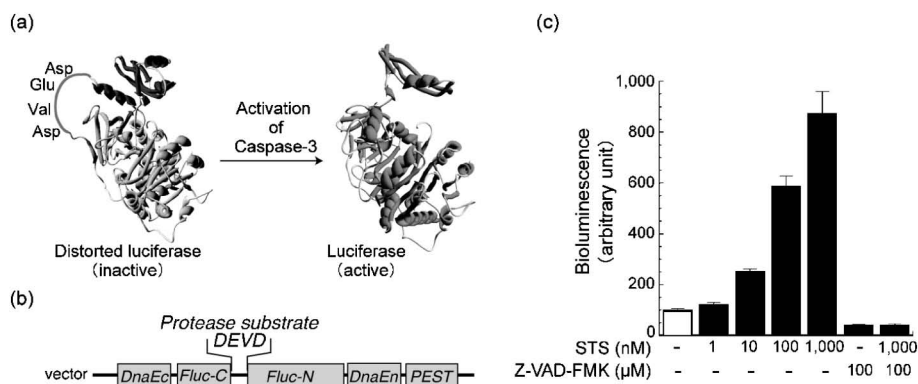


Fig. 4. Strategy for the Detection of Caspase-3 Activity

(a) Principle for monitoring the activity of caspase-3 by using cyclic luciferase; (b) Schematic structures of the cDNA construct of the probe; (c) *In vitro* characterization of cyclic luciferase. The probe was expressed in COS-7 cells and the cells were stimulated with staurosporine (STS) and/or an inhibitor, Z-VAD-FMK.

in living mice. Effective concentrations of the chemical compounds can be estimated noninvasively by using this imaging method.

CONCLUSION

In conclusion, split-reporter reconstitution methods have been applied to advance our understanding of many important molecular and cellular functions in living cells and animals. Novel techniques and related applications of the methods are constantly being developed and reported. At the present time, the development of the split reporter constructs is rather semi-rational and laborious. This would be improved if systematic design and high-throughput screening of the probes were possible. Further development of a variety of attractive applications with the new reconstitution methods will appear in the near future.

Acknowledgments This work was supported by the Japan Science and Technology Agency (JST), and the Japan Society for the Promotion of Science (JSPS).

REFERENCES

- Shaner N. C., Steinbach P. A., Tsien R. Y., *Nat. Methods.*, **2**, 905–909 (2005).
- Zhang J., Campbell R. E., Ting A. Y., Tsien R. Y., *Nat. Rev. Mol. Cell Biol.*, **3**, 906–918 (2002).
- Massoud T. F., Gambhir S. S., *Genes Dev.*, **17**, 545–580 (2003).
- Ozawa T., Kaihara A., Sato M., Tachihara K., Umezawa Y., *Anal. Chem.*, **73**, 2516–2521 (2001).
- Ozawa T., Nogami S., Sato M., Ohya Y., Umezawa Y., *Anal. Chem.*, **72**, 5151–5157 (2000).
- Evans T. J. T., Xu M. Q., *Chem. Rev.*, **102**, 4869–4884 (2002).
- Ghosh I., Hamilton A. D., Regan L., *J. Am. Chem. Soc.*, **122**, 5658–5659 (2000).
- Kaihara A., Kawai Y., Sato M., Ozawa T., Umezawa Y., *Anal. Chem.*, **75**, 4176–4181 (2003).
- Paulmurugan R., Umezawa Y., Gambhir S. S., *Proc. Natl. Acad. Sci. USA*, **99**, 15608–15613 (2002).
- Remy I., Michnick S. W., *Nat. Methods*, **3**, 977–979 (2006).
- Hu C. D., Kerppola T. K., *Nat. Biotechnol.*, **21**, 539–545 (2003).
- Magliery T. J., Wilson C. G., Pan W., Mishler D., Ghosh I., Hamilton A. D., Regan L., *J. Am. Chem. Soc.*, **127**, 146–157 (2005).
- Kanno A., Yamanaka Y., Hirano H., Umezawa Y., Ozawa T., *Angew. Chem. Int. Ed. Engl.*, **46**, 7595–7599 (2007).
- Pham T. K., Wright P. C., *Expert Rev. Proteomics*, **4**, 793–813 (2007).
- Ozawa T., Sako Y., Sato M., Kitamura T., Umezawa Y., *Nat. Biotechnol.*, **21**, 287–293 (2003).
- Ozawa T., Natori Y., Sako Y., Kuroiwa H., Kuroiwa T., Umezawa Y., *ACS Chem. Biol.*, **2**, 176–186 (2007).
- Ozawa T., Nishitani K., Sako Y., Umezawa Y., *Nucleic Acids Res.*, **33**, e34 (2005).

- 18) Kim S. B., Ozawa T., Watanabe S., Umezawa Y., *Proc. Natl. Acad. Sci. USA*, **101**, 11542–11547 (2004).
- 19) Kanno A., Ozawa T., Umezawa Y., *Anal. Chem.*, **78**, 8076–8081 (2006).
- 20) Kim S. B., Ozawa T., Umezawa Y., *Anal. Chem.*, **77**, 6588–6593 (2005).
- 21) Kim S. B., Ozawa T., Umezawa Y., *Anal. Biochem.*, **347**, 213–220 (2005).
- 22) Kim S. B., Takao R., Ozawa T., Umezawa Y., *Anal. Chem.*, **77**, 6928–6934 (2005).
- 23) Ozawa T., Natori Y., Sato M., Umezawa Y., *Nat. Methods*, **4**, 413–419 (2007).
- 24) Cheong C. G., Hall T. M., *Proc. Natl. Acad. Sci. USA*, **103**, 13635–13639 (2006).
- 25) Wang X., McLachlan J., Zamore P. D., Hall T. M., *Cell*, **110**, 501–512 (2002).
- 26) Conti E., Franks N. P., Brick P., *Structure*, **4**, 287–298 (1996).
- 27) Li X. Q., Zhao X. N., Fang Y., Jiang X., Duong T., Fan C., Huang C. C., Kain S. R., *J. Biol. Chem.*, **273**, 34970–34975 (1998).