Available online at www.sciencedirect.com
SCIENCE DIRECT®
Phytochemistry 65 (2004) 1549–1555
PHYTOCHEMISTRY
www.elsevier.com/locate/phytochem

Genome-scale, biochemical annotation method based on the wheat germ cell-free protein synthesis system

Tatsuya Sawasaki [a], Yoshinori Hasegawa [a], Ryo Morishita [a], Motoaki Seki [b], Kazuo Shinozaki [b], Yaeta Endo [a,*]

[a] *Cell-Free Science and Technology Research Center, The Venture Business Laboratory, Ehime University, Matsuyama 790-8577, Japan*
[b] *Plant Mutation Exploration Team, Plant Functional Genomics Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan*

Received 9 January 2004; received in revised form 31 March 2004

Abstract

Since the complete genomic DNA sequencing of various species, attention has turned to the structural properties, and functional characteristics of proteins. Current cell-free protein expression systems from eukaryotes are capable of synthesizing proteins with high speed and accuracy; however, the yields are low due to their instability over time. This report reviews the high-throughput, genome-scale biochemical annotation method based on the cell-free system prepared from wheat embryos. We first briefly reviewed our highly efficient and robust wheat germ cell-free protein synthesis system, and then showed an application of the system for materialization and characterization of genetic information taking a cDNA library of protein kinase from *Arabidopsis thaliana* as an example. The procedure consists of: (1) fusion of the gene-of-interest to a purification-tag, amplified by the split-primer PCR method; (2) transcription and purification of mRNA; (3) cell-free protein synthesis in the bilayer system using 96-well titer plate; (4) affinity purification and activity measurement. We took 439 cDNAs encoding kinases among 1064 genes annotated so far, and they were translated in parallel into protein. Subsequent assay revealed 207 products having autophosphorylation activity. Furthermore, seven proteins out of 26 calcium-dependent protein kinase genes tested did phosphorylate a synthetic peptide substrate in the presence of calcium ion, demonstrating that the translation products, retained their substrate specificity. The information on biochemical function of gene products accumulated should revolutionize our understanding of biology and fundamentally alter the practice of medicine and influence other industries as well.

Keywords: Wheat germ cell-free protein synthesis system; *Arabidopsis* cDNA library; Protein kinase; Autophosporylation; Kination of synthetic substrate

1. Introduction

With the sequencing of genomes of various species, scientists and technologists have turned their attention to the structure, properties and functional activities of proteins. However, rapid progress in the area of proteomics is premised on the following two requirements: (1) the availability of many proteins, and (2) the availability of sufficient amounts of proteins in naturally folded states. A cell-free translation system can synthesize large number of proteins with speed and accuracy approaching those of in vivo translation, and can also express the proteins which interfere with host cell physiology. However, such systems have generally been unstable and therefore inefficient (Henrich et al., 1982; Goff and Goldberg, 1987; Chrunyk et al., 1993; Kurland, 1982; Pavlov and Ehrenberg, 1996; Roberts and Paterson, 1973). Recently, we found that conventional wheat germ extracts contain the RNA *N*-glycosidase tritin and other translation inhibitors such as thionin, ribonucleases, deoxyribonucleases and proteases, and that these enzymes originate from the endosperm

* Corresponding author. Tel.: +81-89-927-9936; fax: +81-89-927-9941.
E-mail address: yendo@eng.ehime-u.ac.jp (Y. Endo).

0031-9422/$ - see front matter © 2004 Published by Elsevier Ltd.
doi:10.1016/j.phytochem.2004.04.023

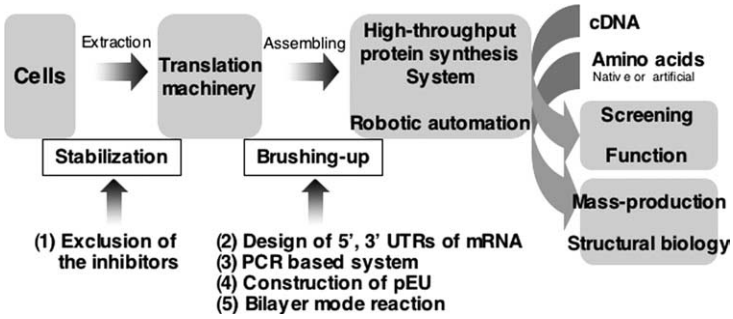

Fig. 1. Schematic diagram of development of the wheat germ cell-free protein synthesis system. A set of inhibitors of the translation machinery contaminated from endosperm was eliminated from embryos by washing (1). To bring the method up to practical use, other elemental technologies were brushed up. Those were design of 5′ and 3′ UTRs of mRNA (2), generation of transcription template based on newly developed "Split-primer PCR" (3), construction of the cell-free expression vector, pEU (4), and a bilayer reaction (5).

(Ogasawara et al., 1999; Madin et al., 2000; Sawasaki et al., 2002a,b). As illustrated in Fig. 1, our wheat germ extract, with careful elimination of endosperm contaminants, has enabled the realization of a cell-free protein synthesis system of high stability and activity (Madin et al., 2000). In addition, our system has several advantages over prokaryotic expression systems. Multi-domain proteins, found more often in eukaryotes, tend to fold into their correct conformations in eukaryotic translation systems much better than in prokaryotic ones (Kolb et al., 2000; Hartl and Hayer-Hartl, 2002). Another advantage of our system is that since it does not have significant contaminants of ribonucleases and proteases, high product quality and quantity can be achieved in comparison to the other cell-free systems. However, in order to adapt the cell-free system to address the high-throughput needs of modern proteomics, several critical improvements were needed. More recently, we improved system performance by examining the following critical design issues (Sawasaki et al., 2002b) as is illustrated in Fig. 1. They are: (1) optimization of the 5′- and 3′-untranslated regions (UTRs) of mRNA with a concomitant elimination of the 5′-^{7}mGpppG (cap) and poly(A)-tail (pA); (2) design of PCR primers to generate transcription templates directly from *Escherichia coli* cells carrying cDNA, thus bypassing the time-consuming cloning steps; (3) construction of a expression vector, named pEU, specialized for the massive production of proteins; (4) continuous translation reactions, named bilayer translation method, that is based on a diffusion principle (Sawasaki et al., 2002a). Assembling all of these elements, we could develop a cell-free system for high-throughput, genome-scale proteomics. Taking advantage of the ability of the cell-free system, namely efficiency, robustness, and bypasses of many of the time-consuming steps, we could bring it to a robotic automation. Here we reviewed the performance of our system when it was applied to biochemical characterization of the protein kinase gene family of *A. thaliana*.

2. Results

2.1. High-throughput protein synthesis and product quality

Fig. 2 shows the flow chart of cell-free expression process and of the kinase assay. The method begins with generation of DNA templates for the transcription by the "Split Primer" PCR (Sawasaki et al., 2002b). In this method, each gene was fused with the streptavidin (STA) gene as a purification tag on it's 3′ side, and to the SP6-promoter and a DNA fragment encoding the mRNA translation enhancer (omega) from the tobacco mosaic virus. After the transcription, parallel protein synthesis was carried out in 96-well microtiter plate based on bilayer reaction mode, each product was partially purified by biotin-coated affinity beads. Following

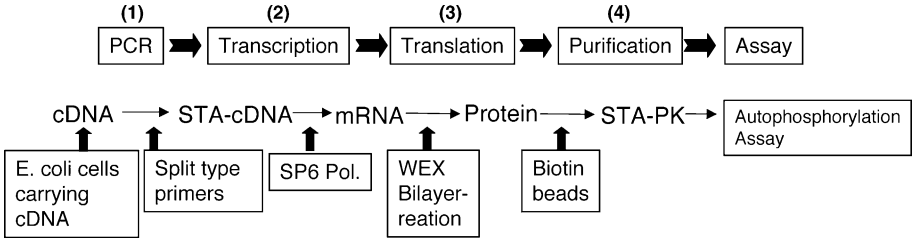

Fig. 2. Flow chart of the cell-free expression procedures for high-throughput biochemical annotation of genetic information. The method begins with generation of DNA templates generated by the "Split-primer" PCR (1), is then followed by in vitro transcription using phage coded SP6 RNA polymerase (2), translation in bilayer reaction (3), and purification of the product through the affinity tag (4). All of the steps were carried out in 96-well microtiter plates.

the phosphatase treatment of the eluates, samples were incubated with [γ-^{32}P]ATP and magnesium.

One of the most important requirement for any protein expression system is that the system can produce proteins in naturally folded state. It has been reported that the fundamental ability of eukaryotic translation machinery is to support effective co-translational protein folding on ribosomes (Kolb et al., 2000; Hartl and Hayer-Hartl, 2002). Thus the wheat germ cell-free system may provide suitable for the eukaryotic gene expression, especially for those multi-domain proteins.

We first addressed this point by synthesizing PHOT1 gene product from *Arabidopsis thaliana* (Christie et al., 1998; Briggs et al., 2001). The PHOT1 protein is a 120-kDa serine-threonine protein kinase localized in the plant plasma membranes to function as a photoreceptor for phototropism. The protein noncovalently binds to flavin mononucleotide (FMN) and has been suggested to act as a chromophore involved in light-dependent autophosphorylation. For the biochemical study the protein has been commonly expressed in a recombinant baculovirus system, in spite of the fact that only a small amount of the protein is localized in the soluble fraction but most of the fraction was insoluble (Christie et al., 1998). We therefore investigated whether PHOT1 protein synthesized in the cell-free system retains autophosphorylation activity in response to light irradiation. The mRNA encoding PHOT1 protein was transcribed from PCR generated DNA template, and was translated in the bilayer mode reaction either in the presence or absence of FMN in darkness. The productivity (20 μg/ml reaction) and the solubility (89–97%) were the same in the two cases (data not shown), however subsequent analysis revealed a dramatic difference on the light responsible activity (Fig. 3). Soluble protein samples were obtained and incubated with [γ-^{32}P]ATP for 15 min. Autoradiogram showed that PHOT1 protein synthesized in the presence of FMN was in fact highly autophosphorylated under the blue light irradiation compared to one which was kept under dim red light. In contrast, not significant levels of the light irradiation dependent autophosphorylation were observed among those proteins synthesized in the absence of FMN. These results demonstrated that (1) the 120-kDa PHOT1 protein probably a multi-domain protein could be synthesized in soluble form, and (2) FMN is an essential requirement during the folding into the holoenzyme. The latter supports the notion of co-translational folding taking place on ribosomes in our wheat germ cell-free system. We believe that the wheat germ cell-free system provide a powerful tool for materializing genetic information of eukaryotes.

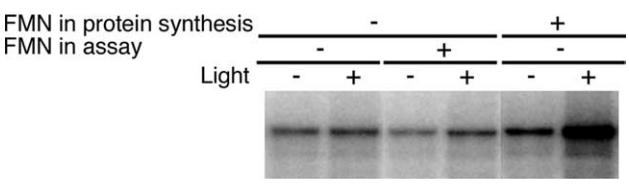

Fig. 3. Co-translational folding of multi-domain enzymes. Methods for protein synthesis and for analysis were done as described in the text.

2.2. Parallel biochemical characterization of kinase gene products from of Arabidopsis thaliana

To show the performance of our approach for high-throughput synthesis of gene products and biochemical characterization we carried out cell-free protein synthesis starting with *E. coli* cells carrying cDNAs (Sawasaki et al., 2002b). All pipetting steps were performed automatically by the synthesizer, named Protein Island Matsuyama (PIM), with standard 96-well microtiter plates for both transcription and translation reactions (Endo and Sawasaki, in press). We chose 439 genes encoding protein kinases of *A. thaliana* out of the 1064 annotated *A. thaliana* protein kinases listed in the PlantP database (http://plantsp.sdsc.edu/) (Table 1). PCR was done using the split-primer method, and the reaction was started with *E. coli* cells. After transcription in microtiter plates, samples were precipitated and washed with ethanol to remove dNTPs, NTPs and magnesium ions. The mRNAs in microtiter plates were directly dissolved in translation reaction mixture containing the substrate solution made of energy, amino acids and other essential chemical elements and were incubated at 26 °C for 17 h (Fig. 4). After incubation, samples were spun down at 30,000*g* for 15 min (some insoluble gene products were lost into the precipitate at this step). Supernatant recovered was mixed with biotin coated beads to isolate the product fused with STA. The partially purified proteins were treated with phosphatase and each sample was added histone H1 and myelin basic protein (MBP) as external phosphorylating marker and the mixture was incubated with [γ-^{32}P]ATP and magnesium. Aliquots of the samples were separated by SDS–PAGE, and protein bands were stained with CBB and were used for autoradiography. Fig. 4 showed some examples; a clearly visible stained protein band (asterisks in Fig. 4(a)) among those of endogenous wheat protein bands. Autoradiogram of the same gel showed clear radioactive bands each of which corresponded to the CBB band (Fig. 4(b)). Since histone H1 and MBP are commonly used as substrates for serine and threonine kinases, we used both proteins as model substrates to find out if any specificity of the kinase activity could be observed among the products. But no substrate specificity was observed. For all 439 cDNAs the system worked very well, producing mostly soluble form with a single product, while in 152 (35%) cases the amounts of proteins were below the CBB detectable level, some of them may have been insoluble and were lost during the centrifugation step. The products were correct in size as estimated from their

Table 1
Summary of autophosphorylation activity of 439 *Arabidopsis* protein kinases

Class[a]	Group[a]	Total clones	Synthesized clones	Autophosphorylated clones
Class 1 – transmembrane receptor kinase and related non-transmembrane kinase	Group 1.1	27	9	5
	Group 1.2 – receptor like cytoplasmic kinase	66	34	25
	Group 1.3	33	4	2
	Group 1.4 – crinkly 4 like kinase	8	2	0
	Group 1.5	37	13	2
	Group 1.6	41	18	13
	Group 1.7 – S-domain and Duf26 domain kinase	90	16	3
	Group 1.8 – leucine rich repeat kinase	49	5	0
	Group 1.9 – CRPK1 like kinase (types 1 and 2)	37	9	3
	Group 1.10 – receptor like cytoplasmic kinase VI	15	8	6
	Group 1.11 – legume lectin domain kinase	41	6	1
	Group 1.12 – leucine rich repeat kinase	83	48	7
	Group 1.13 – leucine rich repeat kinase	44	20	6
	Group 1.14	23	9	3
	Group 1.15	9	4	1
	Group 1.16 – receptor like cytoplasmic kinase	12	6	2
	Group 1.17 – wall associated kinase	6	1	1
	Group 1. Other – singletons and small groups without close relatives	66	21	5
Class 2 – ATN1/CTR1/EDR1/GmPK6 like kinase	Group 2.1 – ATN1/CTR1/EDR1/GmPK6 like kinase	54	29	17
	Group 2.2 – unknown function kinase	2	1	0
Class 3 – casein kinase I	Group 3.1 – casein kinase I	18	13	12
Class 4 – non-transmembrane protein kinases	Group 4.1 – mitogen activated protein kinase kinase kinase (MAP3K)	52	25	12
	Group 4.2 – calcium response kinases	133	76	45
	Group 4.3 – unknown function kinase	4	2	0
	Group 4.4 – MAP3K	27	6	1
	Group 4.5 – MAPK/CDC/CK2/GSK kinases	83	52	34
Class 5 – other kinases	Group 5.1 – other kinase	3	2	1
	Group 5.2 – other kinase	1	0	0
Total		1064	439	207

[a] Their classification were according to PlantP database (http://plantsp.sdsc.edu/).

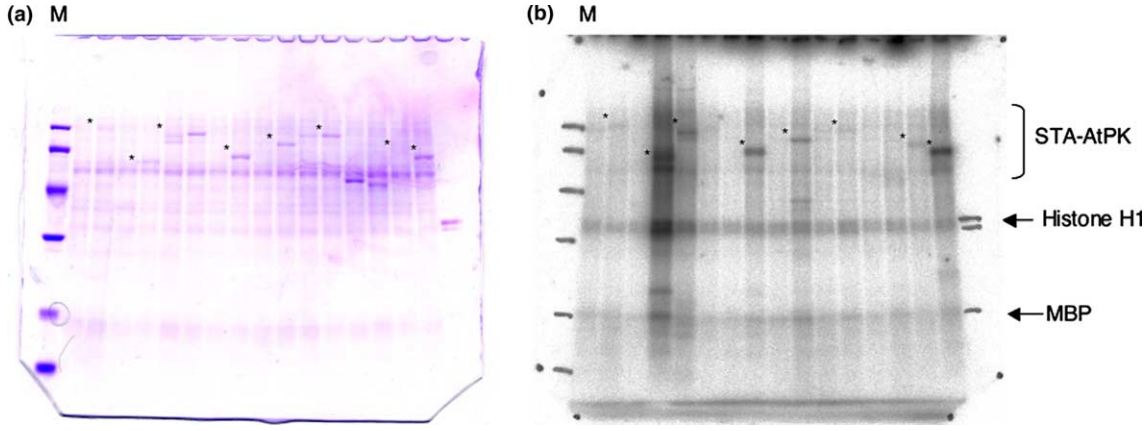

Fig. 4. A high-throughput production and biochemical screening of products from *Arabidopsis* cDNAs encoding protein kinases. Methods for protein synthesis and for analysis were done as described in the text. Sixteen products were separated by SDS–PAGE and stained with CBB (∗) in (a), and the autoradiogram (b). Arrow and arrowhead pointed were histone H1 and myelin basic protein, respectively. M denotes protein size marker.

mobility, and the yield was from 10 to 100 μg/ml of the bilayer reaction volume, as estimated from densitometric scanning of the bands (Table 1). Autoradiography revealed that 207 out of 439 proteins tested displayed autophosphorylation activity, indicating that at least the kinase domain of these proteins is folded into its native

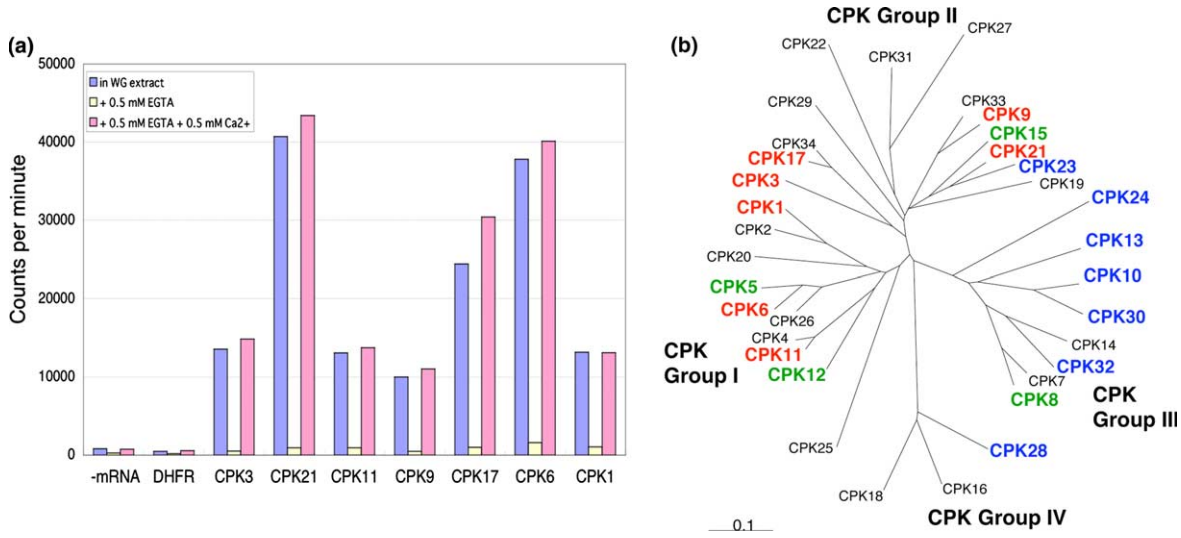

Fig. 5. Calcium dependent phosphorylation of a peptide substrate by calcium-dependent protein kinase cDNA products. Method for the analysis were described in the text. In (a), reaction mixtures contained 0.1 mM Ca^{++} (white bars), in 0.5 mM EGTA (light bars), and 0.5 mM EGTA plus 0.5 mM Ca (dark bars). In (b), phylogenic tree of CDPK family is shown. Red-, blue- and green-colored letters were used to identify the kinases in which both autophosphorylation and peptide phosphorylation activity was detected, only autophosphorylation activity and no activity detected, respectively. Black letter shows the kinases that we did not use.

form in the wheat cell-free system. This screening was performed by one person during two weeks.

2.3. Substrate specificity of the synthesized kinases

To further examine whether those kinases retained their functional integrity as enzyme responsible for the substrate recognition, we selected 26 members of the CDPK family which counts a total of 34 members (Harmon et al., 2001, Group 4.2 in Table 1) and assessed if they had retained their biological activity. Since a model substrate is not known for these proteins in the plant kingdom, we took a biotylated-synthetic peptide (biotin-KKARRQETVDAL, underline showed the phosphorylation site), which is specific for calmodulin-dependent protein kinases. After synthesis, the kinase activity was determined using a kit (SignaTECT, Promega), seven out of 26 protein products (Fig. 5(a)) did specifically phosphorylate the peptide. As can be seen in Fig. 5(b), these enzymes fell in two families in the phylogenetic tree, 3 were in the CPK Group I and 4 were in the CPK Group II. The enzymes belonging to the CPK Group III and IV did not show any phosphorylation activity on the synthetic substrate. These results support that the data obtained reflects the genetic information and biochemical properties of CDPK families rather than being caused by an artificial or random event in the process of the expression system.

3. Discussion

A high-throughput biochemical screening method based on the wheat germ cell-free protein synthesis system has been reviewed. Using a cDNA library of *A. thaliana* genes encoding protein kinases as a model, we showed that the system is powerful for the biochemical characterization of genetic information. In addition, since wheat seeds are available at low cost and may have minimal problems in bio-pollution and ethical issues because it is a foodstuff, our methodology may provide a tool to overcome most of the limitations associated with conventional protein expression systems.

Due to the fact that the full genomic sequence of several organism is now publicly available, we believe that proteomic will be the main focus of the next decade and therefore we anticipate that protein synthesis technologies will experience a one-hundred-fold increases in throughput and the same in decreasing cost. The information on structure and function of gene products accumulated should revolutionize our understanding of biology and fundamentally alter the practice of medicine and influence other industries as well.

4. Experimental

4.1. General

Details of the following procedures were either described or cited previously (Ogasawara et al., 1999; Madin et al., 2000; Sawasaki et al., 2002a,b): isolation of wheat germs and preparation of the extract, generation of DNA template by PCR using split-primers, synthesis of mRNA, protein synthesis in parallel, estimation of amount of protein synthesized by means of densitometric scanning of the Coomassie brilliant blue (CBB)-stained band, and autoradiography.

4.2. Parallel cell-free protein synthesis

The unique primer for the split-primer method for each of the 439 DNAs from 506 RIKEN *Arabidopsis* full-length protein kinase clones was designed according to the sequence in the database (Seki et al., 2002). For the introduction of the STA sequence, a double-stranded DNA with the 3′-portion of SP6 promoter followed by GAAΩ, the STA ORF, and 5′-CCACCCACCAC-CACCA was generated by PCR, purified, and used (0.2 μM) instead of primer-2. Three microliter of *E. coli* suspension grown overnight in a 96-well plate was used for the 60 μl PCR (Sawasaki et al., 2002b), where the amplification reaction reaches the saturation level, and the resulting DNA was transcribed as above. The transcript in each well of the microtiter plate was precipitated with ethanol and was spun-down by centrifugation of the plate using a Hitachi R10H rotor to collect precipitates. Each washed mRNA (usually 30–35 μg) was transferred into 50 μl of the translation mixture and the reaction was done in the bilayer mode as described earlier (Sawasaki et al., 2002a) with slight modifications. Briefly, the translation mixture (the bottom layer) contained $3A_{260}$ units of extract ($A_{260}/A_{280} = 1.55$) and final concentrations of the various ingredients in translation solution are: 24 mM Hepes/KOH, pH 7.8, 1.2 mM ATP, 0.25 mM GTP, 16 mM creatine phosphate, 2 μg creatine kinase, 2 mM DTT, 0.4 mM spermidine, 0.3 mM of each of the 20 amino acids, 2.8 mM magnesium acetate, 100 mM potassium acetate, 0.005% NaN_3 (referred to as substrate solution). For the upper layer, 125 μl of the substrate solution was loaded onto the translation mixture, and incubations were done at 26 °C for 17 h. The synthesizer, PIM, can cover all these steps starting with the transcription up to protein production, being equipped with all the essentials for pipetting, mixing, incubating, centrifuging for the purification of mRNA and then translation.

4.3. Analysis of autophosphorylation activity of parallel expressed kinases and calcium dependent protein kinases

PHOT1 gene protein was synthesized in the presence (0.1 mM) or absence of FMN in the bilayer reaction in darkness and the subsequent sample manipulation and assay in the presence of [γ-^{32}P]ATP were performed under dim red light. To determine the light-dependent autophosphorylation, the reaction mixtures were incubated with [γ-^{32}P]ATP under the blue light irradiation at a total fluorescence of 3000 μmol/m^2. For the assay of autophosphorylation of parallel expressed kinases, 2 μl of the reaction mixture was mixed with 8 μl of biotin magnetic beads (Spherotech Inc., IL), and was washed twice with protein kinase buffer [50 mM Tris–HCl (pH 7.6)/100 mM potassium acetate/10 mM $MgCl_2$/1 mM DTT] (Sawasaki et al., 2002b). Each partially purified protein retained on beads was treated with 20 units of λ protein phosphatase in 20 μl of commercial buffer (New England Biolabs, Inc., Sawasaki et al., 2002b) at 30 °C for 30 min then was suspended in the kinase buffer and washed twice. Beads were suspended in the kinase buffer and the reaction mixture (10 μl) containing [γ-^{32}P]ATP as the phosphate donor was incubated 30 °C for 30 min. Samples were boiled in the SDS-denaturing buffer and proteins were separated on SDS–polyacrylamide gel. Assay for calcium dependent protein kinases was done according to manufacturer's instructions using a biotylated-synthetic peptide as substrate (SignaTECT, Promega). Each synthesized kinases on the beads as described above was incubated in the presence or absence of calcium ion with [γ-^{32}P]ATP. Phosphorylated synthetic peptide, biotin-KKARRQET(P)VDAL, was bounded on the commercial membrane and the radioactivity was determined.

Acknowledgements

We thank Yuko Matsubara for expert technical assistance and Germain Hugo for critical reading of the manuscript. This work was supported by the Special Coordination Funds for Promoting Science and Technology by the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

Briggs, W.R., Beck, C.F., Cashmore, A.R., Christie, J.M., Hughes, J., Jarillo, J.A., Kagawa, T., Kanegae, H., Liscum, E., Nagatani, A., Okada, K., Salomon, M., Rudiger, W., Sakai, T., Takano, M., Wada, M., Watson, J.C., 2001. The phototropin family of photoreceptors. Plant Cell 13, 993–997.

Christie, J.M., Reymond, P., Powell, G.K., Bernasconi, P.B., Raibekas, A.A., Liscum, E., Briggs, W.R., 1998. Arabidopsis NPH1: a flavoprotein with the properties of a photoreceptor for phototropism. Science 282, 1698–1701.

Chrunyk, B.A., Evans, J., Lillquist, J., Young, P., Wetzel, R., 1993. Inclusion body formation and protein stability in sequence variants of interleukin-1 beta. J. Biol. Chem. 268, 18053–18061.

Endo, Y., Sawasaki, T. High-throughput, genome-scale protein production method based on the wheat germ cell-free expression system. J. Struct. Funct. Genomics, in press.

Goff, S.A., Goldberg, A.L., 1987. An increased content of protease La, the lon gene product, increases protein degradation and blocks growth in *Escherichia coli*. J. Biol. Chem. 262, 4508–4515.

Harmon, A.C., Gribskov, M., Gubrium, E., Harper, J.F., 2001. The CDPK superfamily of protein kinases. New Phytol. 151, 175–183.

Hartl, F.U., Hayer-Hartl, M., 2002. Molecular chaperones in the cytosol: from nascent chain to folded protein. Science 295, 1852–1858.

Henrich, B., Lubitz, W., Plapp, R., 1982. Lysis of *Escherichia coli* by induction of cloned phi X174 genes. Mol. Gen. Genet. 185, 493–497.

Kolb, V.A., Makeyev, E.V., Spirin, A.S., 2000. Co-translational folding of an eukaryotic multidomain protein in a prokaryotic translation system. J. Biol. Chem. 275, 16597–16601.

Kurland, C.G., 1982. Translational accuracy in vitro. Cell 28, 201–202.
Pavlov, M.Y., Ehrenberg, M., 1996. Rate of translation of natural mRNAs in an optimized in vitro system. Arch. Biochem. Biophys. 328, 9–16.
Madin, K., Sawasaki, T., Ogasawara, T., Endo, Y., 2000. A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes. Proc. Natl. Acad. Sci. USA 97, 559–564.
Ogasawara, T., Sawasaki, T., Morishita, R., Ozawa, A., Madin, K., Endo, Y., 1999. A new class of enzyme acting on damaged ribosomes: ribosomal RNA apurinic site specific lyase found in wheat germ. EMBO J. 18, 6522–6531.
Roberts, B.E., Paterson, B.M., 1973. Efficient translation of tobacco mosaic virus RNA and rabbit globin 9S RNA in a cell-free system from commercial wheat germ. Proc. Natl. Acad. Sci. USA 70, 2330–2334.
Sawasaki, T., Hasegawa, Y., Tsuchimochi, M., Kamura, N., Ogasawara, T., Endo, Y., 2002a. A bilayer cell-free protein synthesis system for high-throughput screening of gene products. FEBS Lett. 514, 102–105.
Sawasaki, T., Ogasawara, T., Morishita, R., Endo, Y., 2002b. A cell-free protein synthesis system for high-throughput proteomics. Proc. Natl. Acad. Sci. USA 99, 14652–14657.
Seki, M., Narusaka, M., Kamiya, A., Ishida, J., Satou, M., Sakurai, T., Nakajima, M., Enju, A., Akiyama, K., Oono, Y., Muramatsu, M., Hayashizaki, Y., Kawai, J., Carninci, P., Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A., Shinozaki, K., 2002. Functional annotation of a full-length *Arabidopsis* cDNA collection. Science 296, 141–145.