Protein Engineering, Design & Selection vol. 25 no. 12 pp. 813–820, 2012
Published online August 30, 2012 doi:10.1093/protein/gzs051

Generation of a dual-functional split-reporter protein for monitoring membrane fusion using self-associating split GFP

Hirohito Ishikawa[1,3], Fanxia Meng[3], Naoyuki Kondo[4,5], Aikichi Iwamoto[1,2] and Zene Matsuda[1,3,6]

[1]Research Center for Asian Infectious Diseases, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan, [2]Division of Infectious Diseases, Advanced Clinical Research Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan, [3]China-Japan Joint Laboratory of Structural Virology and Immunology, Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Chaoyang District, Beijing 100101, China, [4]Department of Pediatrics, Emory University School of Medicine, 2015 Uppergate Dr, Atlanta, GA 30322, USA and [5]Present address: Department of Molecular Genetics, Institute of Biomedical Science, Kansai Medical University, 10-15 Fumizono-cho, Moriguchi, Osaka, 570-8506, Japan

[6]To whom correspondence should be addressed.
E-mail: zmatsuda@ims.u-tokyo.ac.jp

Received June 4, 2012; revised July 24, 2012; accepted August 4, 2012

Edited By Stefano Gianni

Split reporter proteins capable of self-association and reactivation have applications in biomedical research, but designing these proteins, especially the selection of appropriate split points, has been somewhat arbitrary. We describe a new methodology to facilitate generating split proteins using split GFP as a self-association module. We first inserted the entire GFP module at one of several candidate split points in the protein of interest, and chose clones that retained the GFP signal and high activity relative to the original protein. Once such chimeric clones were identified, a final pair of split proteins was generated by splitting the GFP-inserted chimera within the GFP domain. Applying this strategy to *Renilla reniformis* luciferase, we identified a new split point that gave 10 times more activity than the previous split point. The process of membrane fusion was monitored with high sensitivity using a new pair of split reporter proteins. We also successfully identified new split points for HaloTag protein and firefly luciferase, generating pairs of self-associating split proteins that recovered the functions of both GFP and the original protein. This simple method of screening will facilitate the designing of split proteins that are capable of self-association through the split GFP domains.

Keywords: green fluorescent protein/luciferase/membrane fusion/split protein

Introduction

Split reporter proteins are designed to recover reporter activity when they reassociate. This provides an assay with very low background as the split proteins do not have reporter activity when separated. This is useful for monitoring biological events such as membrane fusion (Holland *et al.*, 2004; Wang *et al.*, 2009; Kondo *et al.*, 2010, 2011), protein–protein interaction (Fields and Song, 1989; Paulmurugan and Gambhir, 2003), and protein targeting to organelles (Kim *et al.*, 2004; Ozawa *et al.*, 2003, 2005). The basic design is to split a reporter protein into several fragments, usually two, that will reassociate when a biological event of interest occurs. The event is monitored by recovery of the activity of the original reporter protein. Split reporters include GAL4 (Fields and Song, 1989), beta-galactosidase (Ullmann *et al.*, 1967; Holland *et al.*, 2004), GFP (Magliery *et al.*, 2004; Wang *et al.*, 2009), *Renilla reniformis* luciferase (RL) (Paulmurugan and Gambhir, 2003; Kim *et al.*, 2004; Kondo *et al.*, 2010), firefly luciferase (FL) (Paulmurugan *et al.*, 2002) and others (Shekhawat and Ghosh, 2011). Some are enzymes and others are not. Although enzymatic split reporters require exogenous substrates, they allow quantitative measurement. Some split proteins self-associate spontaneously while others require the presence of an appropriate self-association module.

Recently, we described a pair of split reporter proteins called dual split proteins (DSPs). DSPs are chimeras of an enzymatic split RL and a non-enzymatic split GFP (Kondo *et al.*, 2010). The latter compensates for the weak self-association of split RL (Paulmurugan and Gambhir, 2003; Cabantous *et al.*, 2005). The availability of a membrane-permeable substrate for RL allows quantitative, real-time monitoring of membrane fusion without cell disruption. The signal from split GFP provides locational information about the membrane fusion (Magliery *et al.*, 2004; Wang *et al.*, 2009; Kondo *et al.*, 2010). The membrane fusion assay based on self-associating split reporter proteins generates a signal faster than transcription-dependent assays, and more easily distinguishes between true fusion events and mere aggregation than dye transfer assays (Barbeau *et al.*, 1998; Blumenthal *et al.*, 2002; Kondo *et al.*, 2010).

The difficulty of finding an optimal split point is a bottleneck in designing split reporter proteins. Often, regions predicted to be unstructured are targeted (Remy and Michnick, 2006), but a positive outcome is not guaranteed and the process remains rather serendipitous. Making many split protein pairs is time consuming and split-protein expression is difficult to examine if antibodies to protein fragments are not readily available.

We developed a new methodology to screen candidate split points and generate pairs of self-associating split reporter proteins. Our method involves three simple steps: inserting a GFP module into a target protein, evaluating the function of the protein with the GFP insert and generating split proteins. As GFP is used like a scanning device, we named our approach the GFP-scanning method.

We applied the GFP-scanning method to RL and identified several new split points with better RL activity recovery than the original DSP. In a fusion assay, the new DSPs were 100 times more sensitive than the previous pair. We also applied the method to HaloTag protein (Los and Wood, 2006), which has a similar protein structure to RL, and to FL, which has a different structure. In both cases, we successfully identified several new split points and generated dual functional split reporter proteins that can be used to monitor the process of membrane fusion.

Materials and methods

Construction of plasmids

Expression vectors were derivatives of phRL-CMV (Promega, Madison, WI). First, the BamHI site of phRL-CMV was disrupted by the QuikChange method (Stratagene, La Jolla, CA). A linker containing BspEI and BamHI sites, TCCGG AGGATCC, was inserted into the target protein gene using QuikChange. The GFPopt (Cabantous *et al.*, 2005) gene with a few flanking amino acid sequences was inserted as a BspE1-BamHI fragment at the introduced linker site. The resultant GFP inserted-protein had the following order [N-terminus of protein]-SGGGG-[GFPopt]-GGGGS-[C-terminus of protein].

The expression vectors for split proteins were made by digesting the GFP inserted gene with NheI-AflIII or with PvuII-XbaI and cloning the fragments into the original $pDSP_{1-7}$ and $pDSP_{8-11}$ plasmids (Kondo *et al.*, 2010) at the corresponding restriction enzyme sites.

Solvent accessibility calculation

The accessible surface area of each residue of the protein of interest was calculated using Surface Racer 5.0 (Tsodikov *et al.*, 2002) using the Rluc8 structure (Loening *et al.*, 2007) (PDB code 2PSD) or the FL structure (Franks *et al.*, 1998) (PDB code 1BA3). The value was divided by the accessible surface area calculated for residue X in the tripeptide G-X-G (Chothia, 1976) and expressed as solvent accessibility (SA) as a percentage.

Cell culture and transfection

293FT cells (Invitrogen, Carlsbad, CA) or 293CD4 cells, which is 293 cells constitutively expressing human CD4 (Miyauchi *et al.*, 2005), were grown in Dulbecco's modified Eagle's medium (Sigma-Aldrich, St. Louis, MO) supplemented with 10% fetal bovine serum (HyClone, Thermo Fisher Scientific, Waltham, MA). Cells were kept under 5% CO_2 in a humidified incubator (SANYO, Tokyo, Japan). Cells were seeded into 6-well or 96-well plates (BD Falcon, BD Biosciences, San Jose, CA) one day before transfection. Transfection was performed with FuGENE HD transfection reagent (Roche Diagnostic, Indianapolis, IN). 293FT cells (2×10^5) were transiently transfected using FuGENE HD (Roche Diagnostic, Indianapolis, IN) with 2 or 0.1 μg DNA in 6-well or 96-well culture plates, respectively.

Immunoblotting

Immunoblotting was performed as previously described (Long *et al.*, 2011). In brief, 2 days after transfection, transfected cells were lysed with RIPA lysis buffer (Thermo Fisher Scientific, Waltham, MA) for sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis immediately after images of the transfected cells were taken using microscope IX71 (OLYMPUS, Tokyo, Japan) (10 × objective lens). Cell lysates were subjected to electrophoresis (10% SDS-PAGE, Bio-Rad Ready-Gel J; Bio-Rad, Hercules, CA) and transferred to polyvinylidene fluoride membranes (Immobilon-PSQ, Millipore, Billerica, MA) using a passive transfer method. Blots were probed with anti-GFP monoclonal antibody (Proteinstar, Beijing, China or Santa Cruz biotechnology, Santa Cruz, CA). Anti-mouse immunoglobulin (GE Healthcare, Piscataway, NJ) was used as the secondary antibody. Blots were incubated with a streptavidin-horseradish peroxidase conjugate (GE Healthcare Bio-Sciences AB) and signal was generated with Lumi-LightPLUS (Roche Diagnostic, Indianapolis, IN). Images were obtained with LAS-3000 imaging system (Fujifilm, Tokyo, Japan).

Measurement of enzymatic activity

Two days after transfection, the RL activities of samples were measured using the *Renilla* Luciferase Assay System (Promega, Madison, WI) and a GloMax 96 Microplate Luminometer (Promega, Madison, WI). The FL activities of samples were measured using the Luciferase Assay System (Promega, Madison, WI) and a GloMax 96 Microplate Luminometer (Promega, Madison, WI). To detect the HaloTag, TMRDirect Ligand (Promega, Madison, WI) was added to the medium one day before observation by microscope IX71.

DSP assay

An RL−DSP assay was performed as described previously (Kondo *et al.*, 2010, 2011). In brief, 293FT cells (1.3×10^4 per well) were prepared in 96-well plates and 293CD4 cells (9×10^5 per dish) in UpCell dishes (6 cm diameter, CellSeed, Tokyo, Japan), one day before transfection. The RL−DSP_{1-7} expression vector and pHIVenvOPT (Liu *et al.*, 2010) were cotransfected into 293FT cells, and the RL−DSP_{8-11} expression vector was transfected into 293CD4 cells. After 48−60 h, the culture medium was replaced with fresh medium containing 60 μM EnduRen (Promega, Madison, WI). Two hours later, transfected 293CD4 and 293FT cells were cocultured, and RL activity was measured with a GloMax 96 Microplate Luminometer for 0−5 h.

The procedure for the Halo-DSP assay was the same as the RL−DSP assay, except the Halo-DSP_{1-7} expression vector and pHIVenvOPT were cotransfected into 293FT cells; the Halo-DSP_{8-11} expression vector was transfected into 293CD4 cells, and the TMRDirect Ligand was added to 293FT cells instead of EnduRen, one day before the assay. Images were captured by IN Cell Analyzer 1000 (GE Healthcare, Piscataway, NJ) using a 10 × objective lens.

Results

Overview of the GFP-scanning method

We were interested in generating split proteins that self-reassociate through attached self-associating split GFP modules. The structure of the reassociated DSP reporters is expected to resemble that of GFP-inserted RL (RL-GFP). Therefore, RL-GFP with a high RL activity should generate the DSP reporters that recover a high RL activity when RL-GFP is

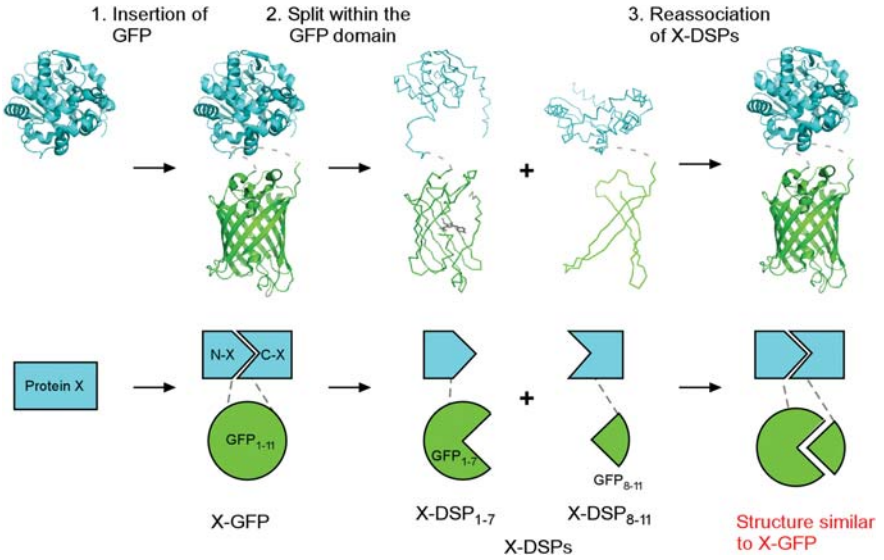

Fig. 1. Basic outline of the GFP-scanning method. The first step is inserting GFP into the target protein (X) to screen for potential split points. The generated chimeric protein, X-GFP, has the structure [N-terminal portion of X]-[GFP_{1-11}]-[C-terminal portion of X]. Subscript numbers indicate beta-strand numbers in GFP. X-GFP that retains high X activity is screened. Second, X-GFP is split between the seventh and eighth beta-sheets of the GFP domain to generate the pair of split proteins X-DSP_{1-7} and X-DSP_{8-11} (subscript indicates beta-sheets in each split protein). Third, the structure of X-DSPs after reassociation is predicted to be similar to the X-GFP structure. Activity of X-DSPs after reassociation is predicted to be screenable by measuring X-GFP activity. GFP, green; protein of interest X, cyan.

split within the GFP module (Fig. 1). Self-association of DSP is promoted through the strongly self-associating split GFP modules. On the basis of this assumption, our strategy was to insert GFP modules in the target protein and screen chimeric proteins with a high RL activity plus a strong GFP signal. These candidates are split within the GFP module. This method has two advantages: (i) screening uses a single expression vector instead of two for each candidate site, and (ii) expression of the chimeric proteins can be easily determined by GFP signal.

In the GFP tertiary structure, the N- and C-terminal ends are close. Therefore, we expected that GFP could be accommodated at an insertion point without severely disturbing the structure of the target protein. As GFP was acting like a scanning device at a potential split point, we called our approach the GFP-scanning method. If no GFP signal was observed after inserting at a particular insertion site, that candidate site was eliminated from further analyses to avoid split proteins that were unstable and inexpressible. Target proteins with GFP insert (X-GFP) that showed GFP signal were examined for retention of the original property of X.

Split protein pairs were generated by splitting an X-GFP that retained high X activity in the GFP portion (between the seventh and eighth beta-strands). The architecture of the split protein pair was [N-terminal portion of X]-[GFP_{1-7}]; [GFP_{8-11}]-[C-terminal portion of X]. Here, the subscript indicates the beta-strand number. We used the term X-DSP_{1-7} for the [N-terminal portion of X]-[GFP_{1-7}], and X-DSP_{8-11} for the [GFP_{8-11}]-[C-terminal portion of X] portion (Fig. 1). In previous work (Kondo *et al.*, 2010), we used the term DSP_{1-7} for [Residues 1-229 of RL]-[GFP_{1-7}] as DSP_{8-11} for [GFP_{8-11}]-[Residues 230-311 of RL]. However, because different DSPs with different split points were made in this study, we renamed them $RL_{229-230}DSP_{1-7}$ and $RL_{229-230}DSP_{8-11}$ to expand the definition of our DSP. These pairs of DSPs reassociated efficiently because of the strong self-association activity of the split GFP portion.

GFP-scanning method identified potential new split points in RL

Candidate split points for RL were selected based on information from the structure of Rluc8 (PDB code 2PSD) (Loening *et al.*, 2007), which is a variant of RL with eight mutations in a total of 311 residues. The SA of each residue was calculated and the points where the two adjacent residues had a high average SA (between amino acids 154-155, 208-209, 227-228, and 248-249) were chosen as the candidate split points. We hypothesized that the more exposure to the solvent, the better the point for splitting. Loop regions with moderate SA (69-70, 274-275) or low SA (123-124, 223-224) were also chosen as candidate split points for comparison. We also chose split points flanking the split point 154-155 (153-154, 155-156, 156-157, 157-158) for more detailed analysis.

Each RL-GFP construct was transfected into 293FT cells, and the GFP signal (Fig. 2a and Supplementary Fig. S3a) and RL activity (Table I) were measured. Judging from the GFP signal, all proteins were expressed well. All constructs showed some dense green spots, which might have been from protein aggregation (Supplementary Fig. S1). The GFP signal was not likely due to an internal translation initiation of GFP, because immunoblotting with anti-GFP antibody detected only a band near 62 kDa, which corresponded to the molecular weight of RL-GFP, but not GFP alone (26 kDa) (Fig. 2b and Supplementary Fig. S3b). Therefore, the GFP signal served as an expression marker of RL-GFP. The amount of RL_{69-70}GFP and $RL_{123-124}$GFP, by immunoblotting, seemed to be low compared with the other RL-GFPs, while their apparent amount seemed to be similar, by GFP signal. A possibility of the inaccessibility of the

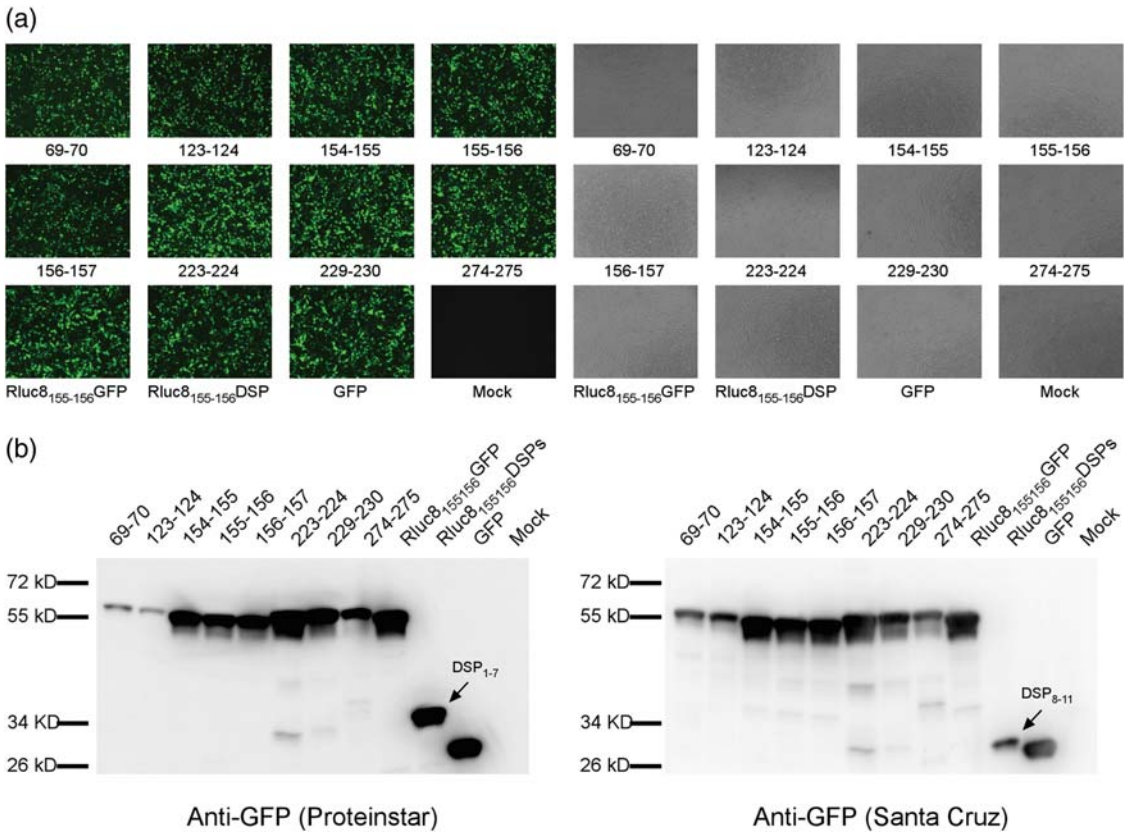

Fig. 2. Expression of RL–GFPs in 293FT cells. (**a**) GFP signal and bright field image of RL–GFP constructs transfected into 293FT cells. Numbers below the image indicate insertion points of GFP within RL. (**b**) Immunoblotting analysis of RL-GFP constructs expressed in 293FT cells. Number above the blot indicates the insertion points of GFP within RL. Antibody used in the left panel recognizes an epitope in GFP_{1-7}. Antibody used in the right panel recognizes an epitope in GFP_{8-11}. Band corresponding to $Rluc8_{155-156}DSP_{1-7}$ is shown as DSP_{1-7}; band $Rluc8_{155-156}$ is DSP_{8-11}. Position of the molecular weight marker is on the left.

Table I. RL activity of RL–GFP constructs

Insertion point[a]	Average SA (%)[b]	RL activity (million RLU)[c]
69-70	38	0.04 (±0.01)[d]
123-124	0	0.02 (±0.005)
153-154	50	39 (±4)
154-155	75	155 (±11)
155-156	42	643 (±6)
156-157	21	16 (±2)
157-158	60	12 (±1)
208-209	74	16 (±1)
223-224	5	0.03 (±0.006)
227-228	82	42 (±5)
229-230	31	69 (±0.5)
248-249	77	9 (±1)
274-275	35	0.2 (±0.06)
Intact RL	–	874 (±88)
GFP	–	0.04 (±0.006)

RLU, relative light units.
[a]Numbers indicate insertion points of GFP within RL.
[b]Indicates average solvent accessibility (SA) of the two adjacent residues at the insertion point.
[c]Each construct was transfected into 293FT cells and RL activity was measured.
[d]Number in parentheses indicates SD of three measurements.

anti-GFP epitope in $RL_{69-70}GFP$ and $RL_{123-124}GFP$ was unlikely because the results with two different anti-GFPs with different epitope specificities were similar (Fig. 2b, left and right). Another possibility is that while $RL_{69-70}GFP$ and $RL_{123-124}GFP$ were more soluble and generated GFP signals within the cell, they became insoluble and were lost into the precipitate fraction during sample preparation. Indeed, cell lysate samples from $RL_{69-70}GFP$ and $RL_{123-124}GFP$ tended to generate more green precipitate than other samples. $RL_{154-155}GFP$, $RL_{155-156}GFP$ and $RL_{229-230}GFP$ showed high RL activity and $RL_{69-70}GFP$, $RL_{123-124}GFP$, $RL_{223-224}GFP$ and $RL_{274-275}GFP$ showed low RL activity (Table I, RL activity). Contrary to our hypothesis, the observed RL activity and the value of SA did not clearly correlate (Fig. 3a). Among the constructs, $RL_{155-156}GFP$ showed the highest activity, but the average SA of residues 155 and 156 was moderate (42%). The insertion point of $RL_{229-230}GFP$ also had a moderate average SA (31%), but showed the third highest activity, which was significantly higher than RL-GFPs with highly exposed insertion points: $RL_{208-209}GFP$ (74%), $RL_{227-228}GFP$ (82%) and $RL_{248-249}GFP$ (77%). The RL-GFPs with a low SA insertion point, $RL_{123-124}GFP$ (0%) and $RL_{223-224}GFP$ (5%), showed no RL activity. Of note is that $RL_{155-156}GFP$ and $RL_{156-157}GFP$, whose GFP insertion points differed by only one amino acid, had differences in RL activity of more than 40 times (Table I). The reason for this was not clear, but was not because of differences in the expression level (Fig. 2b).

Generation of new split RL pairs

We found several points that tolerated the insertion of the GFP module. These sites were potentially good new split

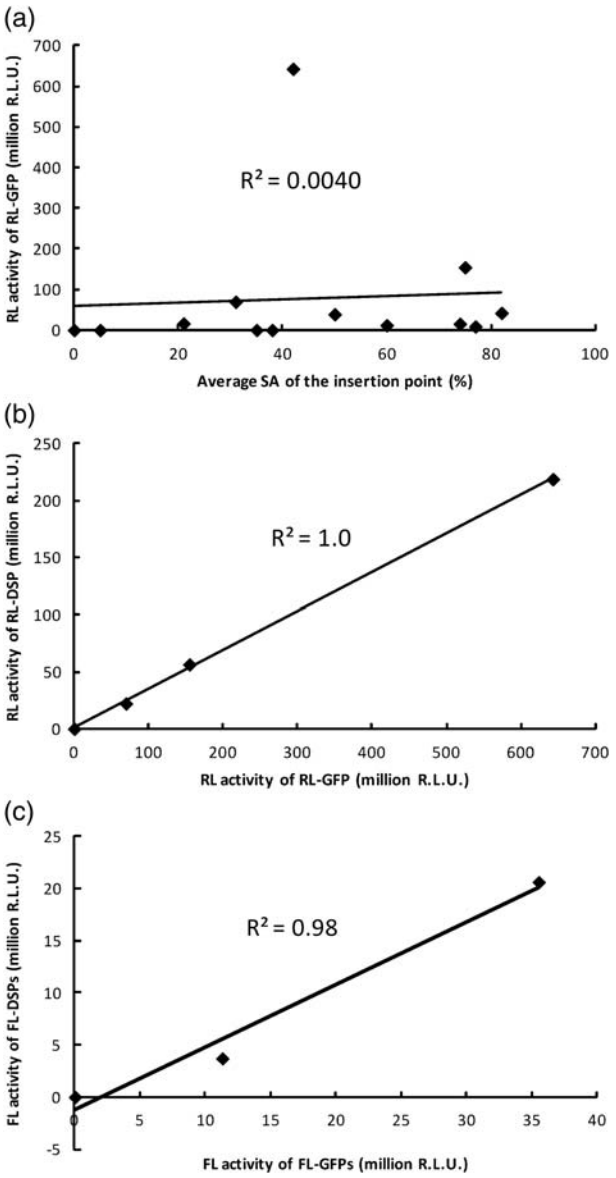

Fig. 3. RL activity of RL−GFPs. Correlation between (**a**) RL activity and SA values of the RL-GFPs. (**b**) RL activity of RL−GFPs and resultant RL−DSPs. (**c**) FL activities of FL−GFPs and FL−DSPs. R^2 indicates correlation efficiency.

points (between amino acids 154-155 and 155-156), so we made the split version of the proteins. Constructs were cotransfected into 293FT cells and GFP and RL signals were analyzed. The GFP signal confirmed protein expression (data not shown) and RL activity was measured (Table II). $RL_{155\text{-}156}DSP$ pair ($RL_{155\text{-}156}DSP_{1\text{-}7}$ with $RL_{155\text{-}156}DSP_{8\text{-}11}$) showed the highest activity, which was 10 times higher than the RL−DSP pair with a previously used split point between residues 229 and 230 (Kondo *et al.*, 2010). The cotransfected $RL_{155\text{-}156}DSP$ pair retained ∼25% of the intact RL activity. We also changed the RL portion to a Rluc8, which has four times higher activity than the wild-type RL (Loening *et al.*, 2006). The $Rluc8_{155\text{-}156}DSP$ pairs showed ∼30 times higher RL activity than $RL_{229\text{-}230}DSP$. The cellular distribution of the Rluc8 construct seemed to be more homogenous, with no dense green spots as seen in the original RL construct (Supplementary Fig. S1). Therefore, the Rluc8 construct seemed to express proteins that were more soluble than the original RL.

Table II. RL activities of split RL−DSP constructs

Split point[a]	RL activity (million RLU)[b]		
	DSP1-7	DSP8-11	Cotransfection
123-124	0.002 (± 0.002)[c]	0.0002 (± 0.0001)	0.006 (± 0.003)
154-155	0.008 (± 0.006)	0.0005 (± 0.0002)	57 (± 0.2)
155-156	0.02 (± 0.003)	0.0006 (± 0.0003)	219 (± 22)
229-230[d]	0.006 (± 0.005)	0.002 (± 0.002)	22 (± 0.2)
Rluc8 155-156	0.008 (± 0.007)	0.002 (± 0.0005)	697 (± 47)

RLU, relative light units.
[a]Numbers indicate the split points of RL in the RL−DSP constructs.
[b]Each construct was transfected individually or cotransfected into 293FT cells and RL activity was measured.
[c]Number in parentheses indicates SD of three measurements.
[d]Kondo *et al.* (2010).

The RL activities of RL−GFPs and the resultant split RL−DSP pairs correlated well (Fig. 3b). Therefore, splitting at the candidate points identified by GFP insertion resulted in functional split proteins. Neither $RL_{155\text{-}156}DSP_{1\text{-}7}$ nor $RL_{155\text{-}156}DSP_{8\text{-}11}$ showed any activity independently (Table II), so the $RL_{155\text{-}156}DSP$ regained RL activity only after reassociation of the DSP pairs. Also, the split RL construct without split GFP portion showed no RL activity (Supplementary Table SI).

Application of the method to other proteins

After successfully applying the GFP-scanning method to RL, we next tested the method in other proteins. First, we applied this method to HaloTag protein. Although the structure of HaloTag is not yet solved, the tertiary structure is expected to be similar to RL. HaloTag has 92% sequence identity with *Rhodococcus rhodochrous* haloalkane dehalogenase DhaA (PDB code 3FBW), which can be superimposed on the Rluc8 structure with Z-score of 40.8, and an r.m.s.d. of 1.8 Å for 287 Cα atoms using DaliLite (Holm and Park, 2000). Therefore, we split HaloTag at points corresponding to RL (amino acids 139-140, 194-195, and 212-213 in HaloTag correspond to 155-156, 208-209, and 227-228 in RL) with an additional split point (155-156). We made each GFP-inserted HaloTag expression vector and transfected it into 293FT to check that these positions in HaloTag could accommodate GFP insertion. All constructs showed strong GFP signal and could be stained by TMRDirect ligand, which is a membrane permeable fluorescent ligand for HaloTag protein (Ohana *et al.*, 2011) (data not shown). Staining efficiency was similar among constructs, which might be because an excess of ligand was used with a prolonged incubation time. On the basis of these observations, we tested only $Halo_{155\text{-}156}GFP$ for making a split protein pair. $Halo_{155\text{-}156}DSPs$ ($Halo_{155\text{-}156}DSP_{1\text{-}7}$ and $Halo_{155\text{-}156}DSP_{8\text{-}11}$) had no GFP activity and could not be stained with TMR ligand individually. However, it regained both activities when cotransfected into 293FT cells (Fig. 4). TMR ligand signals were much brighter than GFP signals when pictures were taken with same exposure time under the same microscope (Fig. 4).

We next applied the method to FL, which has a completely different structure from the previous two proteins. Potential split points were selected based on visual inspection

of the structure (PDB code 1BA3) and on the SA value, similar to the RL case. All candidate split points had high SA values (Table III). Another split point between residues 264 and 265, with 0% SA, was chosen as a negative control. FL–GFP constructs were made and transfected into 293FT cells, and FL activity was measured. $FL_{233\text{-}234}$GFP showed the highest activity among the FL–GFP constructs. No clear correlation was observed between FL activity and SA, similar to the RL–GFPs. Several FL–DSP constructs were made and the $FL_{233\text{-}234}$DSP pair showed the highest activity among the tested constructs (Table IV). FL–DSP showed neither FL nor GFP activity individually. FL activities of GFP inserted- and split-FL correlated well with FL activity (Fig. 3c). Therefore, the positive correlation of the activity between GFP inserted and split proteins was not limited to RL.

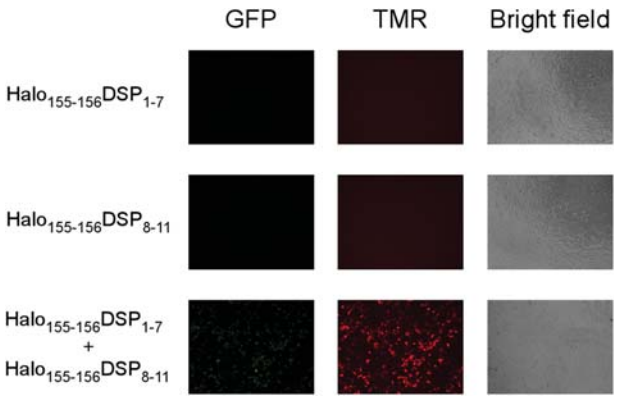

Fig. 4. GFP and TMR signal of $Halo_{155\text{-}156}$DSPs. At 48 h after transfection, GFP and TMR signal, as indicated, were examined using a IX71 microscope (10 × objective lens). Bright field images are also shown. Transfected construct names are at left.

Table III. FL activities of FL–GFPs constructs

Insertion point[a]	Average SA (%)[b]	FL activity (million RLU)[c]
174-175	105	11 (±1)[d]
233-234	87	36 (±3)
264-265	0	0.5 (±0.08)
377-378	85	4 (±0.4)
445-446	82	0.03 (±0.004)
489-490	99	9 (±1)
Intact FL	–	113 (±17)
GFP	–	0.00001 (±0.000001)

RLU, relative light units.
[a]Numbers indicate the insertion points of GFP within FL.
[b]Indicates average solvent accessibility (SA) of the two adjacent residues at the insertion point.
[c]Each construct was transfected into 293FT cells and FL activity was measured.
[d]Number in parentheses indicates SD of three measurements.

Application of the new DSPs for monitoring membrane fusion

We used the new $RL_{155\text{-}156}$DSP pairs for monitoring membrane fusion (Fig. 5a) using HIV Env-mediated cell–cell fusion as performed previously (Kondo *et al.*, 2010). The $Rluc8_{155\text{-}156}$DSP pair showed 100 times higher activity than the original $RL_{229\text{-}230}$DSP pair. With the new $Rluc8_{155\text{-}156}$DSP, we

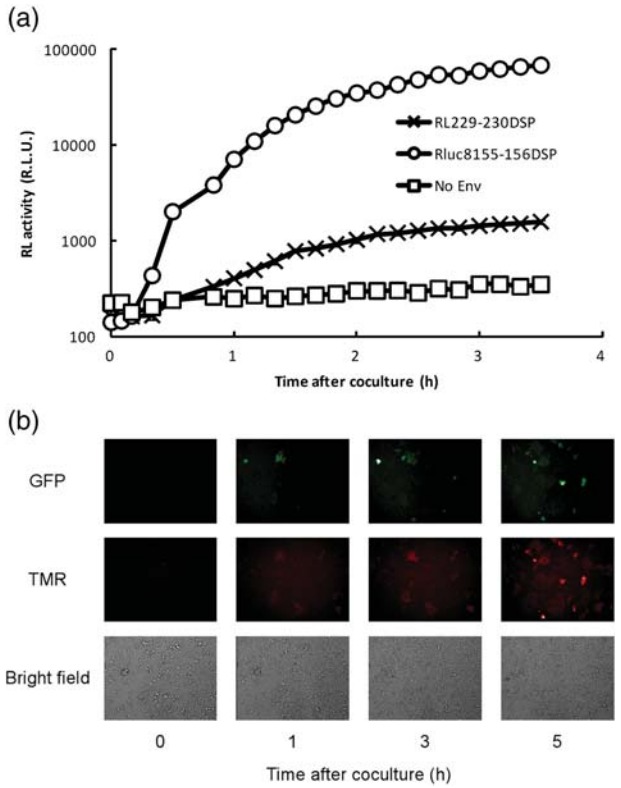

Fig. 5. Membrane fusion monitored by the new DSP (**a**) DSP assay of HIV-1 Env-mediated membrane fusion with original DSP and new $Rluc8_{155\text{-}156}$DSP. Error bars indicate standard deviation of three measurements. RLU, relative light units. (**b**) Membrane fusion assay induced with HIV-1 Env using Halo-DSPs. GFP, TMR signals and bright field images are shown.

Table IV. FL activities of the split FL–DSP constructs

Split point[a]	FL activity (million RLU)[b]		
	DSP1-7	DSP8-11	Cotransfection
174-175	0.008 (±0.0009)[c]	0.001 (±0.0008)	4 (±0.2)
233-234	0.00002 (±0.000005)	0.001 (±0.0002)	20 (±2)
445-446	0.00005 (±0.00004)	0.00003 (±0.000007)	0.008 (±0.0008)

RLU, relative light units.
[a]Numbers indicate the insertion points of FL in the FL–DSP constructs.
[b]Each construct was transfected individually or cotransfected into 293FT cells and FL activity was measured.
[c]Number in parentheses indicates SD of three measurements.

can detect fusion activity of lower fusogenic Env mutants because of the high sensitivity.

$Halo_{155\text{-}156}$DSP pairs were also used to monitor membrane fusion (Fig. 5b). When 293FT and 293CD4 cells fused, fused cells could be stained with the TMR ligand because of reassociation and functional recovery of Halo-DSPs. The TMR ligand signal appeared faster and brighter than the GFP signal. As HaloTag staining is more stable than GFP at low pH, the split HaloTag could be used to detect membrane fusion process in more acidic conditions, such as membrane fusion at endosomes.

Discussion

We developed a new method for designing a pair of split proteins that uses GFP like a scanning device. This method allows fast and easy screening of potential split points. The generated split proteins have a strong self-association activity through split GFP modules and recover the dual functions of the split protein and GFP upon reassociation.

The suitability of potential split points was examined by inserting a GFP module at candidate sites. We employed a two-step method to achieve GFP insertion: insertion of the linker, then cloning of the GFP gene. We found good correlation between the constructs with a linker constructs and GFP inserted, for both RL and FL (Supplementary Fig. S2a and b). Therefore, insertion of a short linker sequence might be sufficient by itself to examine the suitability of the split points. If the split proteins retain self-association capacity or if strong self-association via the split GFP module is not required, the suitability of the split point can be examined by splitting the protein at the identified point(s). If the self-associating split GFP will be used, direct insertion of the GFP gene using the megaprimer QuikChange method (Makarova *et al.*, 2000) could shorten the required time. In our hands, however, the success rate was higher for the two-step method than the megaprimer approach. A method using transposons to insert GFP randomly into the target gene (Gregory *et al.*, 2010) might also be preferable for larger-scale screenings.

Inserted GFP was as an easily detectable marker for expression of the engineered protein in live cells. This avoided other, more time-consuming screening methods such as immunofluorescence or immunoblotting. As we used GFPopt (Cabantous *et al.*, 2005; Pedelacq *et al.*, 2006), which has an extremely efficient folding and solubility, the GFP module seemed to fold independently of the folding status of the rest of the protein. For example, $RL_{123\text{-}124}$GFP had almost no RL activity (Table I), which might indicate that the RL portion was not properly folded, but the GFP signal was as strong as other RL-GFPs (Fig. 2a). Use of GFPopt was advantageous because the GFP signal indicated the total amount of expressed protein regardless of the folding status of the non-GFP portion(s).

Compared with other self-association modules, split GFP has some unique characteristics. It not only facilitates the association of the attached proteins, but also regains GFP activity upon reassociation. Some split proteins self-associate (e.g. beta-galactosidase), and do not require the attachment of a self-associating module. The use of split GFP as a self-associating module, however, provides the split proteins with dual functions: that of the target protein and GFP. The GFP signal can be used to provide positional information within cells such as the site of membrane fusion, as described previously (Kondo *et al.*, 2010). If generation of the GFP signal from the reassociated protein is not desirable, the split GFP chromophore can be inactivated or mutated. Split BFP (blue fluorescent protein) with a mutated chromophore self-associated with a similar efficiency (data not shown).

When connecting a split protein and self-association module, the order of connection and the sequence and length of the connecting linker must be carefully designed. For example, a pair of parallel coiled-coil motifs, such as those in the Velcro required a long linker to obtain the correct orientation of the split proteins in some cases (Kondo *et al.*, 2010). For a MyoD/Id pair whose associated structure was unknown, it was difficult to determine the optimum order and linker length *a priori* (Paulmurugan and Gambhir, 2003). However, the GFP module could be successfully inserted into target proteins using a short linker because the N- and C-termini of GFP are spatially close. Since the activity of the split DSPs and GFP-inserted proteins correlated well (Fig 3b and c), the self-associated DSP might mimic the structure of RL with inserted GFP. Split proteins can be generated by simply dividing the protein with the GFP insert within the GFP domain. The resulting split protein has the order [N-terminal portion of X]-$[GFP_{1\text{-}7}]$ and $[GFP_{8\text{-}11}]$-[C-terminal portion of X], and should be able to retain a favorable orientation for GFP self-association.

The reassociation mediated by split GFPopt is rapid and strong; the RL activity in DSP was recovered within 7 min (Kondo *et al.*, 2010). A split intein system, on the other hand, in which an intact reporter protein is regenerated by protein splicing, can take about an hour to splice (Ozawa, 2006). By cotransfection, the DSP protein recovered ~25% of the intact RL and 27% of intact FL, comparable to other self-association modules (Paulmurugan and Gambhir, 2003). The association of split GFP is irreversible in most cases (Kerppola, 2006). This feature is desirable when event accumulation needs to be monitored. The reassociated protein will not dissociate by the dilution of the reaction system.

Currently, no established strategy determines candidate split points in a target protein. The structure of the protein might or might not be available. For example, the split point between amino acids 229 and 230 in RL was identified even though the structure of RL was not determined (Paulmurugan and Gambhir, 2003). In this study, we selected new candidate split points by visual inspection of the available structure and the SA value. However, the actual results and the prediction were not clearly correlated for either RL or FL (Fig. 3a and Tables I and III). The only exception was that the introduction of split points at completely buried residues did result in not favorable pairs.

The best split point for RL in this study was between residues 155 and 156, which falls within a loop region in the determined structures. The loop region (residues 153–163) cannot be observed in one of the five Rluc8 structures, and another has a large conformational change (Loening *et al.*, 2007). This region is suggested to be flexible for substrate binding in the structurally similar haloalkane dehalogenases (Schanstra and Janssen, 1996). This loop connects the α/β hydrolase domain and cap domain. Another favorable split point ($RL_{229\text{-}230}$DSP) was localized in another loop (residues 223–229) connecting the two domains. For FL, the preferred

split point was also in the region between two domains (Ozawa, 2006). As the loop has higher *B*-factors than the other region of Rluc8 (Loening *et al.*, 2007), one parameter to be considered in selecting candidate sites is the *B*-factors and/or flexibility of the residues in the determined structure. However, the difference in activity was significant for split proteins within the short segments ($RL_{155\text{-}156}$GFP was four times higher than $RL_{154\text{-}155}$GFP and 60 times higher than $RL_{156\text{-}157}$GFP). Almost no difference was observed in the average *B*-factor in this region (averages for Cα were 30.4 for residues 154–155, 30.2 for 155–156, and 29.8 for 156–157). Therefore, pinpointing the best split point based solely on the *B*-factor is difficult. The best way to identify an optimum split point is to target a flexible loop region connecting domains in the protein of interest, and examine several split points within the region. Therefore, a number of trials are needed even if the protein structure is known. If structural information is not available for a protein of interest, more sample screening is necessary. Therefore, for a method like GFP scanning, a simple strategy of inserting GFP to search for the split point instead of making two independent expression vectors for a pair of candidate split proteins, reduces the work. Although our approach was aimed at making pairs of split proteins, the results could provide information on the flexibility of the inserted regions whose structure is not solved.

Newly identified split RL–DSP was suitable for monitoring membrane fusion process or organelle targeting of proteins. The new split RL, $Rluc8_{155\text{-}156}$DSP, had 100 times higher RL activity and detected a lower incidence of membrane fusion than a previous version. Similarly, Halo-DSP could monitor membrane fusion in acidic conditions.

Supplementary data

Supplementary data are available at *PEDS* online.

Acknowledgements

The authors thank Dr Kunito Yoshiike for critical reading of the manuscript. Rluc8 plasmid was a kind gift from Dr Sanjiv Sam Gambhir.

Funding

This work was supported by a contract research fund from the Ministry of Education, Culture, Sports, Science, and Technology for Program of Japan Initiative for Global Research Network on Infectious Diseases.

References

Barbeau,B., Fortin,J.F., Genois,N. and Tremblay,M.J. (1998) *J. Virol.*, **72**, 7125–7136.
Blumenthal,R., Gallo,S.A., Viard,M., Raviv,Y. and Puri,A. (2002) *Chem. Phys. Lipids*, **116**, 39–55.
Cabantous,S., Terwilliger,T.C. and Waldo,G.S. (2005) *Nat. Biotechnol.*, **23**, 102–107.
Chothia,C. (1976) *J. Mol. Biol.*, **105**, 1–12.
Fields,S. and Song,O. (1989) *Nature*, **340**, 245–246.
Franks,N.P., Jenkins,A., Conti,E., Lieb,W.R. and Brick,P. (1998) *Biophys. J.*, **75**, 2205–2211.
Gregory,J.A., Becker,E.C., Jung,J., Tuwatananurak,I. and Pogliano,K. (2010) *PLoS ONE*, **5**, e8731.
Holland,A.U., Munk,C., Lucero,G.R., Nguyen,L.D. and Landau,N.R. (2004) *Virology*, **319**, 343–352.
Holm,L. and Park,J. (2000) *Bioinformatics*, **16**, 566–567.
Kerppola,T.K. (2006) *Nat. Rev. Mol. Cell. Biol.*, **7**, 449–456.
Kim,S.B., Ozawa,T., Watanabe,S. and Umezawa,Y. (2004) *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 11542–11547.
Kondo,N., Miyauchi,K. and Matsuda,Z. (2011) *Current Protocols in Cell Biology*. John Wiley & Sons, Inc., Hoboken.
Kondo,N., Miyauchi,K., Meng,F., Iwamoto,A. and Matsuda,Z. (2010) *J. Biol. Chem.*, **285**, 14681–14688.
Liu,S., Kondo,N., Long,Y., Xiao,D., Iwamoto,A. and Matsuda,Z. (2010) *Retrovirology*, **7**, 100.
Loening,A.M., Fenn,T.D. and Gambhir,S.S. (2007) *J. Mol. Biol.*, **374**, 1017–1028.
Loening,A.M., Fenn,T.D., Wu,A.M. and Gambhir,S.S. (2006) *Protein Eng. Des. Sel.*, **19**, 391–400.
Long,Y., Meng,F., Kondo,N., Iwamoto,A. and Matsuda,Z. (2011) *Protein Cell*, **2**, 369–376.
Los,G.V. and Wood,K.V. (2006) *Methods in Molecular Biology*. Humana Press, Inc., Totowa, NJ, pp. 195–208.
Magliery,T.J., Wilson,C.G.M., Pan,W., Mishler,D., Ghosh,I., Hamilton,A.D. and Regan,L. (2004) *J. Am. Chem. Soc.*, **127**, 146–157.
Makarova,O., Kamberov,E. and Margolis,B. (2000) *BioTechniques*, **29**, 970–972.
Miyauchi,K., Komano,J., Yokomaku,Y., Sugiura,W., Yamamoto,N. and Matsuda,Z. (2005) *J. Virol.*, **79**, 4720–4729.
Ohana,R.F., Hurst,R., Vidugiriene,J., Slater,M.R., Wood,K.V. and Urh,M. (2011) *Protein Expr. Purif.*, **76**, 154–164.
Ozawa,T. (2006) *Anal. Chim. Acta*, **556**, 58–68.
Ozawa,T., Nishitani,K., Sako,Y. and Umezawa,Y. (2005) *Nucleic Acids Res.*, **33**, e34.
Ozawa,T., Sako,Y., Sato,M., Kitamura,T. and Umezawa,Y. (2003) *Nat Biotech.*, **21**, 287–293.
Paulmurugan,R. and Gambhir,S.S. (2003) *Anal. Chem.*, **75**, 1584–1589.
Paulmurugan,R., Umezawa,Y. and Gambhir,S.S. (2002) *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 15608–15613.
Pedelacq,J.D., Cabantous,S., Tran,T., Terwilliger,T.C. and Waldo,G.S. (2006) *Nat. Biotech.*, **24**, 79–88.
Remy,I. and Michnick,S.W. (2006) *Nat. Meth.*, **3**, 977–979.
Schanstra,J.P. and Janssen,D.B. (1996) *Biochemistry*, **35**, 5624–5632.
Shekhawat,S.S. and Ghosh,I. (2011) *Curr. Opin. Chem. Biol.*, **15**, 789–797.
Tsodikov,O.V., Record,M.T. and Sergeev,Y.V. (2002) *J. Comput. Chem.*, **23**, 600–609.
Ullmann,A., Jacob,F. and Monod,J. (1967) *J. Mol. Biol.*, **24**, 339–343.
Wang,J., Kondo,N., Long,Y., Iwamoto,A. and Matsuda,Z. (2009) *J. Virol. Methods*, **161**, 216–222.