

Award Accounts

The Chemical Society of Japan Award for Young Chemists for 2003

Methods of Analysis for Protein Dynamics in Living Cells Based on Protein Splicing

Takeaki Ozawa

Department of Chemistry, School of Science, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033

PREST, Japan Science and Technology Agency, 4-1-8 Honcho Kawaguchi, Saitama 332-0012

Received September 21, 2004; E-mail: ozawa@ims.ac.jp

Protein–protein interactions and protein localization have key roles in many essential biological processes in living cells. We have developed novel reporter proteins with general applicability for detecting the biological processes in living cells and animals. The principle is based on reconstitution of split reporter proteins by protein splicing, which involves a self-catalyzed excision of a protein splicing element, intein, from flanking polypeptide sequences, exteins, leading to ligation of the flanking exteins by a peptide bond. As the exteins, N- and C-terminal halves of rationally-split green fluorescent protein (GFP) or split bioluminescent proteins were used. The N- and C-terminal reporters connected respectively with a pair of interacting proteins worked as indicators for protein–protein interactions in bacteria and mammalian cells. The split GFP reporter provided a genetic method for identifying mitochondrial proteins from large-scale complementary DNA libraries. The split bioluminescent reporter enabled high-throughput sensing and noninvasive imaging of nuclear transport of target proteins in living animals. This basic concept of split reporter reconstitution by protein splicing provides a wide variety of applications not only for fundamental biological studies, but also for assay and screening methods for chemicals that inhibit or facilitate the biological processes in living cells.

Developments of novel methods and advances in biotechnology have provided many basic tools that allow quantifying and imaging intracellular signals and events in living subjects. To investigate the biological processes, fluorescent and bioluminescent proteins, which are called "reporters", have been used as a read out.[1–3] One of the most useful reporters is a green fluorescent protein (GFP) derived from *Aequorea victoria*, or its variants.[4] It is particularly valuable due to its structural stability and the fact that its chromophore is spontaneously formed in an autocatalytic cyclization that does not require any cofactor. Generally, the gene encoding GFP is fused with a gene of interest so that the fusion protein emits fluorescence in living cells. The fusion protein can then be followed by optical imaging to deduce spatial and temporal information on gene expressed from *Photinus pyralis* and *renilla* luciferase from *Renilla reniformis*.[5] The firefly and *renilla* luciferase reporters have potential to allow very high-sensitive and background-free detection. The luciferase reporters have been used as the reporter gene assays, where the activation of transcription factors has been assessed by their binding to promoters that upregulate the expression of the reporter proteins. GFP and luciferase reporters are now crucial for understanding complex cellular processes that depend on when, where, and how much of a target protein is present. However, analysis of the cellular processes in living cells with the fluorescent and bioluminescent proteins is limited only to the use of the correctly-refolded proteins that possess their fluorescence and luminescence activity.

While a wide variety of applications with the fluorescent and bioluminescent proteins were emerging in the 1990s, I read an attractive report concerning the formation of vacuolar H^+-ATPase in *Saccharomyces cerevisiae*.[6,7] Protein biosynthesis was initially thought to be a simple process; the genetic information in DNA is directly copied into messenger RNAs (mRNAs), which in turn direct the biosynthesis of proteins. However, in *Saccharomyces cerevisiae*, a nascent 120 kDa translational product of the *VMA1* gene autocatalytically excised out a 50 kDa site-specific endonuclease (*VMA1*-derived endonuclease; VDE, also called as PI-*Sce*I) and spliced the two external polypeptides (exteins) to form a 70 kDa catalytic subunit of vacuolar H^+-ATPase (Fig. 1A). After the first report about protein splicing, further biochemical evidence has led to the conclusion that protein splicing did not occur at the RNA level, but instead at the protein level without any cofactors.[8] Only the internal protein segment (termed an intein) of VDE was found to be essential for self-splicing activity. Since the discovery of the protein splicing, more than 170 putative inteins have been identified in eubacteria, archea, and eukaryotic unicellular organisms (http://www.neb.com/neb/inteins.html).[9]

The protein splicing is a multi-step processing event involving excision of an intein segment from a primary translation product and concomitant ligation of the flanking sequences

Published on the web May 6, 2005; DOI 10.1246/bcsj.78.739

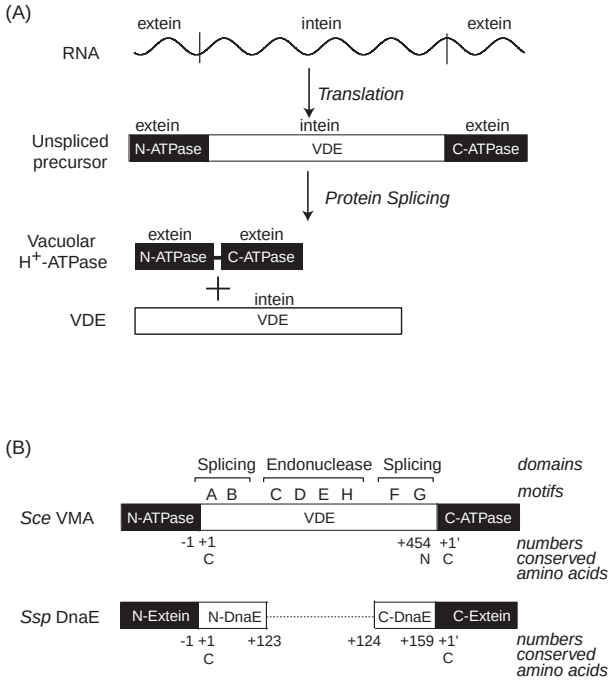

Fig. 1. Intein domains, motifs, conserved amino acids, and mechanism of protein splicing. (A) Posttranslational modification by protein splicing. Genetic information in DNA is directly copied into mRNA, which in turn directs the biosynthesis of a precursor protein (translation). An intervening sequence, termed intein (VDE), is excised from an internal site in a precursor protein and the surrounding polypeptides (exteins) are ligated to form the matured vacuolar H^+-ATPase. (B) Conserved amino acids and motifs in typical inteins. Domains A and B contain the N-terminal splicing motifs, while domains F and G contain the C-terminal splicing motifs. The first amino acid residue upstream of the intein is numbered -1, and the intein is numbered sequentially beginning with the N-terminal amino acid residue. The C-extein residue is numbered $+1'$. Conserved amino acid residues at position $+1$ and $+1'$ are indicated.

of external proteins, exteins.[10,11] A typical intein segment consists of 400–500 amino acid residues and contains four conserved protein splicing motifs, A, B, F, and G, as well as a homing endonuclease sequence embedded between the motifs B and F (Fig. 1B). The endonuclease sequence can be deleted from the intein sequence while still retaining the activity of protein splicing. At intein–extein junctions, conserved amino acid residues are directly involved in the protein splicing reaction: They include Cys or Ser at the intein amino terminus, and Asn or Gln at the intein carboxy terminus, and Cys, Thr, or Ser at the beginning of the C-extein. While VDE and most of inteins are composed of a single polypeptide chain, a pair of functional and naturally split intein-coding sequences have been found from the split *dnaE* genes in the genome of *Synechocystis* sp. PCC6803.[12] The *Ssp* DnaE intein 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 and ligation of the flanking exteins without any exogenous cofactors or energy sources such as ATP or GTP. This feature has encouraged great ideas of reconstitution of split reporter proteins and their important applications.

1. Basic Concept of Reporter Protein Reconstitution

In 1998, we developed a new idea of novel reporters, in which protein splicing generates a reconstituted reporter protein from two of its split reporter fragments. One of the examples is reconstitution of the split GFP.[13] GFP consists of 238 amino acids and its structure is of eleven strands of β-sheet that form a barrel structure with short α-helices forming lids on each end.[14,15] The fluorescence-active center of GFP is located inside the barrel. If the structure of GFP is split into two fragments at a peptide bond, the fluorescence of GFP is completely lost. Our interest was whether the two fragments of split GFP are reconnected together by protein splicing in living cells and thereafter, whether its connected GFP fragment refolds correctly and recovers its fluorescence (Fig. 2A). We examined whether protein splicing occurs in a single polypeptide consisting of VDE and N- and C-terminal fragments of GFP dissected at an amino acid position between 128 and 129 (Fig. 2B). To analyze the products after protein expression in *E. coli*, crude and affinity-purified extracts from *E. coli* were subjected to SDS-PAGE. As shown in Fig. 2C, anti-VDE and anti-GFP antibodies reacted with the excised 50-kDa VDE (lane 4) and 53-kDa GST-GFP fusion protein (lane 5), respectively. This result indicated that a 103-kDa precursor of the GST fusion protein was processed into a 50-kDa VDE and 53-kDa GST-GFP fusion protein. The identities of the GST-GFP fusion protein were further verified by purification with a GST-affinity column. The crude extract was applied to the affinity column and bound proteins to the resin were excised with PreScission protease. The excised proteins were analyzed by SDS-PAGE. A single 25-kDa band was observed (lane 3), which was almost the same as the calculated mass of GFP. Fluorescence spectra of the affinity-purified protein were further measured; excitation and emission maximums were found to be 488 and 510 nm, respectively. Such locations were consistent with those of GFP. Bacteria that included pGEX vector showed strong fluorescence in living cells (Fig. 2D). From these results, we conclude that spliced intervening VDE was excised out and that the two external regions of the N- and C-terminal halves of GFP were ligated with a peptide bond, and remarkably, that the ligated protein of GFP folded correctly and the fluorophore was formed in living cells. The success of the GFP reconstitution was found to largely depend on the dissection point of GFP. The reconstitution occurred only when GFP were dissected at amino acid positions of 129, 145, and 157 from the amino terminus.

We have demonstrated that the basic concept of this GFP reconstitution by protein splicing is of general use for any reporter proteins. Firefly luciferase oxidizes its substrate D-luciferin to result in light emission (bioluminescence). The luciferase is used as a general reporter for immunoassays and reporter gene assays. The 3-D structure of firefly luciferase is folded into two compact domains: the large N-terminal globular domain, and the small C-terminal domain (Fig. 3).[16] The N- and C-terminal domains are separated by a wide cleft, which is the location of the active site of the luciferase. We showed that when the lu-

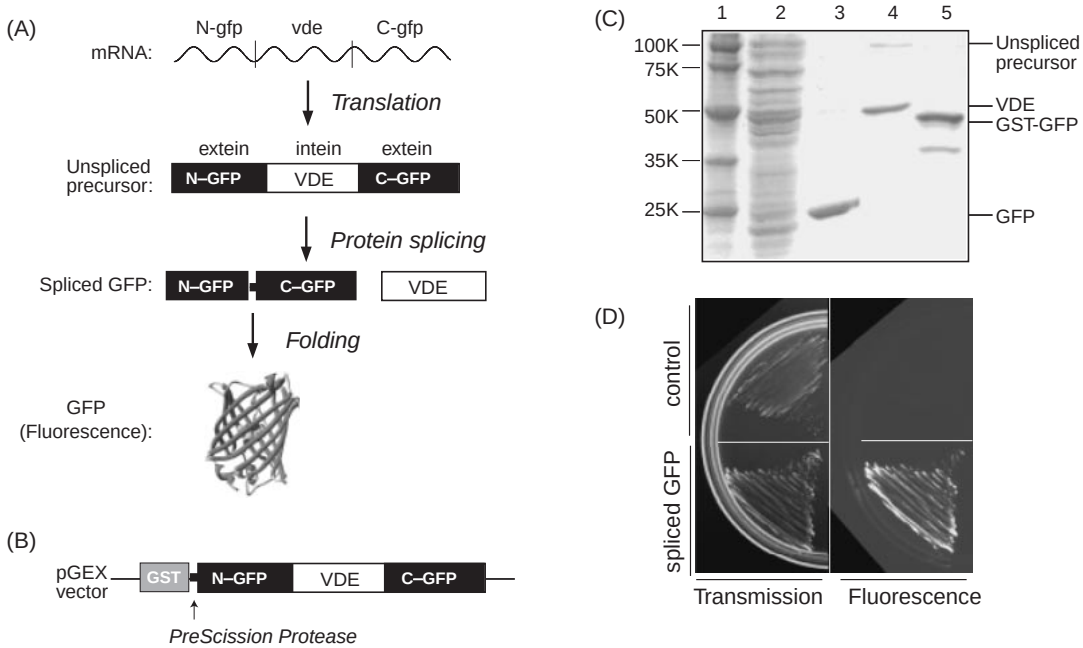

Fig. 2. Split GFP reconstitution by protein splicing with VDE. (A) Basic scheme. A single polypeptide, composed of 128 amino acid residues from the N-terminal half of GFP (N-GFP), 454 amino acid residues from VDE, and 110 amino acid residues from C-terminal half of GFP (C-GFP), undergoes protein splicing and thereby N- and C-terminal fragments of GFP ligate by a peptide bond. The matured GFP thus formed folds correctly and its fluorophore is formed inside the barrel. (B) Schematic representation of a cDNA construct. Glutathione S-transferase (GST) was connected with N-GFP for affinity purification. An amino acid position dissected by PreScission Protease is indicated with an arrow. (C) SDS-PAGE analysis of proteins expressed in *E. coli* transformed with pGEX vector. Lane 1–3, Coomassie Blue-stained SDS-PAGE. Lane 1, protein molecular mass standards with their molecular masses (kDa) shown on the left; lane 2, crude extract before loaded onto a GST-affinity column; lane 3, a GST-affinity purified sample. Lanes 4 and 5, Western blot analysis of crude cell extracts using antibodies specific for VDE (lane 4) and GFP (lane 5). (D) Visual appearance of *E. coli* carrying pGEX (lower panels) and control vector (upper panels). The bacteria were streaked onto LB agarose, transmitted with white light (left panel), or illuminated with blue-LED at 470 nm (right panel). The fluorescence was detected with a cooled CCD camera through an emission band-path filter of 530 nm.

ciferase was split into the N- and C-terminal fragments, the bioluminescence activity was completely lost.[17] However, when DnaE was connected with the N- and C-terminal fragments of split luciferase, the two fragments were linked with a peptide bond by protein splicing, and its activity of bioluminescence was fully recovered. We have also recently reported reconstitution of split *renilla* luciferase by protein splicing.[18] Although the crystal structure of *renilla* luciferase has not been known until now, it has been clear that its N-terminal and several cysteine residues are important for its bioluminescence activity.[19] Based on the previous report, we found a pair of split *renilla* luciferase fragments, which were inactive until spliced together, allowed reconstitution of the luciferase by protein splicing.

An important point of this split GFP or luciferase reporter is that protein splicing generates an intact fluorescent or bioluminescent protein from two of its protein fragments in living cells without any cofactors or external enzymes. Therefore, the reporter can be used anywhere in the cells. Using the reconstitution of split GFP and luciferase reporters, we have developed novel methods with general applicability for detecting protein–protein interactions and protein localization in living cells and animals.

2. Methods for Detecting Protein–Protein Interactions

In living cells, protein–protein interactions constitute essential regulatory steps that modulate the activity of signaling pathways. Identification of the interactions and characterization of their physiological significance is one of the main goals of current research in different fields of biology. Towards this goal, several technologies have been developed for detecting protein–protein interactions without the need for disrupting living cells. The yeast two-hybrid assay pioneered the field of protein interactions,[20] which facilitated the identification of potential protein–protein interactions and was used as a method for the generation of protein interaction maps. The limitation of the two-hybrid technique is that activation of the transcription factors occurs only in the nucleus, so it is impossible to detect key molecules and steps of intracellular signaling outside the nucleus.[21] To overcome this limitation, we have developed a method for detecting protein–protein interactions based on reporter protein reconstitution of split GFP and split firefly luciferase.

The basic principle of detecting protein–protein interactions using the VDE intein is shown in Fig. 4.[13] An N-terminal fragment of VDE (N-VDE; 1–184 amino acids) is fused to N-terminal fragment of GFP (N-GFP; 1–128 amino acids), and the C-terminal fragment of VDE (C-VDE; 389–454 amino acids) is fused to the rest of GFP (C-GFP; 129–239 amino acids). Each of these fusion proteins is linked to a protein of interest (protein A) and its target (protein B). When interaction occurs

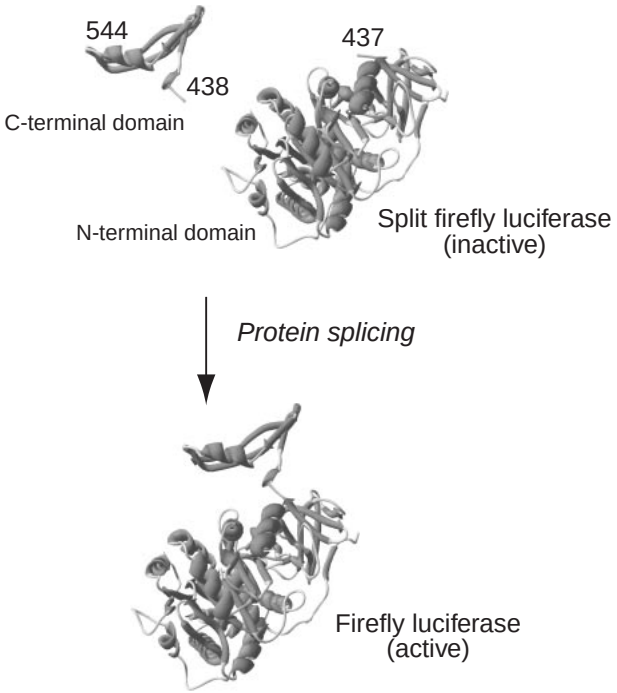

Fig. 3. Split firefly luciferase reconstitution by protein splicing. Split firefly luciferase is composed of N-terminal (1–437 amino acids) and C-terminal (438–544 amino acids) domains. Each domain does not possess bioluminescence activity. When the N- and C-terminal domains are connected together by protein splicing, its bioluminescence activity can be recovered in vivo.

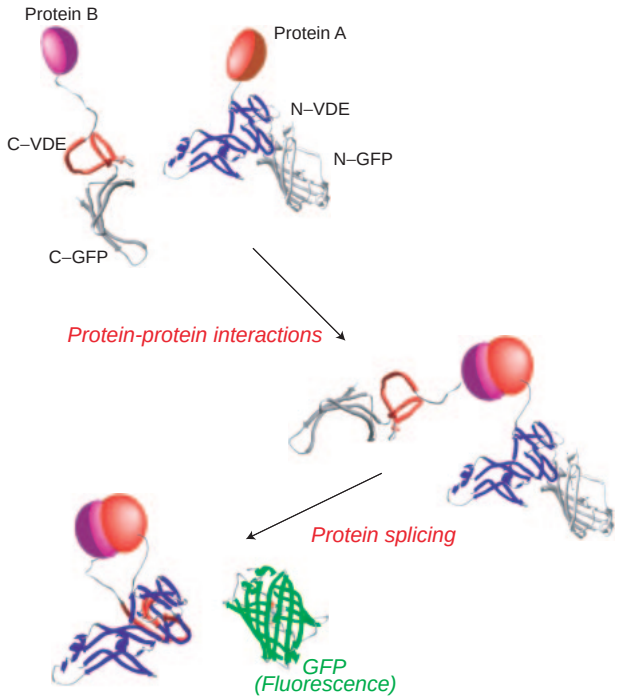

Fig. 4. Split EGFP as a probe for detecting protein–protein interactions. Ribbon diagrams of the N-terminal half of VDE (1–184 amino acids; blue) and the C-terminal half of VDE (389–454 amino acids; red) are each connected with the N-terminal half of GFP (1–128 amino acids; light gray) and C-terminal half of GFP (129–238 amino acids; dark gray), respectively. Interacted protein A and protein B are linked to opposite ends of the split VDE. Interaction between protein A and protein B induces the folding of N- and C-terminal halves of VDE and protein splicing occurs. The N- and C-terminal halves of GFP are linked together by a peptide bond to yield fluorescent GFP.

between the two proteins, the N- and C-terminal fragments of VDE are brought in close proximity and undergo correct folding, which induces a splicing event. The N- and C-terminal fragments of GFP directly link to each other by a peptide bond, and thereby the matured GFP forms its fluorophore with its emission maximum at 510 nm. The extent of the protein–protein interaction can be evaluated by measuring the magnitude of fluorescence intensity originated from the formation of GFP. As a proof of this principle, calmodulin (CaM) and its target peptide (M13) were used as an interaction pair. The interaction of CaM and M13 in *E. coli* resulted in the formation of fluorescent GFP.

To improve the splicing efficiency, we replaced the VDE intein with the DnaE intein because of special features of DnaE: (I) DnaE is composed of a total of 159 amino acids, which is shorter than VDE by 90 amino acids;[12] (II) The solubility of DnaE is higher than that of VDE, which is known to make an inclusion body upon expression in *E. coli*;[8] (III) For efficient splicing to occur, the requirement for amino acid sequences of DnaE around the splicing junctions is less stringent than for VDE.[22,23] We found that enhanced GFP (EGFP) dissected at the position between 157 and 158 from the N-terminus showed strong fluorescence intensity upon protein interactions.[24] A time required for the formation of EGFP after protein interactions was only 4 hours, as compared to 3 days with the VDE intein. To demonstrate the screening procedure with this probe, we mixed plasmids pETm157(/) and pETm157(C/M) in a molar ratio of 1:1 and cotransformed this mixture in *E. coli* (Fig. 5). The transformants were plated on LB agar plates. Around 50% of the colonies exhibited strong fluorescence, and the rest of the colonies fluoresced weakly. We confirmed that the strong fluoresced colonies included the cDNAs encoding a pair of interacting proteins. Thus, we demonstrated that this split GFP reporter can be used for rapid bacterial screening and selection to look for interacting proteins. Several bacterial one- and two-hybrid systems have been proposed earlier.[25,26] These include a common principle: when the proteins interact, they trigger a transcriptional activation of a reporter gene and produce a signal protein that is accumulated in the bacteria. The protein interactions were thereby observed in bacterial colonies on LB agar plates. Unlike the bacterial two-hybrid systems, the split-GFP system involves the reconstitution of GFP, and does not require that the protein–protein interactions take place near the cell nucleus or that an enzyme substrate be present.

We have further applied this split GFP reconstitution system to the screening of endocrine disrupting chemicals (EDCs) using living eukaryotic cells.[27] As an example, the interaction between the androgen receptor (AR) and cSrc is demonstrated. The male sex hormone androgen, dihydrotestosterone (DHT), has been known to bind the AR in the cytosol and to induce

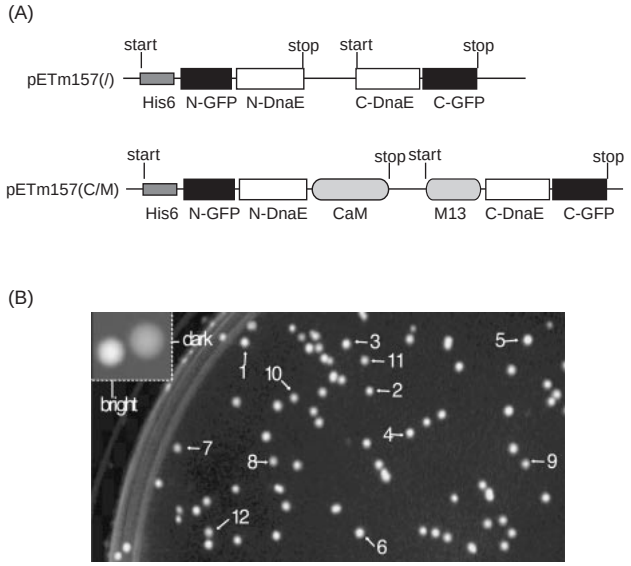

Fig. 5. Bacterial screening of interacting proteins. (A) Schematic structures of cDNA constructs. His6 indicates polyhistidine tag. CaM is *Xenopus* calmodulin and M13 is the CaM-binding peptide derived from skeletal muscle myosin light-chain kinase. (B) Fluorescence images of bacterial colonies on LB agar plates. BL21(DE3)pLysS cells were cotransformed with a mixture of plasmids, pETm157(/) and pETm157(C/M); plated on LB agar plates containing 1.0 mM IPTG, 0.1 mM ampicillin, and 0.1 mM chloramphenicol; and incubated for 16 h at 37 °C. Inset: An expanded image of bright and dark bacterial colonies on LB agar plates.

cell proliferation.[28] Furthermore, DHT increased the kinase activity of cSrc located adjacent to the cell membrane in the cytosol of prostate cancer cells. On the bases of these earlier findings, the N- and C-terminal halves of the split GFP reporter were attached to cSrc and AR, respectively, and expressed in NIH3T3 cells. Upon stimulation of the cells with DHT, the cells emitted fluorescence of GFP, indicating that DHT induced the interaction between cSrc and AR inside the cell membrane and that GFP was generated as a result. In addition, flutamide, an AR antagonist, decreased the number of fluorescent cells, indicating that cSrc-AR interaction was inhibited by flutamide. Thus, the split GFP reconstitution system enabled detection of particular intracellular protein–protein interactions nondestructively in living eukaryotic cells.

In order to extend the basic concept of the split reporter proteins for wider applications, we have developed split firefly luciferase as an optical probe for detecting protein–protein interactions.[17] Amino terminal DnaE is fused to an N-terminal half of luciferase (1–437 amino acids) and C-terminal DnaE to the rest of luciferase (438–544 amino acids). Interacted proteins are linked to opposite ends of those DnaE. When a specific protein interaction occurs, the N- and C-terminus of DnaE are brought into close proximity and undergo correct folding, which induces protein splicing. The N- and C-terminal halves of luciferase are linked together by a peptide bond, and the matured luciferase thus formed creates an active center to emit light, which can be detected by a luminometer. To demonstrate the applicability of this split luciferase reconstitution, we used insulin-induced interaction between known binding partners, phosphorylated Y941 peptide derived from insulin receptor substrate 1 (IRS-1) and its target N-terminal SH2 domain of PI 3-kinase. We found that the firefly luciferase activity obtained with the stimulation of insulin for more than 3 h was 4 times higher than the background signal (Table 1a and 1b). The magnitude of the luminescent intensity increased with increasing concentrations of insulin from 1.0×10^{-10} to 1.0×10^{-7} M ($1\ \mathrm{M} = 1\ \mathrm{mol\,dm^{-3}}$; Table 1c). Upon expression of the Y941 mutant that is not subjected to phosphorylation by the insulin receptor, the magnitude of the normalized luminescent intensity was the same as the background level, demonstrating that phosphorylation of Y941 peptide is indispensable for producing the luciferase enzyme by the protein splicing reaction. From these results, we concluded that enzymatic luciferase activity triggered by insulin served to monitor the interaction between IRS-1 and the SH2 domain in an insulin dose-dependent manner. While this report aimed at demonstrating the detection of phosphorylation in the insulin-signaling pathway, the method of split luciferase reconstitution has been used for noninvasively imaging MyoD-Id interaction in living mice.[29] There will be numerous other possible applications of this particular method using split luciferase reconstitution such as quantitative evaluation of interactions between membrane- or organelle-associated proteins in different signaling pathways in living cells and animals.

These split GFP and split luciferase approaches have an advantage in that protein interactions can be monitored anywhere in the cells, whereas the traditional two-hybrid approaches are limited to interactions only in the nucleus. In addition, GFP or luciferase accumulates in a target cell until it degrades, and information of the interaction is thereby integrated in the cell. Both split GFP and luciferase reconstitution systems provide rational methods to identify and characterize protein–protein interactions in living cells, as exemplified in differing cell signaling pathways. Imaging interacting protein pairs in living subjects may pave the way to functional proteomics in whole animals and provide a new tool for evaluating new pharmaceuticals targeted to modulate protein–protein interactions.

3. Methods for Identifying Organelle-Localized Proteins

One of the most distinct features of eukaryotic cells, especially mammalian and plant cells, is the compartmentalization of each protein. Protein localization is tightly bound to function, such that preferential localization of a protein is often an essential step determining its function. Therefore, functional assays aimed at characterizing the cellular localization of proteins are very important for understanding complicated protein networks. The technique for identifying proteins localized in organelles largely relies on the isolation of compartments through cell fractionation and electrophoresis, combined with mass spectrometry.[30–32] This biochemical method 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 isolate.[33]

To overcome these drawbacks, we have developed a method with a general applicability for identifying mitochondrial proteins from large-scale cDNA libraries.[34] The basic strategy is

Table 1.

a. Time dependent increases in protein splicing events upon addition of insulin

	RLU	
Time	Insulin (100 nM)	Insulin (0 M)
5 min	0.17 ± 0.10	—
3 h	0.73 ± 0.26	0.15 ± 0.04
72 h	1.00 ± 0.28	—

b. Effect of a point mutation in Y941 peptide on the protein splicing

	RLU	
	Insulin (100 nM)	Insulin (0 M)
Y941 peptide	0.50 ± 0.95	0.11 ± 0.04
Y941A mutant	0.16 ± 0.05	0.15 ± 0.05

c. Insulin concentration dependence on RLU

	Concentration (M)	RLU
Y941 peptide	1.0×10^{-12}	0.13 ± 0.07
	1.0×10^{-11}	0.14 ± 0.03
	1.0×10^{-10}	0.17 ± 0.06
	1.0×10^{-9}	0.27 ± 0.06
	1.0×10^{-8}	0.35 ± 0.09
	1.0×10^{-7}	0.50 ± 0.10

a) CHO-HIR cells were cultured in 6-well plates and were transfected with 2 μg of the plasmid, pIRES-DSL(Y/S), and 0.02 μg of control plasmid, pRL-TK. After incubation for 45 h, the cells were subjected to be stimulated with 100 nM of insulin for the indicated time at 37 °C. b), c) CHO-HIR cells were cultured in 6-well plates and were transfected with 1 μg of pIRES-DSL(Y/S) or its mutant together with 0.02 ng of pRL-TK. After incubation for 45 h, the cells were subjected to be stimulated with or without each concentration of insulin for 3 h and were allowed to measure its luminescence.

based on reconstitution of EGFP by protein splicing with DnaE intein (Fig. 6A). A characteristic feature of DnaE used in this method is to have natural splicing ability to ligate accompanying exteins *in trans*. A tandem fusion protein, containing a mitochondrial targeting signal (MTS) and C-terminal fragments of DnaE and EGFP, localizes in the mitochondrial matrix in mammalian cells. cDNA libraries generated from mRNAs are genetically fused to the sequences encoding the N-terminal fragments of EGFP and DnaE. If test proteins expressed from the cDNA libraries contain a functional MTS, the fusion products translocate into the mitochondrial matrix, and 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. This method has the advantage that only "mitochondria-positive" cloned cells yield a fluorescence signal. These fluorescent cells are screened rapidly and collected by a fluorescence-activated cell sorter (FACS). Relevant genes are subsequently identified by genetic PCR and their DNA sequences.

To verify that EGFP formation by protein splicing occurs exclusively for the proteins with MTSs, we tested two representative polypeptides; one is a well-characterized cytosolic protein, calmodulin, and the other is an MTS polypeptide derived from cytochrome *C* oxidase. Western blots of the cells showed that protein splicing occurred exclusively in the cells expressing MTS-tagged N-terminal probes to produce ligated EGFP (Fig. 7B). The localization of EGFP was found to be virtually the same as the localization of mitochondria stained with a cell-permeable mitochondrion-selective dye, tetramethylrhodamine ethyl ester. Moreover, it was evident that the fluorescence intensity of the cells harboring MTS-tagged N-terminal probe was 70-fold higher than the background fluorescence. Thus, the cells that include proteins with MTSs can be separated from the ones without MTSs and collected, and the fluorescence intensity of the EGFP reconstituted in mitochondria is strong enough to be detected by FACS analysis.

Next, we tested whether genes encoding mitochondrial proteins could selectively be isolated from large cDNA libraries. The cDNA libraries which were constructed from mouse liver cells contained 1.1×10^6 independent clones with the size of cDNAs averaging 1.4 kbp. The libraries were converted to retroviruses by using a high-titer retrovirus packaging cell line, Plat-E (Fig. 6C). After infection of the retrovirus to BNL1MEmito cells, in which a tandem fusion protein of C-terminal halves of DnaE and EGFP were expressed in the mitochondrial matrix, the population of the fluorescence cells was measured by FACS. The percentage of the fluorescent cells was found to be 0.1% of the total cells. We collected these fluorescent cells by FACS, and analyzed the subcellular localization of each isolated clone. If the cDNAs were integrated in the host genome, the corresponding proteins should be expressed constitutively in the BNL1MEmito cells and therefore EGFP reconstitution would take place in the mitochondria. The EGFP were localized exclusively in mitochondria, so cDNAs integrated in the genome were analyzed by a DNA sequencer. Thirty mitochondrial proteins including 3 novel proteins were identified as mitochondrial (Table 2).

Our results demonstrate that this method provides a high-throughput approach to identify novel gene products that are compartmentalized in the mitochondrial matrix and inner membrane. This method should allow the identification of groups of proteins localized not only in mitochondria but also

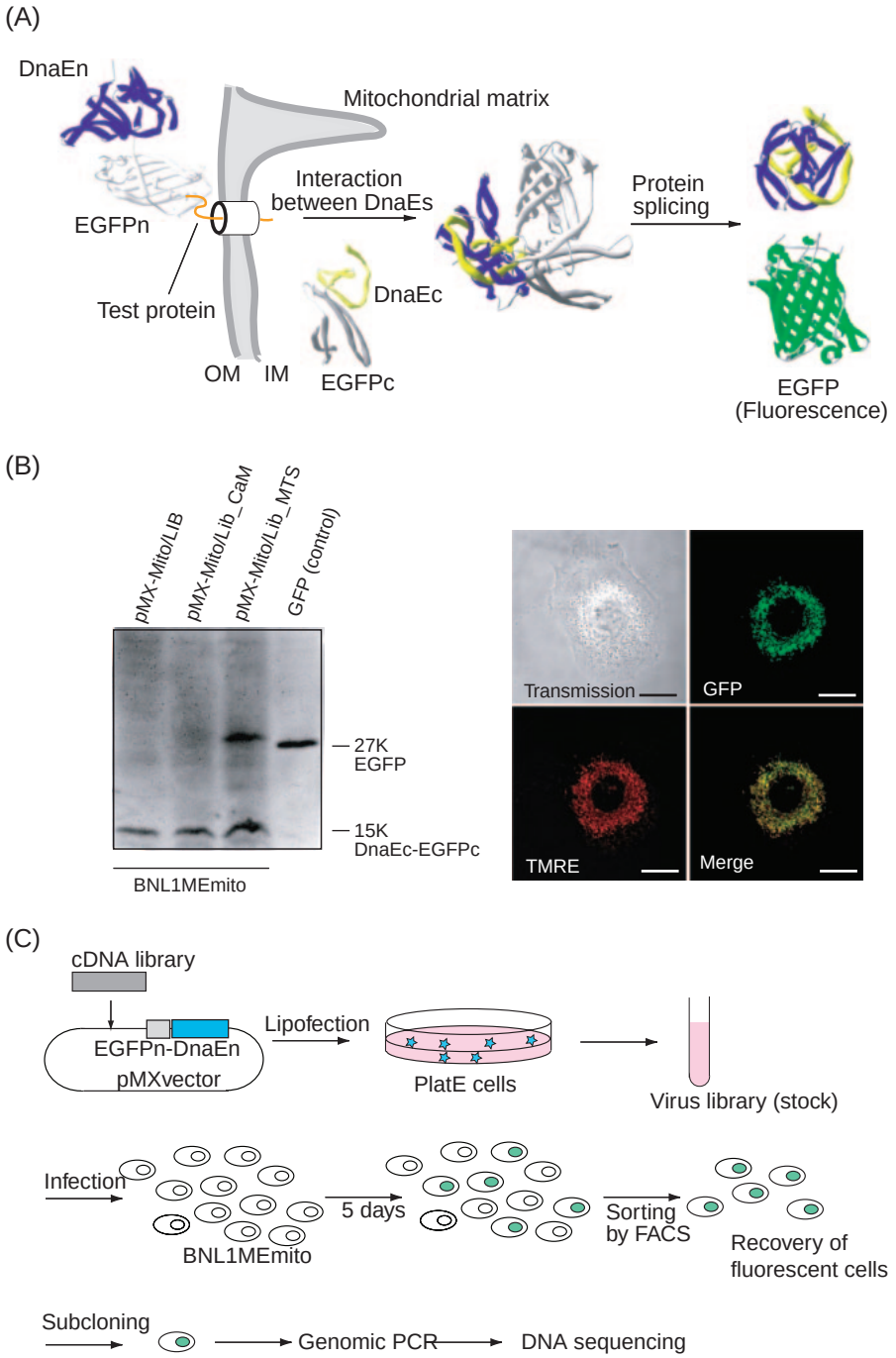

Fig. 6. Split EGFP system for identifying mitochondrial proteins from cDNA libraries. (A) Scheme showing how EGFP is formed by protein splicing of DnaEs when a test protein is localized in the mitochondrial matrix. Three-dimensional structures colored blue and yellow represent DnaEn and DnaEc, while light and dark gray represent N- and C-terminal halves of EGFP, respectively. Orange strand means a test protein. OM and IM are mitochondrial outer and inner membranes, respectively. (B) Selective and sensitive detection of mitochondrial proteins. Left: Whole cell lysates from mouse liver cells expressing CaM-tagged or MTS-tagged N-terminal probe were western blotted with a monoclonal antibody specific to C-terminal half of EGFP. BNL1MEmito indicates BNL1ME cells derived from mouse liver that harboring DnaEc-EGFPc fusion protein in mitochondria. Right: Expression of the MTS-tagged N-terminal probe and localization of reconstituted EGFP to mitochondria. BNL1MEmito cells infected with retroviruses of pMX-Mito/Lib-MTS were cultured for 2 d, and the cells were spread on a glass dish. The cells were imaged live (transmission) and fluorescence of EGFP recorded by a confocal microscope (GFP). After the image was taken, mitochondria in the live cell were stained with tetramethylrhodamine ethyl ester (TMRE). Bar, 10 μm. (C) Strategy for identifying mitochondrial proteins from large-scale cDNA libraries. cDNAs are inserted into pMX vector and transfected into packaging cell line, PlatE cells. After converting the library to retroviruses, BNL1MEmito cells were infected to meet less than 30% infection efficiency. Fluorescent cells were collected by FACS 5 d after infection, cloned and subjected for further analysis.

Table 2. Category of Identified cDNAs

Description
Identical to mouse mitochondrial protein
ATP synthetase H^+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1 (Atp5a1)
Glutaryl-coenzyme A dehydrogenase (Gcdh)
Malate dehydrogenase, mitochondrial (Mor1)
Acetyl-coenzyme A dehydrogenase, long chain (Acad1)
Cytochrome c oxidase, subunit Vb (Cox5b)
ATP synthetase H^+ transporting, mitochondrial F1 complex, beta subunit (Atp5b)
ATP synthetase H^+ transporting, mitochondrial F1 complex, gamma polypeptide 1 (Atp5c1)
Mitochondrial gene for cytochrome b
Uncoupling protein 2, mitochondrial (Ucp2)
Aldehyde dehydrogenase 2, mitochondrial (Aldh2)
Mitochondrial ribosomal protein S11 (Mrps11)
Identical to mouse protein
Ribosomal protein S18 (Rsp18)
Stomatin-like protein 2
Similar to mouse gene
Phosphoenolpyruvate carboxykinase 2
NADH-ubiquinone oxidoreductase 13 kDa-A subunit, mitochondrial precursor (CI-13KD-A)
Inorganic phosphatase
60S ribosomal protein L3 (L4)
Homologue to mammal gene
Heat shock protein 75 (*Homo sapiens*, 89%)
NADH-ubiquinone oxidoreductase 30 kDa subunit precursor (*Homo sapiens*, 88%)
Putative cytochrome c oxidase assembly protein (*Schizosaccharomyces pombe*, 23%)
Mitochondrial 28S ribosomal protein S18-1 (*Homo sapiens*, 77%)
Succinate dehydrogenase complex, subunit B, iron sulfur (*Homo sapiens*, 91%)
Biphenyl hydrolase-related protein (*Homo sapiens*, 75%)
Ubiquinol-cytochrome C reductase complex core protein 2, (complex III subunit II) (*Rattus norbegi*, 93%)
Predicted protein
GI: 12859851
GI: 12840016
GI: 12852607

Description indicates the results of BLASTn and BLASTx searches of the daily updated GenBank entries at NCBI (National Center for Biotechnology Information) as of 10 August 2002. A description or accession number (GI) is given for the most significant matches. When the clone was identical to mouse gene and the gene product was known to be localized in mitochondria, it was categolized in "Identical to mouse mitochondrial protein". The clones identical mouse known proteins were in "Identical to mouse protein". The clones identical to mouse gene and similar to mouse known proteins were annotated as "Similar to mouse gene". The clones identical to mouse gene and similar to other species were "Homologue to mammal gene". The clones identical to mouse gene but having no similarity to mammal genes were classified in "Predicted protein".

in the nucleus, endoplasmic reticulum, golgi bodies, and peroxisome, by replacing the MTS attached to the C-terminal probe with the targeting signals of each. In addition, when combined with cDNA subtraction methods, the method will provide greater flexibility to compare, for example, expressed genes found in healthy and disease states or in different tissues.

4. Detection of Protein Nuclear Transport

Dynamics of the protein movement inside living cells is an important event in eukaryotic cells. A typical example is nuclear transport of a particular protein, which plays a key role in regulating gene expressions in response to extracellular signals. A technique for monitoring the dynamics of the protein movement inside living cells relies on the use of immunocytochemistry or optical imaging with genetically tagged GFP. These analyses are effective for imaging spatial and temporal dynamics of proteins of interest within single living cells. For high-throughput analysis of the protein movement inside the

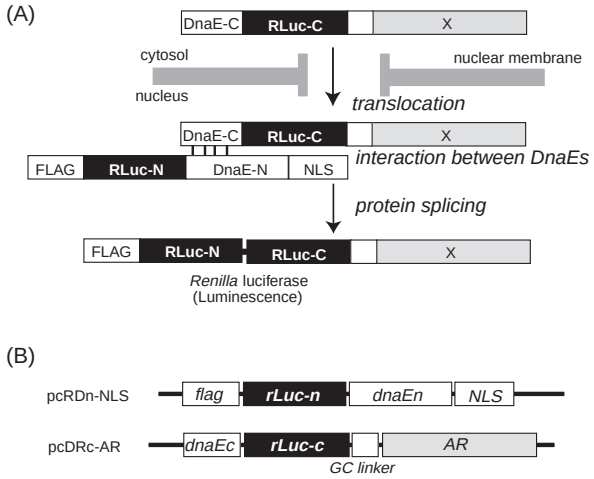

Fig. 7. Principle for monitoring translocation of a particular protein (X) into the nucleus using protein splicing of split *renilla* luciferase (Rluc). The N-terminal half of Rluc (RLuc-N; 1–229 amino acids) is connected with the N-terminal half of DnaE (DnaE-N; 1–123 amino acids) and nuclear localization signal (NLS; $(DPKKKRKV)_3$), which is predominantly localized in the nucleus. The C-terminal half of DnaE (DnaE-C; 1–36 amino acids) is connected with the C-terminal half of Rluc (RLuc-C; 230–311 amino acids) and a protein X, which is localized in the cytosol. When the tandem fusion protein consisting of DnaE-C, Rluc-C and AR translocates into the nucleus, DnaE-C interacts with DnaE-N, and protein splicing results. The Rluc-N and Rluc-C are linked together by a peptide bond, and the reconstituted Rluc recovers its bioluminescence activity. FLAG means epitope (DYKDDDDK). (B) Schematic structures of cDNA constructs. All the Italics in the figure mean genes of their corresponding proteins. *GC linker*, cDNA sequence of amino acids GGGGSG.

cells, automated fluorescence microscopy has been developed. Although such technological progress is important, image acquisition can be slow and tedious. In particular, algorithms to automatically determine nuclear vs cytoplasmic localization in the acquired image still remain imprecise and slow. 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.[35]

In order to solve these problems, we have developed a bioluminescent reporter for detecting transport into the nucleus of a particular protein using a *renilla* luciferase reconstitution system.[18] *Renilla* luciferase has desirable features for a monomeric protein: small size (36 kDa), strong luminescence, and ATP is not necessary for its activity.[36,37] In addition, its substrate, coelenterate luciferin (coelenterazine) rapidly penetrates through cell membranes, which is suitable for in vivo imaging.[38] The luciferase is split into N- and C-terminal fragments, and they are connected with N- and C-terminal fragments of DnaEs, respectively (Fig. 7A). The C-terminal fragment is permanently localized 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-terminal *renilla* luciferase meets the C-terminal *renilla* luciferase in the nucleus, and full-length *renilla* luciferase is reconstituted by protein splicing. In order to demonstrate the usefulness of the indicator, a well-known nuclear receptor, androgen receptor (AR), was examined; it translocates from the cytosol into the nucleus upon binding to 5α-dihydrotestosterone (DHT).

To ensure that the AR protein fused to the C-terminal half of *renilla* luciferase translocates from the cytosol to the nucleus upon addition of DHT and that the N-terminal half of the luciferase permanently resides in the nucleus, we constructed cDNAs of a fusion protein shown in Fig. 7B and examined the localization of the expressed protein in COS-7 cells by immunocytochemistry. In the absence of DHT, the AR-fusion protein was predominantly in cytoplasm, whereas addition of DHT resulted in the nuclear localization of the fusion protein (Fig. 8A). In contrast, the N-terminal fragments of *renilla* luciferase permanently localized in the nucleus both in the presence and absence of DHT. The results of western blot showed that, in the presence of DHT, the AR antibody recognized 115 kDa of an unspliced precursor and 139 kDa of a polypeptide, of which the electrophoretic mobility was consistent with the predicted size of the product after protein splicing (Fig. 8B). The luminescence intensity from the nucleus in the presence of DHT was found to be much higher than that of cytosol (Fig. 8C). Given all these results, we concluded that protein splicing occurred and that the split *renilla* luciferase was reconstituted when the DHT-bound AR was translocated into the nucleus.

Next, we showed that the *renilla* luciferase reconstitution strategy works for quantitative analysis of the extent of AR translocation 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 (Fig. 9A). Several endogenous hormones and synthetic chemicals were tested. Endogenous androgens such as testosterone and 19-nortestosterone induced high luminescence intensities, while 17β-estradiol, progesterone, and a synthetic steroid hormone of cyproterone acetate (CPA) gave slight increases in the observed luminescence intensities. Also, procymidone and a PCB congener (Aroclor 1254) were found to inhibit the DHT-induced strong luminescence. The results demonstrate that this assay system can be used for quantitative analysis of the extent of AR translocation into the nucleus.

For further extending the usefulness of the split *renilla* luciferase reporter, we have developed a technique for locating the distribution of chemical compounds in the organs of living mice and quantifying their effects on AR translocation into the nucleus. The COS-7 cells harboring both split reporters were implanted in the backs of mice. One group of mice were stimulated with 1.0% DMSO (vehicle) for 2 h, whereas the other group of mice were stimulated with DHT for the same period. Luminescence intensities obtained from the mice with DHT stimulation were 3.5 times higher than that with vehicle stimulation (Fig. 9B). The results demonstrate that it is possible to image and quantitatively evaluate the difference in the extent of AR translocation in living mice in the presence of DHT relative to its absence.

Finally, we studied the usefulness of this strategy for examining the effects of inhibitors on AR translocation into the nu-

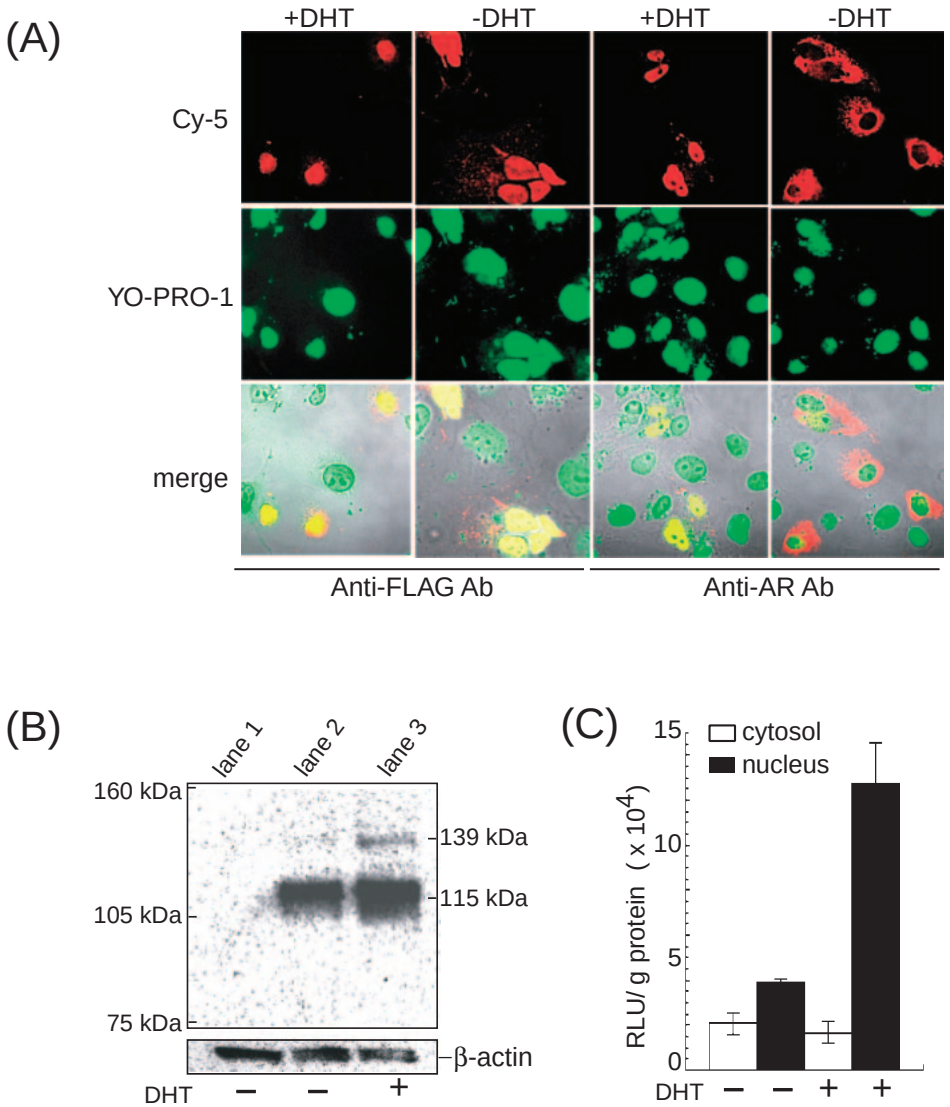

Fig. 8. Characterization of the indicators in vitro. (A) The immunocytochemical images of COS-7 cells transiently transfected with pcRDn-NLS or pcDRc-AR. The COS-7 cells were cultured for 36 h, and then the cells were incubated for 2 h in the absence or presence of 1 μM DHT. The two expressed proteins were recognized by anti-AR and anti-FLAG antibodies, respectively, and stained with Cy-5-labeled secondary antibody (top images). The middle images show the nuclei stained with YO-PRO-1, and their merged images are shown with the transmission (bottom images). (B) Western blot of protein extracts from COS-7 cells (lane 1) and from the cells cotransfected with pcRDn-NLS and pcDRc-AR in the absence (lane 2) or presence (lane 3) of 1 μM DHT. As a reference for the amounts of the electrophoresed proteins, β-actin was stained with its specific antibody. (C) Quantitative analysis of the Rluc activity for nuclear and cytoplasmic fractions. The cellular fractions were obtained from cotransfected cells with pcRDn-NLS and pcDRc-AR in the absence or presence of 1 μM DHT ($n = 3$).

cleus in the mouse brain. The COS-7 cells harboring both split reporters were implanted in the mouse brain at a depth of 3 mm. Induction of AR nuclear transport by DHT resulted in a significant increase in photon count from the brains as compared to the value obtained with vehicle (Fig. 9C). The stimulation with DHT together with procymidone and PCB, respectively, resulted in a decrease in luminescence intensity, indicating that procymidone and PCB had an ability to pass through the blood-brain barrier and hindered nuclear import of AR.

A remarkable point of this method is that it allowed for determination of the subcellular localization of AR by the luminescence signals generated only when the AR localized in the nucleus. AR remaining in the cytosol did not induce reconstitution of split *renilla* luciferase, and therefore no background luminescence was observed. Moreover, 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. Thus, this probe enabled background-free, accurate, precise, and sensitive detection, which is of great advantage to quantitatively evaluate the extent of protein nuclear transport in a high-throughput manner. Some specific xeno-chemicals or drugs can be non-invasively monitored as to their effects on a particular protein translocation into the nucleus in a tissue specific manner.

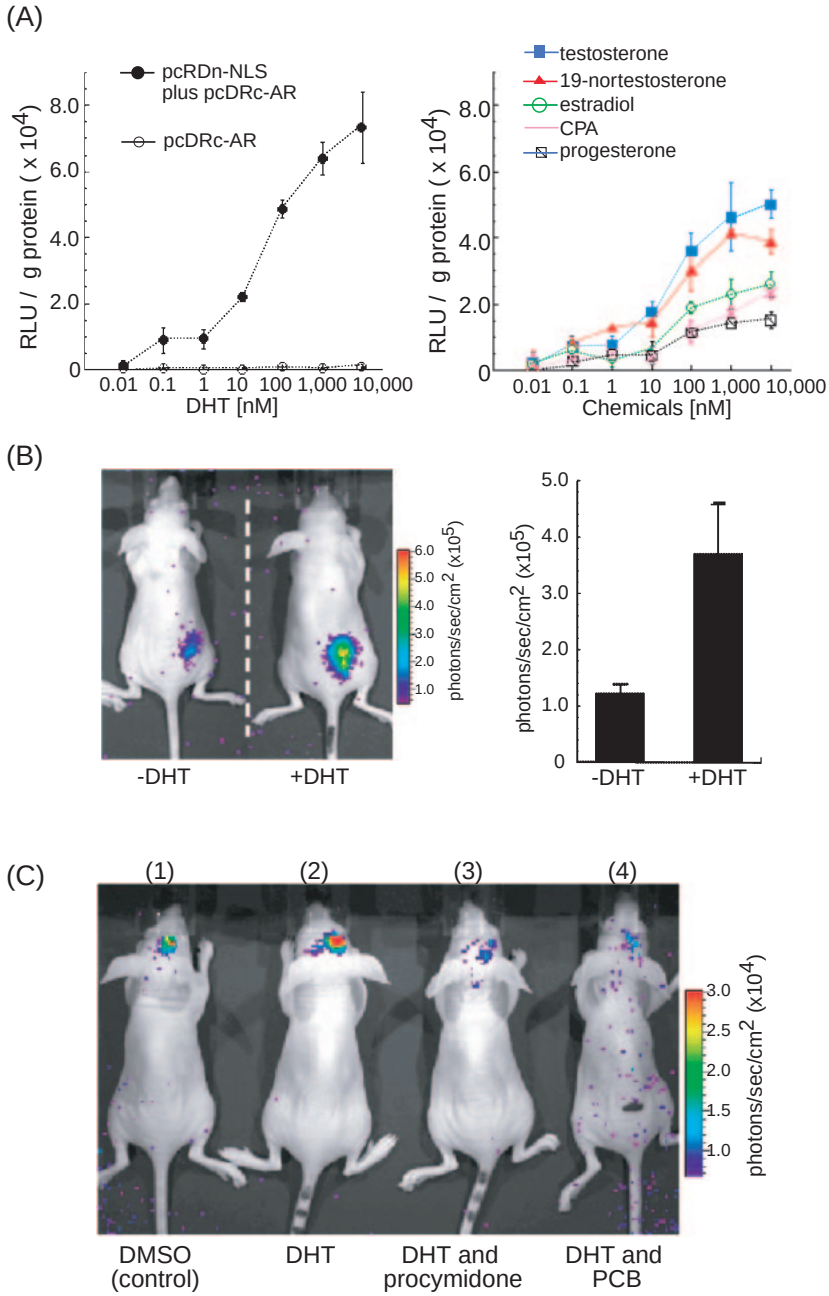

Fig. 9. Quantitative analysis of AR translocation for various chemical compounds. (A) Left: Dose-response curves for DHT based on the luminescence intensity of the reconstituted Rluc. The COS-7 cells were transiently transfected with pcDRc-AR, or cotransfected with pcRDn-NLS and pcDRc-AR, and Rluc activities were tested. The mean luminescence intensities ($n = 3$) were determined at each DHT concentration. Right: Dose-response curves for steroid hormones based on the luminescence intensity of Rluc. The COS-7 cells were cotransfected with pcRDn-NLS and pcDRc-AR, and Rluc activities were tested upon addition of each hormone ($n = 3$). (B) DHT-dependent translocation of AR in living mice. Left: In vivo optical CCD imaging of mice carrying transiently transfected COS-7 cells in the presence or absence of DHT (100 μg/kg body weight). The mice were imaged after *s.c.* implantation of COS-7 cells transiently cotransfected with pcRDn-NLS and pcDRc-AR. One group of mice was injected with DHT and the other group was left uninjected. Right: The graph shows the observed photon counts for uninjected and DHT-injected group of mice with mean values over four mice, respectively. (C) An inhibitory effect of procymidone or PCB on the bioluminescence developed by DHT (10 μg/kg body weight) in the living mice. The COS-7 cells transiently cotransfected with pcRDn-NLS and pcDRc-AR were implanted in the forebrain of the nude mice. Of mouse group (1) to (4), group (1) and group (2) were stimulated with 1% DMSO, while group (3) and group (4) were stimulated with procymidone (10 mg/kg body weight) and PCB (10 mg/kg body weight), respectively. Two hours after the stimulation, mouse group (2) to (4) were then stimulated with DHT (10 μg/kg body weight). Two hours after DHT stimulation, the mice were imaged in two-minute intervals until reaching the maximum photon counts after *i.c.* injection of coelenterazine (1.4 mg/kg body weight).

5. Concluding Remarks

Novel reporters of split GFP and luciferase proteins have been developed with general applicability to detect protein–protein interactions and protein localization in living cells and animals, as exemplified in differing cell signaling pathways. The concept of the reporters is based on reconstitution of the reporter protein from its fragments by protein splicing. Reconstitution of the split GFP upon protein–protein interactions enabled the bacterial screening and selection to look for interacting proteins. The use of split firefly luciferase together with DnaE provided selective and quantitative detection of insulin that led to the phosphorylation and protein interactions relevant to the insulin signaling pathways. The usefulness of the split GFP reporter was further demonstrated as the genetic method for identifying mitochondrial proteins. Finally, we described the feasibility of the split *renilla* luciferase reconstitution for monitoring the nuclear trafficking of target proteins in living cells and animals.

One of the advantages of the use of split GFP reporter is that fluorescence is an attractive readout for rapid and high-throughput approaches because of the availability of technologies such as FACS and multi-well plate readers. Single-cell imaging with the split GFP reporter under fluorescence microscopy will also allow the investigation of potentially interesting intracellular events in living cells. In contrast, split firefly and *renilla* luciferase reporters offer distinct advantages when compared with the split-GFP reporter: the split luciferase reporters have the potential to non-invasively image protein–protein interactions or protein translocations in living subjects owing to the background-free and high-sensitive detection. In addition, bioluminescence detection of intracellular molecular events is quantitative rather than qualitative, and its dynamic range is much wider than the one of GFP detection. The limitation of the split GFP and luciferase reporters is that it takes about three hours to obtain fluorescent and bioluminescent signals generated by the protein splicing reaction. During this reaction, the full-length reporter thus generated may be in part degenerated by proteasome, and thereby the quantitative analysis of the intracellular target would be hampered. Despite such a potential limitation, the application of the split reporters based on protein splicing will still continue to provide exciting new insights into the fields of chemistry and biology. Further understanding of protein splicing mechanisms, discovery of new inteins, and development of new split reporters, will certainly expand the basic tools with protein splicing technology. Development of a variety of attractive applications with the protein splicing technology will increase in the near future.

I would like to express my sincere gratitude to Professor Yoshio Umezawa for his guidance, discussion, and careful reading of this manuscript, and to all collaborators in Umezawa laboratory at the University of Tokyo for their efforts in contributing to these studies.

References

1 J. Zhang, R. E. Campbell, A. Y. Ting, and R. Y. Tsien, *Nat. Rev. Mol. Cell Biol.*, **3**, 906 (2002).
2 S. Grimm, *Nat. Rev. Genet.*, **5**, 179 (2004).
3 T. F. Massoud and S. S. Gambhir, *Genes Dev.*, **17**, 545 (2003).
4 M. Zimmer, *Chem. Rev.*, **102**, 759 (2002).
5 T. Wilson and J. W. Hastings, *Annu. Rev. Cell Dev. Biol.*, **14**, 197 (1998).
6 P. M. Kane, C. T. Yamashiro, D. F. Wolczyk, N. Neff, M. Goebl, and T. H. Stevens, *Science*, **250**, 651 (1990).
7 R. Hirata, Y. Ohsumk, A. Nakano, H. Kawasaki, K. Suzuki, and Y. Anraku, *J. Biol. Chem.*, **265**, 6726 (1990).
8 M. Kawasaki, S.-I. Makino, H. Matsuzaka, Y. Satow, Y. Ohya, and Y. Anraku, *Biochem. Biophys. Res. Commun.*, **222**, 827 (1996).
9 F. B. Perler, *Nucleic Acids Res.*, **30**, 383 (2002).
10 T. J. T. Evans and M. Q. Xu, *Chem. Rev.*, **102**, 4869 (2002).
11 H. Paulus, *Annu. Rev. Biochem.*, **69**, 447 (2000).
12 H. Wu, Z. Hu, and X. Q. Liu, *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 9226 (1998).
13 T. Ozawa, S. Nogami, M. Sato, Y. Ohya, and Y. Umezawa, *Anal. Chem.*, **72**, 5151 (2000).
14 F. Yang, L. G. Moss, and G. N. Phillips, Jr., *Nat. Biotechnol.*, **14**, 1246 (1996).
15 M. Ormo, A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien, and S. J. Remington, *Science*, **273**, 1392 (1996).
16 E. Conti, N. P. Franks, and P. Brick, *Structure*, **4**, 287 (1996).
17 T. Ozawa, A. Kaihara, M. Sato, K. Tachihara, and Y. Umezawa, *Anal. Chem.*, **73**, 2516 (2001).
18 S. B. Kim, T. Ozawa, S. Watanabe, and Y. Umezawa, *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 11542 (2004).
19 J. Liu and A. Escher, *Gene*, **237**, 153 (1999).
20 S. Fields and O. Song, *Nature*, **340**, 245 (1989).
21 T. Ozawa and Y. Umezawa, *Supramol. Chem.*, **14**, 271 (2002).
22 M. Kawasaki, S. Nogami, Y. Satow, Y. Ohya, and Y. Anraku, *J. Biol. Chem.*, **272**, 15668 (1997).
23 T. C. Evans, Jr., D. Martin, R. Kally, D. Panne, L. Sun, I. Ghosh, L. Chen, J. Benner, X.-Q. Liu, and M.-Q. Xu, *J. Biol. Chem.*, **275**, 9091 (2000).
24 T. Ozawa, T. M. Takeuchi, A. Kaihara, M. Sato, and Y. Umezawa, *Anal. Chem.*, **73**, 5866 (2001).
25 G. Karimova, J. Pidoux, A. Ullmann, and D. Ladant, *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 5752 (1998).
26 J. K. Joung, E. I. Ramm, and C. O. Pabo, *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 7382 (2000).
27 Y. Umezawa, T. Ozawa, and M. Sato, *Bull. Chem. Soc. Jpn.*, **75**, 1423 (2002).
28 S. M. Singh, S. Gauthier, and F. Labrie, *Curr. Med. Chem.*, **7**, 211 (2000).
29 R. Paulmurugan, Y. Umezawa, and S. S. Gambhir, *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 15608 (2002).
30 S. W. Taylor, E. Fahy, B. Zhang, G. M. Glenn, D. E. Warnock, S. Wiley, A. N. Murphy, S. P. Gaucher, R. A. Capaldi, B. W. Gibson, and S. S. Ghosh, *Nat. Biotechnol.*, **21**, 281 (2003).
31 A. Sickmann, J. Reinders, Y. Wagner, C. Joppich, R. Zahedi, H. E. Meyer, B. Schönfisch, I. Perschil, A. Chacinska, B. Guiard, P. Rehling, N. Pfanner, and C. Meisinger, *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 13207 (2003).
32 V. K. Mootha, J. Bunkenborg, J. V. Olsen, M. Hjerrild, J. R. Wisniewski, E. Stahl, M. S. Bolouri, H. N. Ray, S. Sihag, M. Kamal, N. Patterson, E. S. Lander, and M. Mann, *Cell*, **115**,

629 (2003).
33 B. Westermann and W. Neupert, *Nat. Biotechnol.*, **21**, 239 (2003).
34 T. Ozawa, Y. Sako, M. Sato, T. Kitamura, and Y. Umezawa, *Nat. Biotechnol.*, **21**, 287 (2003).
35 R. Weissleder and V. Ntziachristos, *Nat. Med.*, **9**, 123 (2003).
36 J. C. Matthews, K. Hori, and M. J. Cormier, *Biochemistry*, **16**, 85 (1977).
37 W. W. Lorenz, R. O. McCann, M. Longiaru, and M. J. Cormier, *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 4438 (1991).
38 S. Bhaumik and S. S. Gambhir, *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 377 (2002).

Takeaki Ozawa was born in Tokyo in 1969. He graduated from the Department of Chemistry, Faculty of Science, University of Tokyo in 1993. He received the M.Sci. (1995), and the Ph.D. (1998) from the University of Tokyo. He was appointed as a research associate at School of Science, University of Tokyo in 1998, and promoted to a lecture in 2002. Since 2003, he has been a PRESTO researcher of Japan Science and Technology Agency (JST). His current researches have focused on the development of novel methods for imaging complex intracellular events and for quantifying biological molecules in living subjects.