

Exploration of Human ORFeome: High-Throughput Preparation of ORF Clones and Efficient Characterization of Their Protein Products

Takahiro NAGASE^{1,*}, Hisashi YAMAKAWA¹, Shinichi TADOKORO², Daisuke NAKAJIMA¹, Shinichi INOUE¹, Kei YAMAGUCHI³, Yasuhide ITOKAWA⁴, Reiko F. KIKUNO¹, Hisashi KOGA¹, and Osamu OHARA^{1,2,5}

Department of Human Genome Research, Kazusa DNA Research Institute, 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan¹; Laboratory of Pharmacogenomics, Graduate School of Pharmaceutical Sciences, Chiba University, 1-8-1 Inohana, Chuo-ku, Chiba 260-8675, Japan²; Graduate School of Science and Technology, Chiba University, 648 Matsudo, Matsudo, Chiba 271-8510, Japan³; Graduate School of Science, Toho University, 2-2-1 Miyama, Funabashi, Chiba 274-8510, Japan⁴ and RIKEN Research Center for Allergy and Immunology, 1-7-22 Suehiro, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan⁵

(Received 10 January 2008; accepted on 29 January 2008; published online 2 March 2008)

Abstract

In this study, we established new systematic protocols for the preparation of cDNA clones, conventionally termed open reading frame (ORF) clones, suitable for characterization of their gene products by adopting a restriction-enzyme-assisted cloning method using the Flexi[®] cloning system. The system has following advantages: (1) preparation of ORF clones and their transfer into other vectors can be achieved efficiently and at lower cost; (2) the system provides a seamless connection to the versatile HaloTag[®] labeling system, in which a single fusion tag can be used for various proteomic analyses; and (3) the resultant ORF clones show higher expression levels both *in vitro* and *in vivo*. With this system, we prepared ORF clones encoding 1929 human genes and characterized the HaloTag-fusion proteins of its subset that are expressed *in vitro* or in mammalian cells. Results thus obtained have demonstrated that our Flexi[®] ORF clones are efficient for the production of HaloTag-fusion proteins that can provide a new versatile set for a variety of functional analyses of human genes.

Key words: cDNA; HaloTag; proteomics; Flexi cloning; ORFeome

1. Introduction

The entire human genome sequence has allowed us to create a provisional catalog of human transcripts and proteins by a combination of computational predictions and information from experimentally collected cDNA sequences.¹ Although approximately 23 000 protein-coding genes have been assigned to the human genome by the Ensembl database (NCBI built 36, Ensembl 48; http://www.ensembl.org/Homo_sapiens/index.html),² the functional roles of

the respective gene products remains to be experimentally explored. Because of the importance of these studies, many groups have started to collect a set of protein-coding sequences, conventionally termed open reading frames (ORFs), in a genome-wide manner, so-called 'ORFeome cloning'. ORF clones serve as versatile reagents for the functional and structural studies of proteins.^{3–5} Therefore, various systems have been developed to enable efficient large-scale cloning and expression of ORF clones. Currently available systems rely on either site-specific recombinases or rare-cutting restriction enzymes for the transfer of ORFs between expression vectors. These systems allow the correct orientation and reading frame to be maintained after transfer.^{6–8} The Gateway and Creator cloning

Edited by Katsumi Isono

* To whom correspondence should be addressed. Tel. +81 438-52-3930. Fax. +81 438-52-3931. E-mail: nagase@kazusa.or.jp

© The Author 2008. Kazusa DNA Research Institute.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org

systems, which are representative of the recombinase-based systems, transfer ORFs by λ -att recombination using the 25-bp *attB* site, and Cre-*loxP* recombination using the 34-bp *lox-P* site upstream of the ORF, respectively.⁷ On the other hand, restriction-enzyme-based systems are less frequently used at present, although a commercial system has recently emerged that is as efficient as those based on recombinases. The Flexi[®] cloning system can manipulate ORFs using rare-cutting restriction enzymes, *SgfI* and *PmeI*, after the addition of these 8-bp restriction sites to the flanking 5' and 3' ends of the ORFs, respectively.⁸ We previously prepared a set of ORF clones for human KIAA genes using a Gateway-type vector.⁹ Furthermore, we created the ORF clones as fusion proteins with green-fluorescent protein to analyze the subcellular localization of the proteins in cultured cells and to evaluate the *in vitro* expression of ORF clones.⁹ However, when we wish to carry out alternative functional studies beyond bioimaging, the Gateway system requires re-construction of appropriate expression plasmids for the creation of differently tagged proteins. This creates a serious bottleneck in the use of our large ORF clone set for functional genomics analysis. Thus, we started to look for a more versatile tag system that would allow us to carry out both bioimaging and biochemical experiments using a single tag. Within this context, we considered a recently emerging technology, commercially designated as HaloTag[®] technology.¹⁰ This technology is quite attractive since HaloTag can be used not only for bioimaging, but also for various proteomic applications involving fusion protein immobilization. Although similar technologies have been developed, such as SNAP-tag which uses a modified human O⁶-alkylguanine DNA alkyl transferase as the tag,^{11,12} we preferred HaloTag[®] technology because we felt it may be more suitable for studies in mammals. Haloalkane dehalogenase, the enzyme from which HaloTag is derived, does not exist in mammalian cells, unlike O⁶-alkylguanine DNA alkyl transferase. Since HaloTag[®] technology was available in the Flexi[®] cloning system, we were strongly motivated to examine whether the Flexi[®] cloning method was suitable for preparation of a large number of ORF clones, and whether it would be useful for downstream functional characterization.

In this study, we prepared nearly 2000 ORF clones using the Flexi[®] cloning method, and the resultant ORF clones were examined with respect to production in cell-free protein synthesis systems as HaloTag-fusion proteins. The results indicated that the Flexi[®] cloning method was as efficient as the Gateway cloning method and that the Flexi ORF clones consistently produced equal or larger amounts of HaloTag-fusion proteins than Gateway clones in *in*

vitro protein production systems. We also examined the subcellular localizations of 40 HaloTag-fusion proteins in HEK293 cells and confirmed that the results were equivalent to those of the corresponding Monster Green[®] Fluorescent Protein (MGFP)-fusion proteins. Taken together, these results indicate that the Flexi ORF clone set can perform efficient functional analyses of human genes and provide an alternative resource for the human ORFeome project.

2. Materials and methods

2.1. Materials

pTD1 plasmid was obtained from Shimadzu (Kyoto, Japan). pF1Kof vector, which was constructed by flipping the *AccIII*-*AgeI* 832-bp DNA fragment including the origin of replication in the pF1KT7 (Promega, Madison, USA) vector, was kindly provided by Promega Corporation. pF3AHT vector was constructed by inserting the blunt-ended *EcoICRI*-*Sall* 946-bp DNA fragment from the pFC8A (HT) vector (Promega) into the *EcoICRI* site of the pF3A WG (BYDV) (Promega) vector backbone in the appropriate order for production of a C-terminal HaloTag-fusion protein. pFC8A (MGFP) was constructed by replacing the HaloTag gene with the MGFP gene from the pHMGFP vector (Promega). For construction of HaloTag-fusion KIAA protein expression clones in the pTD2-Gateway and pTD2-Flexi vectors, KIAA ORFs with the HaloTag sequence were extracted from pF3AHT ORF clones. The ORFs for the synthetic firefly *luciferase* gene (*luc2*), HaloTag, and Venus were obtained from pGL4.10[*luc2*], pFC8A vector (Promega), and (EYFP-F46L/V68L/M153T/V163A/S175G)/pCS2 kindly provided by Dr. Atsushi Miyawaki (Brain Science Institute, RIKEN),¹³ respectively, by PCR and were subcloned into the pF1KT7 Flexi vector. Each ORF was transferred to seven different kinds of modified pTD2 expression vectors whose cloning sites were flanked at both sides with specified sequences. The D5 vector, which carried a 10-bp sequence (AATCGAATTC) in place of the 5'-polyhedrin enhancer sequence, was also used in the preparation of the *luc2* expression plasmid. pFC8KHT-Memb, pFC8KHT-NLS, pFC8KHT-ER, pFC8KHT-Golgi, and pFC8KHT-Mito were constructed by ligation of DNA fragments containing localization signal sequences, with the pFC8K (HaloTag) vector digested by *SgfI* and *EcoICRI* (Carboxy Flexi[®] System, Promega). The localization signal fragments (abbreviated as Memb, NLS, ER, Golgi, and Mito) were amplified by PCR from the following expression clones (BD, Franklin Lakes, USA) with each primer appended by *SgfI* or *PmeI* sites at the 5' end: Memb, a membrane-localized signal from N-terminal sequence of neuromodulin

(pEYFP-Mem); NLS, a nuclear localization signal from a triple repeat of NLS from SV40 large T-antigen (pDsRed2-Nuc); ER, an endoplasmic reticulum (ER) targeting sequence from calreticulin (pDsRed2-ER); Golgi, a Golgi-localized signal from beta 1,4-galactosyltransferase (pEYFP-Golgi); and Mito, a mitochondrial targeting sequence from subunit VIII of cytochrome c oxidase (pDsRed2-Mito). HaloTag Cy5 ligand was prepared by the coupling of HaloTag Amine (O4) ligand (Promega) and Cy5 Mono-Reactive Dye Protein Array Grade (Amersham Biosciences, Piscataway, USA) according to the manufacturer's instructions. Anti-HaloTag IgG was generously provided by Promega Corporation. *Escherichia coli* strain JC8679 (*recB21*, *recC22*, *sbcA23*, *thr-1*, *leuB6*, *phi-1*, *lacY1*, *galk2*, *ara-14*, *xyl-5*, *mtl-1*, *proA2*, *his-4*, *argE3*, *rpsL31*, *tsx-33*, *supE44*, *his-328*) was obtained from the Health Science Research Resources Bank (HSRRB) of the Japan Health Sciences Foundation.

2.2. Preparation of Flexi ORF clones

The ORF of interest was amplified from 1 to 4 ng of a plasmid DNA using 0.5 units of PrimeSTAR™ HS DNA polymerase (Takara Bio, Shiga, Japan) in a 20- μ L PCR reaction mixture that included 0.2 mM each of four dNTPs and 0.5 μ M each of the gene-specific primers (5'-CCCCGCGATCGCCATG N₁₇-3' and 5'-CCCCGTTTAAAC N₂₀-3', where N₁₇ indicates a 17 nt-sense sequence downstream from the start methionine codon of ORF and N₂₀ indicates a 20 nt-anti-sense sequence upstream from the stop codon), under the following PCR conditions: 95°C for 2 min/30 cycles of 98°C for 10 s and 68°C for 3 min/68°C for 5 min. In some cases, 0.5 units of KOD-Plus-DNA Polymerase (Toyobo, Osaka, Japan) was used in a 25- μ L PCR reaction mixture that included 0.2 mM each of four dNTPs, 1 mM MgCl₂ and 0.3 mM each of the gene-specific primers under the following PCR conditions: 94°C for 2 min/30 cycles of 98°C for 10 s, 55°C for 30 s/68°C for 3 min. One-third of the resultant amplified PCR products was digested by *SgfI* and *PmeI* in a 10- μ L reaction mixture of Flexi® buffer with 1 \times Flexi® Enzyme Blend (*SgfI* and *PmeI*) (Promega) for 30 min at 37°C after purification of the PCR products using Wizard SV 96 PCR Clean-Up System (Promega). After inactivation of the restriction enzymes by incubation for 20 min at 65°C, one-sixth of the digested PCR product was ligated with 25 ng of the *SgfI*- and *PmeI*-digested pF1K T7 vector in a 10- μ L reaction of Flexi® Ligase buffer with 10 units of T4 DNA Ligase (Promega) for 60 min at 25°C. One microliter of the ligation reaction was used to transform 10 μ L of JM109 Competent Cells (>10⁸ cfu/ μ g, Promega),

and the transformants were selected on LB agar plates containing 50 μ g/mL of kanamycin. A randomly chosen clone was first examined using the CloneChecker™ System (Invitrogen, Carlsbad, USA) by estimating the size of the supercoiled plasmid DNA. For ORFs cloned by the PCR cloning method, entire sequences were determined with the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, Foster City, USA), with sequencing primers designed every approximate 400 bp in a single strand of ORF. The sequence data were obtained with ABI PRISM 3700 or 3130, and analyzed with SEQUENCHER™ sequence assembly software (Gene Codes Co., Ann Arbor, USA). For ORF Trap cloning, the ORF of interest was transferred to pF1Kof vector using a homologous recombination system in *E. coli* (JC8679), according to the method previously described.⁹ For Flexi® Vector cloning using the pENT entry clones we had previously constructed,⁹ the ORF of interest was recovered by digestion with *BstBI* and *SnaBI* restriction endonucleases whose recognition sites were located upstream and downstream of the ORF, respectively, in the vector portion. It was then inserted between the *BstBI* and *SnaBI* sites of the pSP73Flexi-1 vector, which contains *SgfI* and *PmeI* sites upstream and downstream of the *BstBI* and *SnaBI* sites, respectively. After cloning the ORFs into the pSP73Flexi-1 vector, they were cut with *SgfI* and *PmeI* restriction endonucleases and then re-cloned between the *SgfI* and *PmeI* sites of the pF1K vector.

2.3. In vitro protein production

For *in vitro* translation systems using Wheat Germ Extract Plus (WGEP) (Promega) and a TransDirect insect cell lysate (TD) (Shimadzu, Japan), RNA synthesis was performed using the T7 RiboMAX Express Large Scale RNA Production System (Promega) with linearized template DNAs, according to the supplier's instructions. After purification with the RNeasy Mini Kit (QIAGEN, Hilden, Germany), RNAs were concentrated by ethanol precipitation and dissolved in 15 μ L of DEPC-treated water. *In vitro* translation reactions were performed in a reaction volume of 10 μ L using 2.4 and 3.2 μ g of the RNA for the WGEP and the TD systems, respectively, following the supplier's recommendation unless otherwise stated. For the TNT SP6 High-Yield Protein Expression (HYPER) System (Promega), approximately 400 ng of TempliPhi-amplified pF3AHT HaloTag-fusion protein expression clones were used as a template in a 10- μ L reaction, according to the manufacturer's instructions. For production of radiolabeled proteins, 2.86 μ Ci of ³⁵S-labeled Redivue Promix (a mixture of L-[³⁵S]methionine and L-[³⁵S]cysteine) (GE Healthcare,

USA) was added in the reaction. The TempliPhi (GE Healthcare) reaction was conducted by denaturation of 10–30 ng of DNAs with 5 μ L of Denature buffer for 3 min at 95°C followed by polymerization for 16 h at 30°C. For the HaloTag tetramethylrhodamine (TMR) ligand labeling of the *in vitro* produced HaloTag-harboring proteins, 2.3 μ L of 1 μ M HaloTag-TMR ligand (Promega) was added after the *in vitro* reaction (10 μ L) and the mixture was incubated for 5–10 min at room temperature. After the addition of 4.2 μ L of 4 \times Laemmli's SDS sample buffer, the products were resolved on a MDG-267 Real Gel Plate (concentration gradient: 5–10%; Biocraft, Japan) by polyacrylamide gel electrophoresis in the presence of SDS (SDS–PAGE).

2.4. Examination of the effects of the flanking sequences of ORF on *in vitro* production of luciferase, Venus, and HaloTag

For *in vitro luc2* production, 1.4 μ g of RNAs were used for WGEP and 3.8 μ g for TD in a reaction volume of 12 μ L. The activities of *luc2* in an aliquot of the reaction mixture (10 μ L) were assayed using the Dual-Glo Luciferase Assay System (Promega) according to the supplier's instructions. Luminescence was detected in a GloMAX 96 Microplate Luminometer (Promega). *In vitro* syntheses of Venus proteins were performed in 20 μ L using 4.8 and 6.4 μ g of the template RNA in the WGEP and the TD systems, respectively, after removing turbidity by centrifugation at 15 000 rpm for 2 min. Fluorescent signals were detected in the Applied Biosystems 7500 Fast Real-Time PCR System after 2 and 5 h reactions in the WGEP and TD systems, respectively. HaloTag-TMR ligand binding was monitored for quantification of the active HaloTag protein. Two microliters of 5 μ M HaloTag-TMR ligand was added to an 8- μ L reaction mixture, and the mixture was incubated at 25°C for 5 min. The binding reaction was stopped by adding 2 μ L of 6 \times Laemmli's SDS sample buffer and then boiling at 100°C for 5 min. The solution (7.5 μ L) was analyzed on 12.5% SDS–PAGE. Fluorescent signals from TMR were detected and quantified with FLA-3000 (Fujifilm, Tokyo, Japan) and MultiGauge image analyzing software (Fujifilm).

2.5. Western blotting

For Western blot analysis of HaloTag-fusion proteins, the proteins separated by SDS–PAGE were electrophoretically transferred onto a PVDF membrane (FluoroTrans W; PALL, Portsmouth, UK) using the transfer buffer [25 mM Tris–Cl (pH 8.3), 192 mM glycine, 20% (V/V) methanol] with the BIOCRAFT BE-300 semidry transfer device (Biocraft,

Tokyo, Japan). HaloTag-fusion proteins were detected as follows. The resultant PVDF membrane was washed with TBS [20 mM Tris–Cl (pH 7.5) and 150 mM NaCl] including 0.05% Tween-20 (TBST) for 10 min with gentle agitation. After pre-incubation of the membrane with TBST containing 5% skim milk for 60 min, the membrane was incubated with 1:30 000 rabbit anti-HaloTag IgG antibody (Promega) in TBST containing 1% skim milk for 60 min. After washing with TBST for 5 min four times, the membrane was further incubated with 1:30 000 horseradish-peroxidase-conjugated anti-rabbit IgG antibody (Promega) in TBST containing 1% skim milk for 60 min. After washing the membrane with TBST, HaloTag-fusion proteins were finally detected using ECL plus (GE Healthcare) according to the manufacturer's instructions. The luminescent images were recorded by a Luminescent Image Analyzer LAS3000 (Fujifilm). MagicMark XP Western Protein Standard (Invitrogen) was used for estimation of the apparent molecular masses of the HaloTag-fusion proteins.

2.6. Examination of the subcellular localization of HaloTag-fusion proteins

The human embryonic kidney (HEK) 293 cell line was obtained from HSRRB of the Japan Health Sciences Foundation. HEK293 and COS-7 cell lines were grown in DMEM supplemented with 10% of Tet System Approved Fetal Bovine Serum (BD) and 1 \times Antibiotic–Antimycotic reagent (100 U/mL penicillin + 100 μ g/mL streptomycin + 0.25 μ g/mL amphotericin B; Invitrogen). HEK293 and COS-7 were transfected with expression clones using FuGENE6 Transfection Reagent (Roche Diagnostics, Basel, Switzerland) in an 8-well chambered coverglass (Thermo Fisher Scientific, Rochester, USA) according to the supplier's instructions. Cells (200 μ L, 4 \times 10⁴) were plated 24 h before transfection. After 16–24 h of transfection, HaloTag-fusion proteins and DNA were labeled with medium with 1 μ M of HaloTag-TMR ligand and Hoechst (33342, 3.3 μ g/mL, final concentration; Sigma-Aldrich, St. Louis, USA) for 15 min in a CO₂ (5%) incubator at 37°C. After washing four times with 300 μ L of medium, the cells were incubated for 0.5–24 h in a CO₂ incubator and the subcellular localization of HaloTag- or MGFP-fusion proteins were observed with a Biozero fluorescence microscopy system (Keyence, Osaka, Japan) with 605/55 and 535/50 nm filters. Photos labeled with Hoechst33342 were merged with those of MGFP- or HaloTag-fusion proteins and the subcellular localizations of MGFP- and HaloTag-fusion proteins were compared.

3. Results and discussion

3.1. Construction of Flexi ORF clones

Led by the motivation described in Section 1, we have initiated preparation of Flexi[®] Vector ORF clones using human cDNA clones accumulated by the Kazusa cDNA project.^{14,15} So far, 1929 Flexi ORF clones have been constructed in a pF1K vector format. The list of the Flexi ORF clones is accessible at www.dnaresearch.oxfordjournals.org (Supplementary Table S1). These Flexi ORF clones contained relatively large ORFs derived from 1163 KIAA and 766 known genes, the average sizes of which are 2.8 and 2.4 kb, respectively. ORFs of 341 pF1K clones were PCR-cloned using PrimeStar HS or KOD plus high-fidelity DNA polymerases, as described in Section 2. All ORF clones obtained by PCR cloning were sequence-verified and we found that the PCR mutation occurrence rate was less than 5×10^{-5} mutations per residue under the conditions employed in this study. ORF clones with large ORFs are inevitably at higher risk of having artificial missense mutation(s) introduced by PCR, thus we previously developed and applied the ORF Tap cloning method for the construction of Flexi ORF clones for some relatively long ORFs (mainly those >2 kb).⁹ The method is based on a homologous recombination in *E. coli* JC8679 and thus is very unlikely to introduce mutation(s) into ORFs during cloning. We therefore considered that this method was also highly suited to the preparation of Flexi ORF clones because reconfirmation of the ORF sequence could be avoided.^{16–18} Within this context, 842 pF1K clones were constructed by the ORF Tap cloning method. The remaining 749 pF1K clones, termed Flexi_RBS type, were constructed by transferring ORF sequences from the Gateway entry clones we had previously prepared to the pF1K vector, because the 5'- and 3'-untranslated sequences (UTRs) have been already trimmed out in the Gateway entry clones. These clones contained a 19-bp fragment, including a ribosome-binding site (RBS), between the *SgfI* site and a translational initiation codon. These Flexi clones were distinguished from others by adding 'SD', representing a Shine-Dalgarno sequence as an indication of the presence of an RBS, to their clone IDs, whereas the other pF1K clones were simply termed Flexi type. The ORF flanking sequences in these Flexi clones are shown in Fig. 1A. The *SgfI* site was placed one base upstream of the initiation codon, which allowed the production of recombinant proteins with the native translational initiation site or with N-terminal tags using appropriate Flexi[®] vectors. On the other hand, the *PmeI* site was placed just downstream from the original stop codon of the ORF, which resulted in an attachment

of the Val in Flexi type or the Tyr–Val–Val in Flexi_RBS type to the carboxy end of the native ORF. When an ORF sequence flanked by *SgfI* and *PmeI* was cloned between the *SgfI* and *EcoI*CRI sites of a Flexi[®] vector, the translational stop codon was destroyed and thus the protein was expressed as a carboxy-terminal fusion. ORFs in pF1K clones thus prepared can be easily transferred to other types of Flexi[®] vectors following digestion with *SgfI* and *PmeI* (Fig. 1B).

Information about the Flexi ORF clones (i.e. ORF nucleotide sequences, the predicted primary sequences of the gene products, corresponding Ensembl gene ID, and gene description² in addition to results of computer-assisted analyses of the predicted primary amino acid sequences) is available through our KOP database (<http://www.kazusa.or.jp/kop/>).

3.2. Comparison of the productivities of HaloTag-fusion proteins from Flexi ORF clones and Gateway ORF clones in two *in vitro* protein production systems

Since our focus is on the use of HaloTag[®] technology, we initially prepared a set of Flexi ORF clones as HaloTag-fusions using the Flexi[®] cloning method. To examine the utility of HaloTag[®] technology, HaloTag-fusion products were produced in *in vitro* translation systems. The expressed proteins were analyzed on SDS–PAGE after labeling with a HaloTag-TMR ligand. This simple labeling procedure enabled us to fluorometrically detect only proteins carrying a HaloTag portion on SDS–PAGE, and thereby made it easy to check whether the ORF encoded an appropriately sized protein. Using this analysis, quantities of HaloTag-fusion proteins with ligand-binding activity could be directly estimated from the fluorescence intensities of protein bands on SDS–PAGE. In these experiments, we also compared the performance of two *in vitro* translation systems: one derived from a wheat germ extract and the second from an insect cell extract [commercially named WGEP and TransDirect (TD), respectively]. We used these systems because these eukaryotic cell extracts are known to produce larger amounts of proteins than conventional rabbit reticulocyte lysates. In parallel, we also compared the efficiency of protein synthesis by Gateway-type (pTD2-Gw) and Flexi-type (pTD2-Flx) expression constructs, where ORFs were flanked with different appended sequences that are associated with the respective cloning methods. These appended sequences may affect the translational activity of the HaloTag-fusion constructs *in vitro*. For this purpose, we constructed 13 sets of HaloTag-fusion protein expression plasmids using pTD2-Gw

A

Flexi type

SgfI *PmeI*
 NNN NNN GCG ATC GCC ORF w/o stop codon GTT TAA ACN
 Xaa Xaa Ala Ile Ala Met - - - - - Val Stop

After cloning of *SgfI*-*PmeI* ORF sequence into a C-terminal tag-fusion vector (*SgfI*-*EcoI*CRI site)

NNN NNN GCG ATC GCC ORF w/o stop codon GTT TCT CNN
 Xaa Xaa Ala Ile Ala Met - - - - - Val Ser Xaa

Flexi_RBS type

SgfI *BstBI* *SnaBI* *PmeI*
GCG ATC GCT TTC GAA GGA GAT AGA ACC ORF w/o stop codon TAC GTA GTT TAA ACN
 Ala Ile Ala Phe Glu Gly Asp Arg Thr Met - - - - - Tyr Val Val Stop

After cloning of *SgfI*-*PmeI* ORF sequence into a C-terminal tag-fusion vector (*SgfI*-*EcoI*CRI site)

GCG ATC GCT TTC GAA GGA GAT AGA ACC ORF w/o stop codon TAC GTA GTT TCT CNN
 Ala Ile Ala Phe Glu Gly Asp Arg Thr Met - - - - - Tyr Val Val Ser Xaa

B

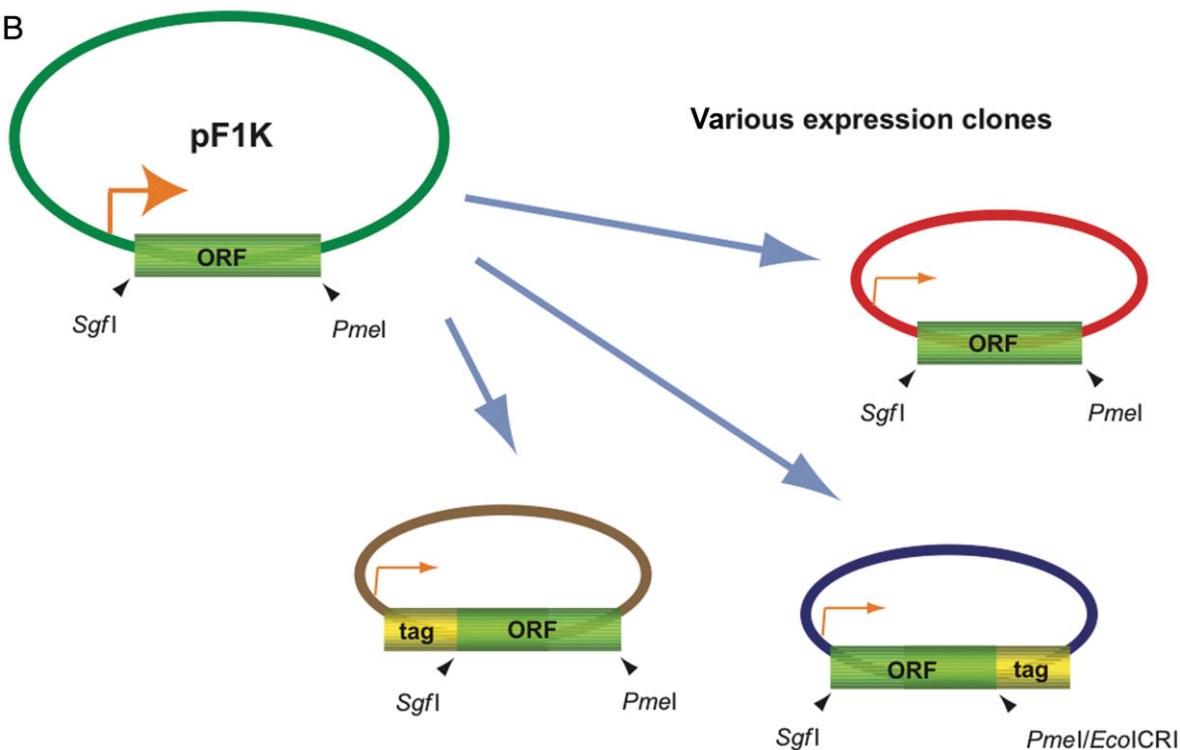


Figure 1. ORF transfer in the Flexi® Vector cloning system. (A) Flanking sequences of ORF in Flexi clones. Recognition sequences of *SgfI* and *PmeI* are indicated as green and red characters, respectively. The nucleotide sequence corresponding to the ribosomal binding site is underlined. The amino acid sequence encoded in the frame in the flanking regions of the ORF is indicated as a three-letter code. Recognition sequences of *BstBI* and *SnaBI*, arising in the vector of Flexi_RBS type are indicated as blue characters. (B) Transfer of the ORF from the pF1K clone to multiple expression vectors. The ORF sequence in the pF1K clone can be easily transferred to a variety of other expression vectors with the correct orientation after digestion by *SgfI* and *PmeI*. For construction of a C-terminal tag-fusion clone, *SgfI*-*PmeI* ORF sequence must be cloned into *SgfI* and *EcoI*CRI sites of the expression vector to omit a stop codon arising in the *PmeI* site. The appropriate promoter is indicated as an orange arrow in the vectors.

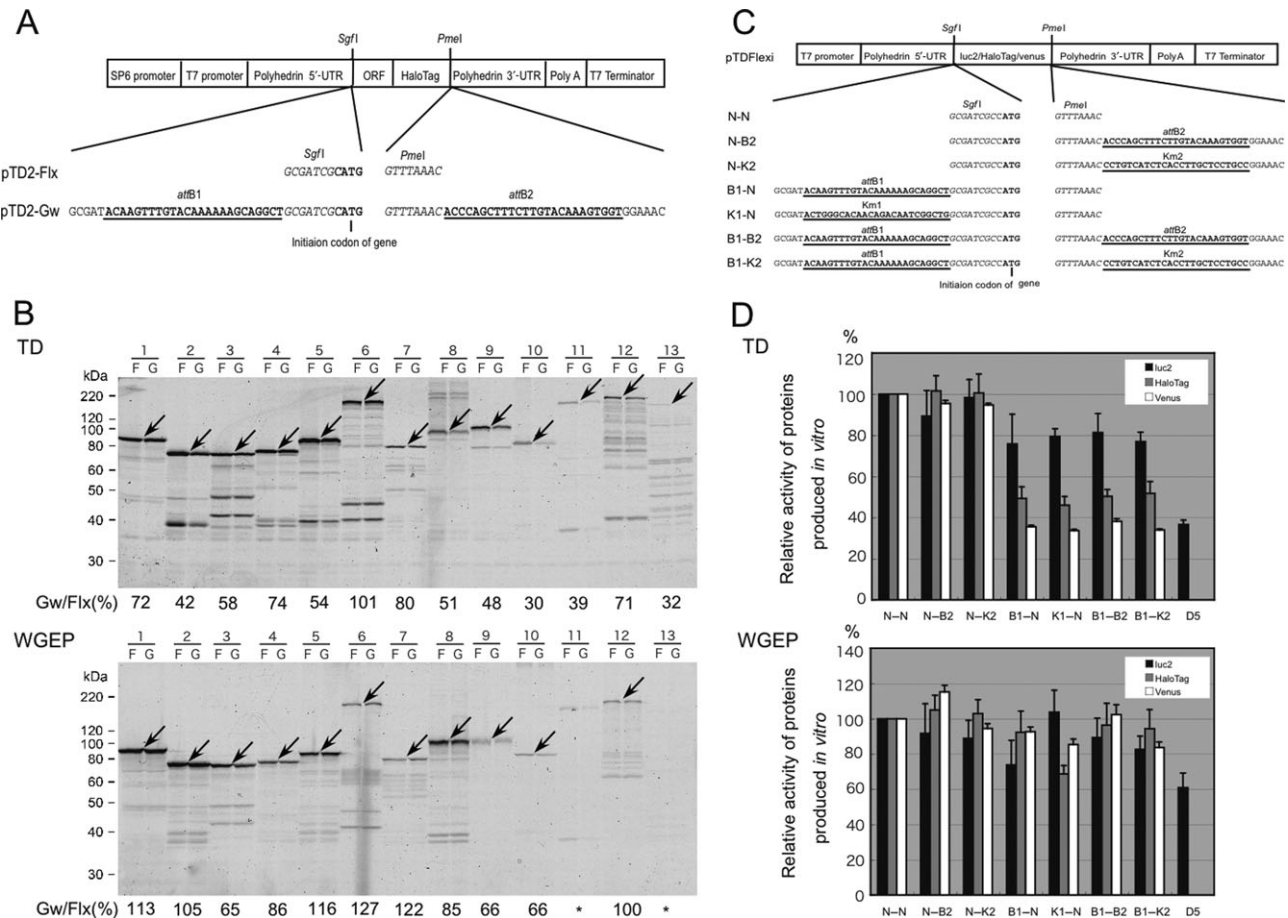


Figure 2. Effects of appended sequences surrounding the ORF on translational activity in cell-free protein synthesis systems. **(A)** Schematic representation of HaloTag-fusion protein expression vectors, pTD2-Flx (Flexi-type), and pTD2-Gw (Gateway-type). Functional elements are indicated in the open boxes. *attB1* and *attB2* represent the Gateway sequences, where specific recombination reactions occur. **(B)** Two types of *in vitro* protein synthesis experiments were carried out using Flexi-type and Gateway-type clones. ORFs from the 13 different genes listed below were cloned into the two different vectors shown in A and used for *in vitro* RNA synthesis, followed by *in vitro* protein synthesis. Products labeled with HaloTag-TMR ligand and their fluorescence intensities were detected on SDS-PAGE using Fluorolmager FLA3000 (indicated by arrows). Upper panel shows the results of the TransDirect insect cell (TD) and lower panel shows those of the WGEP. In each panel, protein products from the Flexi-type (F) and Gateway-type vectors (G) are alternately shown. Ratios of the fluorescence intensities of Gateway products to Flexi products, expressed as percentages, are shown beneath each panel. An asterisk indicates that protein products were under detection limits. The 13 genes examined here and the numbers of amino acid residues without HaloTag are as follows: KIAA0409 (lane 1, 457 a.a.); KIAA1248 (lane 2, 358 a.a.); KIAA0491 (lane 3, 366 a.a.); KIAA1238 (lane 4, 387 a.a.); KIAA1788 (lane 5, 412 a.a.); KIAA1604 (lane 6, 909 a.a.); KIAA1649 (lane 7, 422 a.a.); KIAA1390 (lane 8, 470 a.a.); KIAA2021 (lane 9, 637 a.a.); KIAA0018 (lane 10, 517 a.a.); KIAA1346 (lane 11, 968 a.a.); KIAA1584 (lane 12, 967 a.a.); and KIAA2015 (lane 13, 992 a.a.). **(C)** Schematic representation of *luc2*, HaloTag, and Venus expression clones. Functional elements are indicated in the open boxes. Additional sequences inserted after/before polyhedrin enhancers are shown. Km1 and Km2 were derived from a kanamycin-resistant gene. RBS represents a ribosome-binding sequence. N-N is equivalent to pTDFlexi. **(D)** Results of *in vitro* protein products assay using *luc2*, HaloTag, and Venus expression clones. Their activities were measured as described in Section 2. In the expression clone D5, the polyhedrin 5'-UTR 46-bp sequence was replaced with a 10-bp sequence (AATCGAATTC). Each set of signal intensities was normalized by dividing by that of the N-N, and the averages of three (Venus and HaloTag), six (TD of *luc2*), and seven (WGEP of *luc2*) independent experiments were calculated. Filled bars represent *luc2* activity, gray bars represent the HaloTag-TMR ligand signal, and open bars indicate Venus fluorescence intensity.

and pTD2-Flx vectors. pTD2-Flx is a simple Flexi-type HaloTag-fusion protein expression vector derived from the pTD1 vector for the TD system.¹⁹ This vector has 8-bp restriction enzyme recognition sites (*SgfI* and *PmeI* sites) flanked by the baculovirus polyhedrin 5'- and 3'-UTRs, respectively. pTD2-Gw is a hybrid expression vector derived from pTD2-Flx, which has 25-bp Gateway recombination sites, *attB1*

or *attB2*, between polyhedrin 5'-UTR and *SgfI*, or between *PmeI* and 3'-UTR, respectively (Fig. 2A). In both systems, protein production was directed by *in vitro*-synthesized RNAs. Many fluorescent bands, probably resulting from proteolytic degradation, were observed in most samples. The fluorescence signal intensities of the largest bands were the expected size for the respective protein products,

and each HaloTag-fusion protein was recorded for comparing the protein productivities of pTD2-Flx and pTD2-Gw (Fig. 2B, indicated by an arrow). In the TD system, the fluorescence intensities of 8 out of 13 HaloTag-fusion proteins in the pTD2-Gw format were less than 60% as intense as those derived from pTD2-Flx clones (Fig. 2B, upper lanes: 2, 3, 5, 8–11, and 13). On the other hand, in the WGEP expression system, the fluorescence intensities of HaloTag-fusion proteins from five fusion constructs in the pTD2-Gw vector were slightly reduced compared with those from pTD2-Flx fusion constructs (Fig. 2B, lower lanes: 3, 4, 8–10). Interestingly enough, the fluorescent band patterns of HaloTag-fusions produced in the TD and WGEP systems were considerably different from each other, although the same *in vitro*-synthesized RNAs were subjected to both *in vitro* protein production systems. These results indicated that (1) the protein productivity of both the TD and WGEP systems varied widely from gene to gene, (2) TD and WGEP had different preferences for genes in terms of protein production, and (3) the Gateway recombination site(s) apparently had an inhibitory effect on translational activity, which was more prominent in the TD systems.

3.3. Effects of ORF-flanking sequences on translational activity in cell-free protein synthesis systems

While the different performances of the TD and WGEP systems were anticipated to some extent, it was not anticipated that the Gateway recombination sites would have an inhibitory effect on *in vitro* protein production. Thus, we further pursued the cause of this apparent inhibitory effect of the Gateway recombination sites on *in vitro* protein production. To systematically examine the effects of the appended sequences flanking the ORFs on *in vitro* protein production, we constructed systematically modified versions of expression clones for *luc2*, HaloTag, and Venus auto-fluorescent proteins, with different flanking sequences for each ORF (Fig. 2C). Protein production in the TD and WGEP systems was monitored by measuring luciferase activity, HaloTag-TMR ligand-binding activity, or the fluorescence intensity of Venus protein (Fig. 2D). As shown in Fig. 2C, these clones included flanking sequences derived from the Gateway *attB* sites (B1 and/or B2), or parts of the kanamycin-resistant gene (K1: ACTGGGCACAACAGACAATCGGCTG, or K2: CCTGTCATCTACCTTGCTCCTGCC, as examples of non-specific sequences) between the translational enhancer (polyhedrin 5'-UTR) and the initiation ATG codon, and/or between the stop codon and the polyhedrin 3'-UTR. N–N, an expression clone without any additional sequences, was the control clone. N–B2 and N–K2 carry additional sequences (B2 and K2,

respectively) only at the 3'-flanking site, whereas B1–N and K1–N have additional sequences (B1 and K1, respectively) only at the 5'-flanking site. K1 and K2 were spacer sequences with equal sizes as B1 and B2 and were not expected to be involved with the Gateway recombination. We also compared protein production from the Gateway-type expression vector (B1–B2) and the modified vector (B1–K2) in parallel. In total, we examined seven different expression vectors in these experiments. Figure 2D shows the relative activities of *luc2*, HaloTag, and Venus proteins produced in the TD and WGEP systems compared with the N–N type of pTDFlexi vector. Under these experimental conditions, the luciferase expression clones with *attB1* sequence upstream of the ORFs (Fig. 2, B1–N, B1–B2, and B1–K2) appeared to exhibit only a slight inhibitory effect in the TD and WGEP systems in comparison with the pTDFlexi-*luc2* construct (N–N in Fig. 2C). This reduction in protein productivity seemed to depend on the distance (39 nucleotides) between the 5' translational enhancer and the initiation ATG but not the specific nucleotide sequence, at least in insect TD systems. Expression clones with short nucleotide sequences of the same length derived from the kanamycin-resistant gene also reduced the production of *luciferase* (Fig. 2D, upper, K1–N). Such an inhibitory effect was more evident for HaloTag and Venus protein production in the TD system. In fact, in the TD system, both HaloTag-TMR-ligand-binding activity and the fluorescence intensity of Venus decreased to less than 50% when the additional sequences were located upstream of the ORF (Fig. 2D, upper, B1–N, K1–N, B1–B2, and B1–K2). These results indicated that protein productivity in the TD system was affected by the increase in distance between the polyhedrin 5'-UTR and the start ATG, rather than by the flanking nucleotide sequence of the spacer.

It was reported that the polyhedrin 5'-UTR in pTD1 displayed a strong translational enhancing activity, even in the wheat germ extract and rabbit reticulocyte lysate systems.¹⁹ The inhibitory effect of sequences upstream of the ORF was prominent in the TD system but only minor in the WGEP system (Fig. 2D, lower). Our results also showed that the expression clone lacking 5'-UTR considerably reduced the productivity of *luciferase* in both the TD and WGEP systems, indicating that the translational enhancer of 5'-UTR in pTD1 was functional in each system (Fig. 2D, D5). On the other hand, translational activity was largely unaffected by the spacing between the 3'-translational enhancer and the translational stop codon in all experiments using *luc2*, HaloTag, and Venus proteins (Fig. 2D, N–B2, N–K2). Taking all these results together, we concluded that the spacing between the 5'-translational enhancer and

the translational initiation codon affected protein productivity, at least *in vitro*. Although the mechanism of the inhibitory effect on translational activity by spacing remains to be elucidated, it appears that nucleotide spacing between the enhancer and the initiation codon can sometimes affect translational activity. Because the distance between the 5'-translational enhancer and the initiation codon is shorter in the Flexi ORF clone than in the Gateway ORF clone, these results suggest that the Flexi ORF clone could be the clone of choice to consistently obtain high protein productivity in both the TD and WGEF systems. We consider this as another benefit of using the Flexi ORF clones, since production level is a serious concern, particularly when a recombinant protein is synthesized *in vitro*.

3.4. Quality check of ORF clones from the viewpoint of *in vitro* protein production

Based on the results described above, we confirmed the functionality of a number of Flexi ORF clones produced as HaloTag-fusions in *in vitro* protein production systems. For this purpose, we used a transcription/translation coupled wheat-germ-cell-free protein synthesis system [TNT SP6 High Yield Protein Expression System, hereafter abbreviated as (TNT SP6 HYPES)] because the coupled system is easier and faster to operate and gives a higher yield of recombinant protein production than other uncoupled cell-free protein synthesis systems. Eight hundred and fourteen ORF sequences in pF1K clones were transferred to pF3AHT, C-terminal HaloTag-fusion protein expression vectors, by Flexi[®] cloning methods. One practical problem we encountered was that the coupled system required a large amount of expression plasmids to obtain the best results. Since we planned to use a small volume of *E. coli* culture (about 1 mL) to produce the plasmids, we were concerned that we would not be able to produce enough DNA to obtain good results in the TNT system. To solve this problem, we took advantage of *in vitro* amplification of expression plasmids. We found that DNA amplified with a TempliPhi rolling circle amplification kit could be used as template DNA in the TNT SP6 HYPES without any purification.²⁰ This enabled us to examine the production of a number of HaloTag-fusions in a high-throughput manner. As a result, medium-sized and even relatively large proteins, composed of more than 1000 amino acid residues, were effectively produced and detected by Western blot analysis using anti-HaloTag IgG, with only small amounts of degraded or inappropriately initiated proteins (Fig. 3A). Moreover, most of the HaloTag-fusion proteins retained HaloTag ligand-binding activities (Fig. 3B). Of the 814 pF3AHT

clones tested in the TNT system, 97% produced recombinant proteins that could be detected by TMR HaloTag ligand labeling and Western blot analysis using anti-HaloTag antibody. The apparent molecular masses of the recombinant proteins were estimated by SDS-PAGE followed by Western blot analysis using anti-HaloTag antibodies. Most of the proteins had apparent molecular masses that were consistent with the numbers of amino acid residues of the fused proteins (data not shown). Although the fluorescence intensity of TMR varied widely from protein to protein in a range of arbitrary fluorescence units from 10 to 10 000, HaloTag-fusions that retained the functional characteristics of TMR ligand-binding were efficiently produced. However, we did observe a tendency of HaloTag-TMR ligand binding to display decreases in fluorescence intensity with increases in protein size (Fig. 3C). These results are most likely explained by decreases in the production levels of larger HaloTag-fusions. However, it could also be explained on the assumption that the partner proteins may affect HaloTag-TMR ligand-binding activities. To address this, we examined the relationship between protein yields and their HaloTag ligand-binding activities for 53 HaloTag-fusions. In this experiment, the HaloTag-fusion protein yields were expressed by the radioactivity of [³⁵S]methionine and [³⁵S]cysteine incorporated into the proteins and divided by the number that contained methionine and cysteine, while the HaloTag ligand-binding activities were estimated from the fluorescence intensity of Cy5-labeled HaloTag ligand-bound HaloTag-fusion proteins on SDS-PAGE. The HaloTag-fusion protein yields and the HaloTag ligand-binding activities are plotted against the number of their amino acid residues in Fig. 3D and E, respectively. In Fig. 3F, the ratios of the HaloTag ligand-binding activities to the yields of the HaloTag-fusion proteins are plotted against the number of amino acid residues. The results indicated that the molecular HaloTag ligand-binding activities did not vary widely from fusion to fusion. Based on this result, we consider that the low ligand-binding activities observed resulted from low production levels of HaloTag-fusion proteins, although there may have been some exceptions.

3.5. Subcellular localization of HaloTag-fusion proteins and MGFP-fusion proteins

In general, information on the subcellular localization of proteins provides us with an important clue on potential protein function. Thus, we accumulated data of the subcellular localizations using green fluorescent protein (GFP)-ORF fusions constructed using the Gateway system. It was important for us to

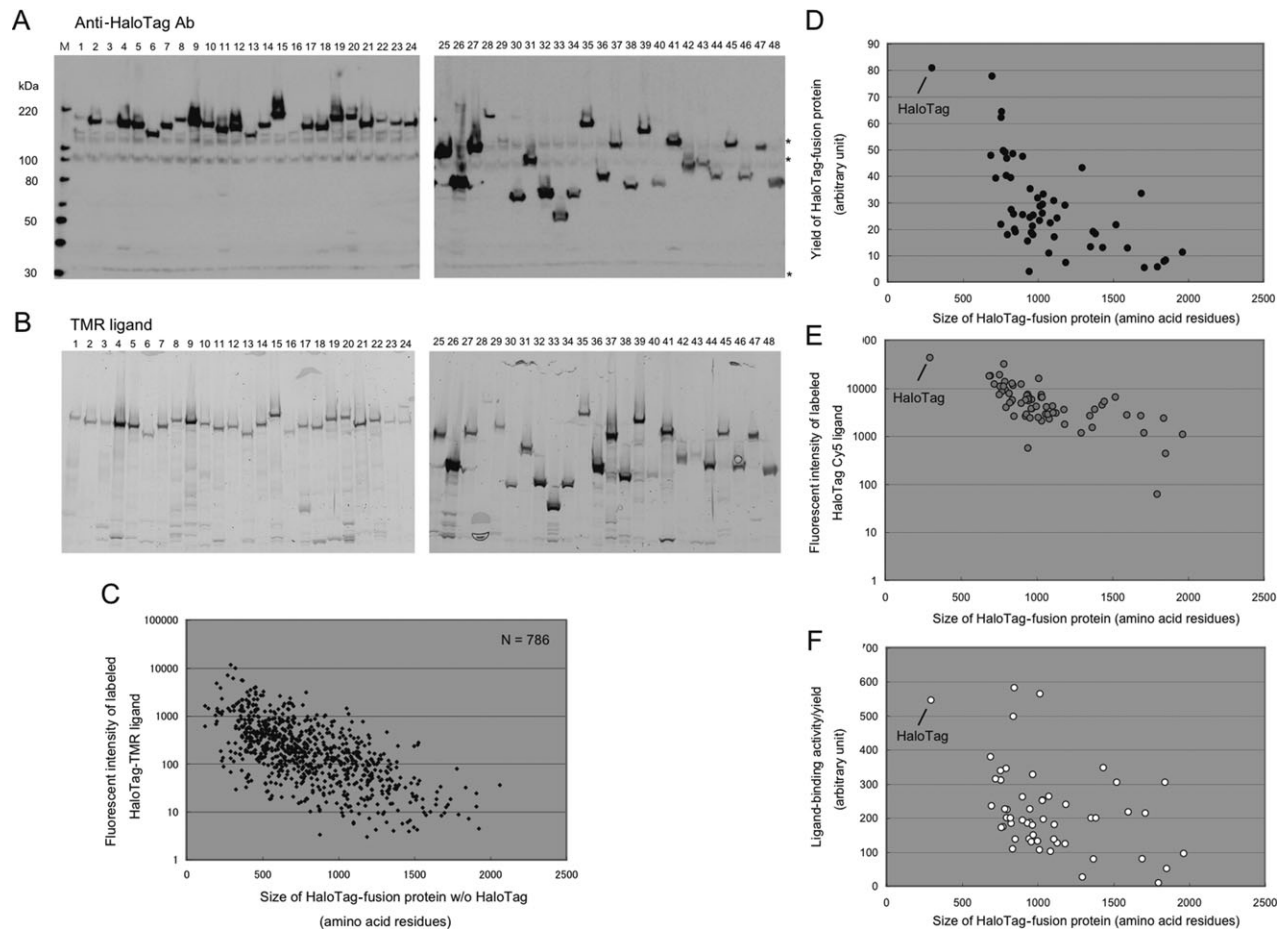


Figure 3. Expression of recombinant HaloTag-fusion proteins in an *in vitro* transcription/translation coupled protein synthesis system. Example of HaloTag-fusion proteins expressed *in vitro*. TempliPhi-amplified 48 pF3AHT DNA clones were used for an *in vitro* transcription/translation system using wheat germ extract (TnT SP6 HYPE System). The recombinant HaloTag-fusion proteins were separated by SDS-PAGE. **(A)** The recombinant proteins were detected by Western blot analysis using anti-HaloTag antibody. Three bands indicated by asterisks are the signals derived from wheat germ endogenous proteins. **(B)** The recombinant proteins were labeled by the HaloTag-TMR ligand after the reaction and detected with FLA3000. The corresponding gene numbers for the HaloTag-fusion recombinant proteins in each lane of (A) and (B) are indicated as follows: 1, KIAA1231; 2, KIAA0720; 3, KIAA1426; 4, KIAA0386; 5, KIAA0651; 6, KIAA0010; 7, KIAA0543; 8, KIAA0987; 9, KIAA1386; 10, KIAA1902; 11, KIAA0594; 12, KIAA1450; 13, KIAA1973; 14, KIAA0055; 15, KIAA0835; 16, KIAA1531; 17, KIAA0750; 18, KIAA1714; 19, KIAA1738; 20, KIAA1106; 21, KIAA1525; 22, KIAA0229; 23, KIAA0182; 24, KIAA0954; 25, KIAA1904; 26, KIAA0020; 27, KIAA0028; 28, KIAA0076; 29, KIAA0150; 30, KIAA0152; 31, KIAA0153; 32, KIAA0168; 33, KIAA0186; 34, KIAA0205; 35, KIAA0223; 36, KIAA0278; 37, KIAA0680; 38, KIAA1184; 39, KIAA1245; 40, KIAA1399; 41, KIAA1871; 42, KIAA1857; 43, KIAA1912; 44, KIAA1916; 45, KIAA1917; 46, KIAA1942; 47, KIAA1949; and 48, KIAA1957. The positions of MagicMark XP Western Protein Standard (M) are indicated in kDa on the left. Relationship between TMR-ligand-binding activities and sizes of HaloTag-fusion proteins synthesized in the TnT SP6 HYPE system. **(C)** Fluorescence intensities of the HaloTag-TMR ligand covalently bound to the recombinant proteins (vertical axis) are plotted against the number of amino acid residues without HaloTag (horizontal axis). **(D)** Yields of the HaloTag-fusion proteins were estimated as the radioactivity of [³⁵S]methionine and [³⁵S]cysteine-labeled proteins calculated by BAS2000 imaging analyzer after SDS-PAGE and the value was divided by the number that contained methionine and cysteine. Yields (vertical axis) are plotted against the number of amino acid residues (horizontal axis) (closed circles). **(E)** HaloTag ligand-binding activity was estimated by measuring the fluorescence intensities of the HaloTag Cy5 ligand covalently bound to the HaloTag-fusion proteins after SDS-PAGE by Fluorolmager FLA3000. The activities (vertical axis) are plotted against the number of amino acid residues (horizontal axis) (gray circles). **(F)** Ratios of HaloTag ligand-binding activities to HaloTag-fusion protein yields (vertical axis) are plotted against the number of amino acid residues (horizontal axis) (open circles). The 53 human genes used for protein production in (D), (E), and (F) are as follows: KIAA0024, KIAA0027, KIAA0057, KIAA0137, KIAA0151, KIAA0199, KIAA0230, KIAA0238, KIAA0352, KIAA0412, KIAA0416, KIAA0422, KIAA0441, KIAA0550, KIAA0557, KIAA0606, KIAA0623, KIAA0644, KIAA0722, KIAA0798, KIAA0813, KIAA0848, KIAA0936, KIAA0939, KIAA0968, KIAA1020, KIAA1048, KIAA1142, KIAA1198, KIAA1225, KIAA1227, KIAA1246, KIAA1253, KIAA1264, KIAA1337, KIAA1369, KIAA1382, KIAA1431, KIAA1469, KIAA1477, KIAA1559, KIAA1572, KIAA1580, KIAA1851, KIAA1852, KIAA1874, KIAA1910, KIAA2003, ABCA3, PAK6, STK38, PINK1, and GPRC5.

confirm that the subcellular localization revealed by HaloTag-fusions was consistent with the subcellular localization observed for GFP-fusions. First, we

examined whether or not the HaloTag protein could be located in an expected subcellular compartment upon fusion of localization signals in COS-7 cells. To

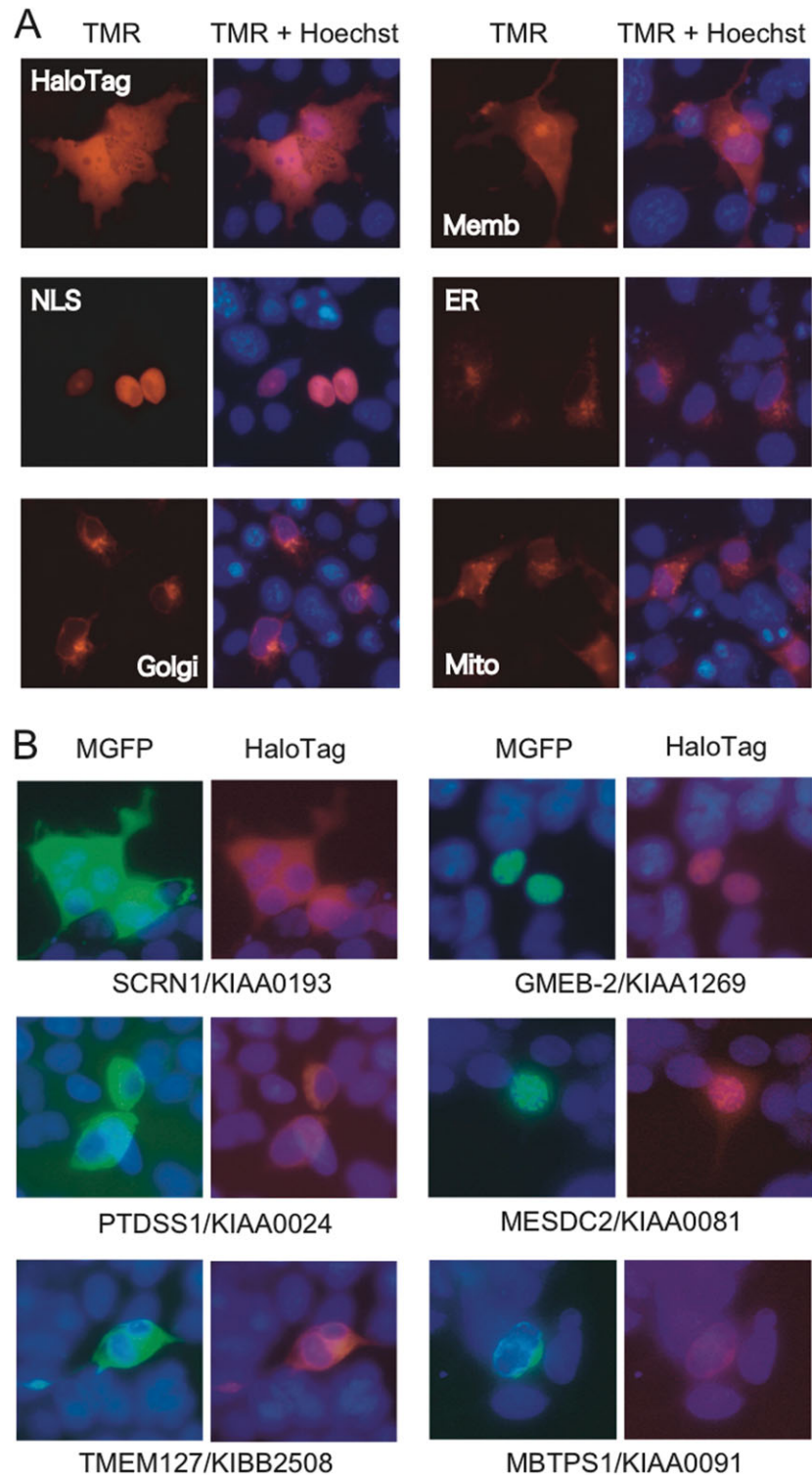


Figure 4. Subcellular localization of HaloTag-fusion proteins in cultured cells. **(A)** Subcellular localization of transiently expressed HaloTag proteins with various signal sequences. COS-7 cells were transfected with pFC8A expression clones for HaloTag proteins containing signal sequences, such as plasma-membrane-localized signal (Memb), nuclear localization signal (NLS), ER targeting sequence (ER), Golgi-localized signal (Golgi), and mitochondria-localized signal (Mito) and HaloTag pHT2 vector which expresses HaloTag ORF (HaloTag). HaloTag-fusion proteins were labeled with medium containing 1-mM HaloTag-TMR ligand (red) for 15 min, washed and incubated for 30 min. Nuclei were stained with Hoechst33342 (blue). Cells were observed by BioZERO fluorescence microscope (TMR). Photos labeled with Hoechst33342 and TMR ligand are merged (TMR + Hoechst). **(B)** Comparison of subcellular localizations between HaloTag- and MGFP-fusion proteins. HEK293 cells were simultaneously transfected with HaloTag-fusion (red) and MGFP-fusion protein (green) expression clones. The HaloTag-fusion proteins (red) and nuclei (blue) were observed as described above, except that the final incubation before observation was 24 h. Photos labeled with Hoechst33342 are merged with those of MGFP- or HaloTag-fusion proteins (MGFP or HaloTag). Gene symbols and alias names are indicated beneath the photos.

do this, we prepared pFC8A expression clones for the HaloTag protein containing signal sequences, such as membrane-localized signal (Memb),²¹ nuclear localization signal (NLS),²² endoplasmic reticulum targeting sequence (ER),²³ Golgi-localized signal (Golgi),²⁴ and mitochondrial targeting sequence (Mito).²⁵ These sequences were all placed at the N-terminal region of the HaloTag proteins. Eighteen hours after transfection of these constructs into COS-7 cells, the HaloTag-fusion proteins were labeled with the TMR ligand and observed with a fluorescence microscopy system. Figure 4A shows that all fused proteins were detected in an expected subcellular compartment, indicating that HaloTag proteins do not inhibit the function of these subcellular localization signals. Next, we directly evaluated whether HaloTag-fusion proteins were co-localized with the same GFP-fusion proteins when both fusions were expressed in the same cells. C-terminal HaloTag- and MGFP-fusion protein expression clones were prepared by the Flexi[®] cloning system using the Flexi ORF clones we prepared and pFC8A (HaloTag) and pFC8A (MGFP) vectors. After co-transfection of HEK293 cells with the same ORF clone fused to the two different fusion tags, HaloTag-fusion proteins were labeled with the TMR ligand and then the fluorescence images were obtained for TMR- (red) and MGFP-labeled (green) proteins for 40 ORFs. Examples of the cellular localization of the fusion proteins are shown in Fig. 4B. Among them, subcellular locations of glucocorticoid modulatory element-binding protein 2 (GMEB-2/KIAA1269), phosphatidylserine synthase 1 (PTDSS1/KIAA0024), membrane-bound transcription factor site-1 protease (MBTPS1/KIAA0091), and transmembrane protein 127 (TMEM127/KIBB2508) are reported in the UniProt Knowledgebase as nucleus/cytoplasm, membrane, endoplasmic reticulum/Golgi apparatus, and membrane, respectively (<http://ca.expasy.org/sprot/>).²⁶ In contrast, the subcellular location of Mesoderm development candidate 2 (MESDC2/KIAA0081) is not mentioned and Secernin-1 (SCRN1/KIAA0193) is predicted to exist in the cytoplasm from similarities found in the database. The subcellular localizations of the 40 HaloTag-fusion proteins analyzed here were the same as those for MGFP-fusion proteins. These results indicate that HaloTag-fusion proteins can be used as an alternative for subcellular localization analysis in place of conventional autofluorescent proteins.

In conclusion, a complete set of ORF resources in various expression-ready clone formats clearly can serve as versatile sets of reagents for functional genomics research. The rapid, easy, and assured transfer of ORF sequences into different expression vectors is highly critical for this purpose, since many ORF clones are used in combination with protein fusion

tags that allow functional analysis. Within this context, we selected a HaloTag protein as a fusion partner because we considered that HaloTag[®] technology would provide a wide range of utilities that include detection and purification as well as adaptability in functional studies. In fact, we have successfully used HaloTag-fusion proteins in other studies for pulse-chase labeling, protein-protein interaction, and chromosomal immunoprecipitation-like analyses along with bioimaging (unpublished data). In this report, we have shown that the Flexi[®] cloning method is suitable for the preparation of a large number of ORF clones into the pF1K vector and allows seamless implementation of HaloTag[®] interchangeable technology. Moreover, we have demonstrated that most human proteins were efficiently expressed as HaloTag-fusions in an *in vitro* coupled protein expression system, as well as in cultured mammalian cells. The resultant HaloTag-fusion proteins were biochemically active and successfully used for analyses of subcellular localization as an alternative to conventional autofluorescent proteins. Based on these results, we will accumulate a number of human ORF clones in the form of expression plasmids for HaloTag-fusion proteins, either C- or N-terminal fusions, in our ORFeome project since they will be directly used for functional studies, as well as for further transfer to another vector. HaloTag[®] technology provides a single fusion tag that can be utilized for a variety of structural and functional proteomic studies. Although there are a number of ORFeome resources for human genes that have been generated by Gateway recombinational cloning,²⁷ a set of human Flexi ORF clones prepared as described would provide the research community with an important alternative for large-scale proteomic studies and a tool for individual research laboratories.

Acknowledgements: We would like to thank J. Hartnett and M. Slater for their gift of the pF1Kof vector and A. Miyawaki for providing us with venus/pCS2 plasmid. We are grateful to B. Bulleit for his valuable comments. We also thank K. Ozawa, K. Sumi, T. Watanabe, K. Yamada, S. Yokoyama, M. Takazawa, N. Kashima, M. Tamura, H. Kinoshita, E. Suzuki, and C. Mori for their technical assistance.

Supplementary material: Supplementary data are available online at www.dnaresearch.oxfordjournals.org.

Funding

This project was supported in part by a grant for Promega-Kazusa scientific collaboration from Promega Corporation and by a special grant for

acceleration of the practical application of biotechnology from Chiba Prefectural Government.

References

1. Brent, M. R. 2006, Genome annotation past, present, and future: how to define an ORF at each locus, *Genome Res.*, **15**, 1777–1786.
2. Hubbard, T. J. P., Aken, B. L., Beal, K., et al. 2007, Ensembl 2007, *Nucleic Acids Res.*, **35**, D610–D617.
3. Rual, J. F., Hill, D. E. and Vidal, M. 2004, ORFeome projects: gateway between genomics and omics, *Curr. Opin. Chem. Biol.*, **8**, 20–25.
4. Brizuela, L., Richardson, A., Marsischky, G. and Labaer, J. 2002, The FLEXGene repository: exploiting the fruits of the genome projects by creating a needed resource to face the challenges of the post-genomic era, *Arch. Med. Res.*, **33**, 318–324.
5. Temple, G., Lamesch, P., Milstein, S., et al. 2006, From genome to proteome: developing expression clone resources for the human genome, *Hum. Mol. Genet.*, **15**, R31–R43.
6. Walhout, A. J., Temple, G. F., Brasch, M. A., et al. 2000, GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes, *Methods Enzymol.*, **328**, 575–592.
7. Marsischky, G. and LaBaer, J. 2004, Many paths to many clones: a comparative look at high-throughput cloning methods, *Genome Res.*, **14**, 2020–2028.
8. Blommel, P. G., Martin, P. A., Wrobel, R. L., Steffen, E. and Fox, B. G. 2006, High efficiency single step production of expression plasmids from cDNA clones using the Flexi Vector cloning system, *Protein Expr. Purif.*, **47**, 562–570.
9. Nakajima, D., Saito, K., Yamakawa, H., et al. 2005, Preparation of a set of expression-ready clones of mammalian long cDNAs encoding large proteins by the ORF trap cloning method, *DNA Res.*, **12**, 257–267.
10. Los, V. G. and Wood, K. 2007, in: Taylor, D. L., Haskins, J. R. and Giuliano, K. (eds.) *Methods in Molecular Biology*, Totowa: Human Press, **356**, pp. 195–208.
11. Keppler, A., Gendreizig, S., Gronemeyer, T., Pick, H., Vogel, H. and Johnsson, K. 2003, A general method for the covalent labeling of fusion proteins with small molecules *in vivo*, *Nat. Biotechnol.*, **21**, 86–89.
12. Gronemeyer, T., Godin, G. and Johnsson, K. 2005, Adding value to fusion proteins through covalent labeling, *Curr. Opin. Biotechnol.*, **16**, 453–458.
13. Nagai, T., Ibata, T., Park, E. S., Kubota, M., Mikoshiba, K. and Miyawaki, A. 2002, A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications, *Nat. Biotechnol.*, **20**, 87–90.
14. Ohara, O., Nagase, T., Ishikawa, K.-I., et al. 1997, Construction and characterization of human brain cDNA libraries suitable for analysis of cDNA clones encoding relatively large proteins, *DNA Res.*, **4**, 53–59.
15. Nagase, T., Koga, H. and Ohara, O. 2006, Kazusa mammalian cDNA resources: towards functional characterization of KIAA gene products, *Brief. Funct. Genomic Proteomic.*, **5**, 4–7.
16. Gillen, J. R., Willis, D. K. and Clark, A. J. 1981, Genetic analysis of the RecE pathway of genetic recombination in *Escherichia coli* K-12, *J. Bacteriol.*, **145**, 521–532.
17. Oliner, J. D., Kinzler, K. W. and Vogelstein, B. 1993, *In vivo* cloning of PCR products in *E. coli*, *Nucleic Acids Res.*, **21**, 5192–5197.
18. Zhang, Y., Buchholz, F., Muyrers, J. P. and Stewart, A. F. 1998, A new logic for DNA engineering using recombination in *Escherichia coli*, *Nature Genet.*, **20**, 123–128.
19. Suzuki, T., Ito, M., Ezure, T., et al. 2006, Performance of expression vector, pTD1, in insect cell-free translation system, *J. Biosci. Bioengineer.*, **102**, 69–71.
20. Dean, F. B., Nelson, J. R., Giesler, T. L. and Lasken, R. S. 2001, Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification, *Genome Res.*, **11**, 1095–1099.
21. Skene, J. H. and Virag, I. 1989, Posttranslational membrane attachment and dynamic fatty acylation of a neuronal growth cone protein, GAP-43, *J. Cell Biol.*, **108**, 613–624.
22. Lanford, R. E., Kanda, P. and Kennedy, R. C. 1986, Induction of nuclear transport with a synthetic peptide homologous to the SV40 T antigen transport signal, *Cell*, **46**, 575–582.
23. Fliegel, L., Burns, K., MacLennan, D. H., Reithmeier, R. A. and Michalak, M. 1989, Molecular cloning of the high affinity calcium-binding protein (calreticulin) of skeletal muscle sarcoplasmic reticulum, *J. Biol. Chem.*, **264**, 21522–21528.
24. Llopis, J., McCaffery, J. M., Miyawaki, A., Farquhar, M. G. and Tsien, R. Y. 1998, Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins, *Proc. Natl. Acad. Sci. USA*, **95**, 6803–6808.
25. Rizzuto, R., Brini, M., Pizzo, P., Murgia, M. and Pozzan, T. 1995, Chimeric green fluorescent protein as a tool for visualizing subcellular organelles in living cells, *Curr. Biol.*, **5**, 635–642.
26. The UniProt Consortium 2007, The universal protein resource (UniProt), *Nucleic Acids Res.*, **35**, D193–D197.
27. Lamesch, P., Li, N., Milstein, S., et al. 2007, hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes, *Genomics*, **89**, 307–315.