Cloning and Transgenesis
ISSN: 2168-9849

Zhang, Clon Transgen 2014, 3:3
DOI: 10.4172/2168-9849.1000131

Research Article Open Access

Construction of Functional GFP-Neo Fusion Protein by Selection from a Peptide Library

Mingdong Liu and Jiayou Zhang*

Department of Microbiology, Immunology and Molecular Genetics, University of Kentucky, KY USA

***Corresponding author: Jiayou Zhang,** University of Kentucky 800 Rose St. Lexington, KY, USA, Tel: +1-859 257-4456; E-mail: jzhan1@uky.edu

Rec Date: September 30, 2014, **Acc Date:** November 28, 2014, **Pub Date:** November 30, 2014

Abstract

Almost all retroviral vectors need to express more than one protein. One of the strategies to express multiple proteins is to construct a fusion protein that encodes two genes from a single reading frame. Since the interaction of two closely linked proteins is poorly understood, there is no guarantee that both of the proteins will be functional when their open reading frames are connected in frame. We inserted random nucleotides encoding peptides between sequences of the green fluorescence protein and the neomycin resistance gene. This peptide library was then transformed into bacteria, and the transformed bacteria were selected for green fluorescence and antibiotic resistance. Several fusion proteins have been selected. When these fused genes are expressed by retroviral vectors, the GFP-Neo fusion proteins are also functional in eukaryotic cells. The fluorescence intensity of our fusion proteins is much greater than a commercial GFP fusion clone, although it is still less than the wild-type GFP. The isolation of a fusion protein described here can be modified and used to isolate other fusion proteins as long as selections for their functions are available.

Introduction

Retroviral vectors are designed to deliver genes. In addition to the gene that needs to be delivered, retroviral vectors need to encode a marker gene in order to determine viral titer as well as for monitoring infected cells after infection. In addition, some projects need to deliver one or more regulatory genes into cells. Therefore, the construction of retroviral vectors frequently requires the expression of multiple genes within a single vector. Retroviral vectors contain two long terminal repeats (LTR), which contain the cis-required sequences. The genes to be expressed are encoded between the two LTRs. These LTRs contain the viral promoter and enhancer so that the first gene can be inserted downstream of the 5' LTR for expression by the standard ribosome scanning model [1]. In order to express more than one gene in a single vector, there are five strategies that can be used to express a second gene:

(1) A splicing donor and splicing receptor can be constructed so that the second gene will be expressed via a spliced subgenomic RNA. Since the sequence of the first gene becomes the intron for the second gene, the efficiency of splicing is difficult to predict [2].

(2) A promoter can be inserted between the first and second gene so that the second gene is expressed from the inserted promoter. Unfortunately, a retroviral vector's size is limited to 7-10-kb. A well-designed retroviral vector must maximize the packaging capacity of the retroviral sequences to express both marker genes and therapeutic or other interest genes. It has been demonstrated that two promoters in close proximity often interfere thereby resulting in a low titer and /or poor expression of the second gene [3].

(3) An internal ribosome entry segment (IRES) can be inserted between the two genes so the second gene is expressed by the IRES [4]. This strategy is widely used to express two genes, however, the insertion of two IRES sequences within a single vector, frequently recombine results in a deletion of the sequence between the two identical IRES [5,6]. Moreover, an IRES sequence is almost as long as a coding reading frame.

(4) A picornavirus "self-cleaving" 2A peptide can be used to generate multiple proteins [7], and two of the identical sequences also can recombine resulting in deletion.

(5) The final strategy is that a vector with fusion proteins is able to encode more functional products than a vector using either promoter or an IRES.

Almost all fusion proteins successfully constructed thus far are simply a connection between two open reading frames (ORF) using restriction digestion and in-frame ligation [8-14]. While straightforward, this traditional approach, which is based on a trial-and-error method, has failed to produce functional proteins in several attempts. A fusion protein gene can be divided into three parts: two protein coding regions and a peptide linker that serves as a bridge between the two proteins. Since the mechanism of a peptide linker connecting the two proteins is poorly understood, we inserted the library of nucleotides to encode a peptide linker, which is made up of random amino acids. Although we found that only 1% of the library can encode functional GFP-Neo proteins, many peptides can act a bridge to connect functional GFP and Neo proteins. In one instance the fusion proteins analyzed resulted in a fusion protein that had GFP and was almost 30% as bright as a wild-type of GFP and fully functional with respect to the neomycin resistance gene.

Methods

Vector constructions

All recombinant techniques were carried out according to conventional procedures [12]. pML41 is a bacterial expression vector (Figure 1A) based on pGFP (Clontech, Pale Alto, CA). The gfp gene within the pGFP was replaced with the gfp sequence from vector

pEFGP-C1 (Clontech) so that the stop codon of gfp was replaced with a multiple cloning site including XhoI. The neo gene was derived from pJZ419 (Figure 2A), which has been previously described [15]. pJZ419 encoded a gfp gene and a neo gene. The start codon of the neo gene has been mutated so that it can be digested with NdeI (CATATG). The sequence pML41-7 (clone 7) has been submitted to GenBank (accession number AY596277). The pML41 was digested with XhoI and NdeI, the 4.5-kb NdeI-XhoI fragment was isolated from an agarose gel (Figure 1B). This fragment was used to ligate an artificially engineered double stranded DNA (Figure 1E).

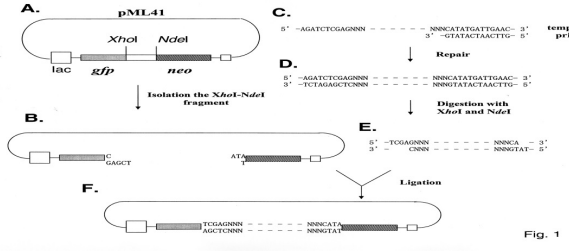

Figure 1: Construction of the GFP-Neo fusion protein. A. A bacterial expression vector, which encodes the gfp and neo sequence. pML41 encodes the bacterial lac promoter, the gfp gene and the neo gene. B. The stop codon of the gfp gene has been deleted and its sequence is ended by the restriction site XhoI. The start codon of the neo gene is within the restriction site NdeI. C. The template and primer oligonucleotide. The template oligonucleotide includes a 5' end with the XhoI site and a 3' end with the NdeI site. 50N (N represents a combination of A, T, C and G.) located between the two ends. The primer oligonucleotide is complementary to the 3' end of the template. D. The formation of the double-stranded DNA after repair by Klenow large fragment. E. Digestion of the double-stranded DNA with XhoI and NdeI. F. Ligation between the vector and random nucleotides.

Sequences of the template and the primer oligonucleotide are described below in the last section of Methods. An equal amount of the two oligonucleotides (50 pmol) was mixed in water and heated to 95°C for 5 min. The two oligonucleotides then were annealed in 24°C for 10 min (Figure 1C). A Klenow large fragment was used to repair the single stranded part of the template into a double-stranded DNA according to the procedure provided by the manufacturer (Figure 1D). The ds DNA was then isolated from an agarose gel and digested with XhoI and NdeI (Figure IE).

ML42 (Figure 2C) is a retroviral vector similar to JZ419 [16], except that pJZ419 encoded a separated gfp gene, neo gene and an IRES sequence between the two genes, while the pML42 encodes the gfp-neo fusion protein without an IRES. The BamHI-RsrII (BamHI is located upstream of the start codon of the gfp gene, while the RsrII is located within the neo gene) fragment from each clone of pML41 was inserted between the BamHI and RsrII sites of the pJZ419. The sequence of pML42 is available upon request.

Cell, transfection and infection

PA317 is an NIH3T3-derived Moloney leukemia virus (MLV) based helper cell line. D17 is a dog osteosarcoma cell line obtained from ATCC (CRL 8468). Transfection, infection, and maintenance of cells have been previously described [17,18].

Sequence analysis

The junction between the gfp and neo gene were sequenced using two primers. They are the upstream primer 5'-GACAACCACTACCTGAGCACC-3' and the downstream primer 5'-CGCTGACAGCCGGAACACGGC-3'.

Fluorescence microscopy and flow cytometric analysis

A fluorescence inverted microscope (Zeiss, Axiovert 25) with a mercury arc lamp (100 watt) and a fluorescent filter set (CZ909) consisting of a 470- to 40-nm exciter, a 515 nm emitter, and a 500 nm beamsplitter were used to detect GFP in living cells.

Flow cytometric analysis was conducted using a Dickinson FACSCalibur/30,000 events were acquired using the FL2-H histogram with the marker set at position 305 on the left. The FL-1 emission channel was used to monitor the intensity of the GFP.

Construction of a random peptide library between the gfp and the neo gene

A bacterial expression vector, pML41, based on pGFP (Clontech) has been constructed (GenBank, accession number AY596277) (Figure 1A). This expression vector contained a lac promoter, a green fluorescence protein gene (gfp), and a neomycin resistant gene (neo). The original gfp sequence in the pGFP was replaced with the gfp sequence from pEGFP-C1 (Clontech), in which the stop codon of the gfp gene was deleted. The neo gene was derived from pJZ419 [16] so that the start codon (ATG) of the neo gene was mutated into the NdeI site (CATATG). The gfp gene without a stop codon was bound at the 3' end by an XhoI site while the neo gene began with an NdeI site so that the gfp gene and the neo gene would be separated by the two restriction sites. This cloning vector was digested with XhoI and NdeI (Figure 1B) and ligated with an artificially engineered double stranded DNA digested with XhoI and NdeI (Figure 1E).

The artificially double-stranded DNA was made by hybridization of two oligonucleotides of different length (Figure 1C). The long oligonucleotide or template was annealed to a short oligonucleotide primer (Figure 1C). This template oligonucleotide was synthesized as 5'-AGATCTCGAG-50N-CATATGATTGAAC-3', where the N represents a combination of A, T, C and G. Therefore, this oligonucleotide is a population comprised of random nucleotides in the middle and bordered by an XhoI (CTCGAG) site at its 5' and an NdeI (CATATG) site at its 3' end. The primer (5'-GTTCAATCATATG-3') was complimentary to the template oligonucleotide at its 3' end and served as a primer for the Klenow fragment to polymerize a double-stranded DNA (Figure 1D). The double-stranded DNAs were then digested with XhoI and NdeI (Figure 1E) and inserted between the XhoI and NdeI sites of the cloning vector, pML41 (Figure 1B) so that the random sequences were located between the gfp and neo gene sequences (Figure 1F). Although, only 50N (about 17 amino acids) were between the XhoI and NdeI sites within these synthesized oligonucleotides, it is possible that multiple copies of the XhoI-NdeI fragments could be inserted between the XhoI and NdeI sites of ML41.

Results

Selection of kanamycin resistant proteins with green fluorescence

Plasmid DNA of pML41 (described in Methods) inserted with the random 50N (17 amino acids) were transformed into bacteria. Transformed bacteria were plated on dishes with ampicillin or kanamycin. The pML41 also encoded the ampicillin resistance gene. Therefore, colonies with pML41 containing any insertion between the XhoI and NdeI sites would be resistant to ampicillin. However, kanamycin-resistant colonies represent clones that encode a functional neo gene.

The number of ampicillin-resistant colonies is about 100 times that of kanamycin-resistant colonies, indicating that only a very small portion of these random peptides can serve as a functional bridge. All ampicillin-resistant colonies and all kanamycin-resistant colonies were green under a fluorescence microscope, indicating that the sequence downstream of the gfp gene did not totally destroy the function of the gfp gene. However, the fluorescence intensity of each clone was different (Table 1).

Clone intensity(a)	Percentage of survived cells in 2 mg/ml neomycin(b)	
D17	3.12 ± 0.56	NA
JZ419	586.92 ± 6.73	96%
JZ523	26.44 ± 3.32	0%
clone 1	14.36 ± 1.09	26%
clone 2	211.05 ± 9.31	22%
clone 3	13.58 ± 3.00	60%
clone 4	352.22 ± 7.13	1%
clone 5	16.86 ± 2.00	38%
clone 6	257.03 ± 10.20	25%
clone 7	168.16 ± 12.65	88%
clone 8	14.12 ± 1.18	62%

Table 1: Fluorescence intensity and neomycin resistance of GFP-Neo fusion proteins. (a) D17 cells infected with each vector virus. Infected cells were selected for neomycin resistance in media with 1 mg/ml of neomycin. Neomycin-resistant colonies formed 10-12 days after selection. The neomycin-resistant cells were pooled and analyzed by flow cytometry. 30,000 cells were required for each analysis. The intensity of each clone is the average of three analyses. D17 cells represent the target cells that were not infected. JZ419 represents D17 cells infected with JZ419 virus, which encodes separate gfp and neo genes. JZ523 encodes a hyg-gfp fusion protein gene. D17 cells infected with JZ523 were selected for hygromycin resistance before analysis by flow cytometry. (b) The D17 cells were infected with each virus. The green cells were selected by flow cytometry two days after infection. Green cells were then cultured in media with different concentration of neomycin.

Determination of the relative functions of green fluorescence and neomycin resistance of the independently derived fusion proteins

To exclude that the random sequence encodes a bacterial promoter, sequences of eight kanamycin-resistant clones were inserted into a retroviral vector. ML42 was constructed based on the Moloney leukemia virus (MLV) vector [18]. Between the two long terminal repeats (LTR), this vector encoded the gfp-neo sequences from the pML41. Plasmid DNAs of retroviral vectors were transfected into the retroviral helper cell line PA317 [18]. Viruses released from transfected PA317 cells were used to infect D17 cells [16]. The MOI (multiplicity of infection) was about 0.01. Infected D17 cells were selected for neomycin-resistance, and the neomycin resistant colonies were visible about 10 days after infection. More than 100 colonies were pooled, and the relative fluorescence of each GFP-Neo fusion protein from the pooled neomycin resistant cells was analyzed by flow cytometry. The green fluorescence intensity of each clone is shown in Table 1. D17 cells infected with JZ419 (Figure 2A) were used as a positive control. Between its two LTRs, the JZ419 virus encodes the gfp and the neo gene independently. An internal ribosome entry site sequence (IRES) is located between the two genes so that the gfp gene is expressed by a traditional scanning model and the neo gene is expressed by the IRES. The fluorescence intensity of wild-type gfp expressed from JZ419 infected cells is 584 units, while the intensity of the uninfected D17 cells was set at zero (Table 1). JZ523, another virus, was used as a negative control for neomycin resistance (Figure 2B). JZ523 encodes a Hyg-GFP fusion protein, which was derived from the pHygEGFP (Clontech). pHygEGFP encoded a hygromycin phosphotransferase (hygromycin resistance protein) and a gfp fusion protein. The fluorescence of the Hyg-GFP fusion protein is only about 4% of wild-type GFP expressed by the JZ419 virus. The intensities of the eight clones isolated were between 2% and 44% (Table 1).

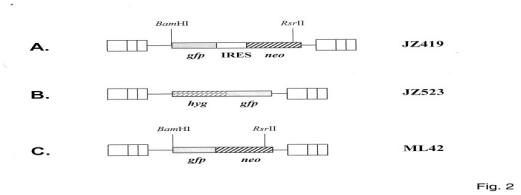

Figure 2: Retroviral vectors encoding the gfp and neo gene. A. JZ419 encodes separated wild-type gfp and neo gene. The gfp gene is expressed by traditional ribosomal scanning model and the neo gene is expressed by the internal ribosome entry segment (IRES). B. JZ523 encodes a Hyg-GFP fusion protein. C. ML42 encodes the GFP-Neo fusion protein.

In order to analyze the function of the neo gene, D17 cells were also infected with a virus derived from each fusion protein clone. Infected cells were selected by flow cytometry, not for neomycin resistance, but for their green color. The green cells were then cultured in a medium with a different concentration of neomycin. A medium containing 1 mg/ml of neomycin is usually used for selection for neomycin resistant D17 cells. The D17 cells infected with JZ523 (encoding Hyg-GFP fusion protein) were used as a negative control. No control cells survived after selection with 1 mg/ml neomycin

(Figure 3). About 96% of the cells infected with the wild-type neo gene (JZ419) survived when cultured with 2 mg/ml neomycin (Figure 3 and Table 1). For all clones encoding fusion proteins analyzed, 1% to 88% of cells survived in the medium containing 2 mg/ml neomycin. The best neomycin surviving clones were found to be clone 7 (88%) and clone 8 (62%) (Table 1 and Figure 3)

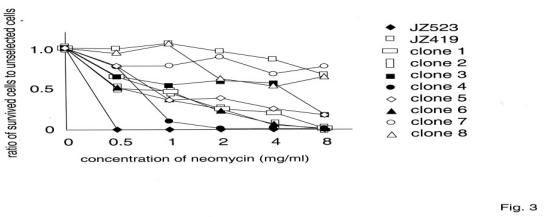

Figure 3: Determination of the neomycin resistance of GFP-Neo fusion proteins. D17 cells infected with each ML42 clone (Figure 2C). Infected cells were isolated using their green fluorescence for selection by flow cytometry. Green cells were cultured in media with different concentrations of neomycin. JZ523 encodes a Hyg-GFP fusion protein and D17 cells infected with JZ523 are neomycin sensitive. JZ419 encodes a separated gfp and neo sequence. D17 cells infected with JZ419 were used as a positive control. The ratio of the number of cells in each plate 12 days after selection to the number of cells in media without neomycin represents the function of the GFP-Neo protein. The result for JZ419 is the average of two analyses while that for clone 7 is the average of three analyses. The horizontal axis is for different concentration of neomycin, while the vertical axis is for ratio of survived cells to unselected cells which represent the neomycin resistance.

The sequences between the gfp and neo gene were analyzed and these two clones (clone 7 and 8) encoded peptides that did not contain stop codons between the two genes (clone 7: SRVAYCSSDMIVPLDEVGAH) (clone 8: SRVYIMCCVIVGLRTTRGCH). It seems no common properties can be found between the 17 residues of the two peptides.

Discussion

Although constructions of many fusion proteins have been reported [8-14], most are only the connection of the open reading frames of the two genes in frame. It is difficult to know the frequency of success by just connecting two genes in frame. We have tried several pairs and all failed. These poor results suggest a low chance that both genes will maintain functionality when they are directly connected. We have successfully constructed a GFP-Neo fusion protein by insertion of random peptides between the gfp and neo sequences, since gfp is the most frequently used color reporter gene and neo is the most frequently used drug resistance gene. This fusion gene will provide a handy tool for construction of retroviral vectors. If the two genes with pML41 are replaced with other gene pairs, this vector can generally be used for construction of other fusion proteins as long as the gene function can be selected in bacteria. It would also be possible to use the pML42 (MLV based vector) construct, and the random peptides could also be inserted between the XhoI and NdeI sites within the pML42 vector and selected for functionality in eukaryotic cells.

The neo genes of the two clones isolated were nearly equivalent in function when compared with the wild-type neo (Table 1 and Figure 3). Clearly clone 7 is the best candidate for a GFP-Neo fusion protein, because it provided wild-type neomycin drug resistance and its fluorescence was 30% that of wild-type GFP (Table 1). Several other clones contained multiple stop codons within the linker sequence. Their sequences are available upon request. It is not surprising that the functions of the neo gene for clones with stop codons were poor, but they were always superior to the absence of the neomycin gene (i.e. JZ523). This could be attributed to frame shift or perhaps to re-initiation of translation at a downstream ATG within the neomycin resistance sequence.

Our initial design for generating the random peptide sequence showed a propensity to generate many stop codons within the random peptides. Since there are three stop codons (TAA, TGA, and TAG) in the 64 codons, the sequence of a random N results in a stop codon from 21 (64/3) codons on average. Using 17 amino acids as the bridge between two proteins seemed to result in few problems. However, since there is no evidence indicating the best length of the bridge is 17 (or a multiple of 17), we need to construct several random sequences with different lengths and look for the best results.

Conclusions

We have successfully isolated a GFP-Neo fusion protein. This method can be modified and used to isolate other fusion proteins as long as selections for their functions are available. It should be much more feasible than just the connection of two genes in frame by restriction enzymes.

References

1. Kozak M (1983) Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol Rev 47: 1-45.
2. Dougherty JP, Temin HM (1986) High mutation rate of a spleen necrosis virus-based retrovirus vector. Mol Cell Biol 6: 4387-4395.
3. Emerman M, Temin HM (1986) Comparison of promoter suppression in avian and murine retrovirus vectors. Nucleic Acids Res 14: 9381-9396.
4. Adam MA, Ramesh N, Miller AD, Osborne WR (1991) Internal initiation of translation in retroviral vectors carrying picornavirus 5' nontranslated regions. J Virol 65: 4985-4990.
5. Zhang J, Ma Y (2001) Evidence for retroviral intramolecular recombinations. J Virol 75: 6348-6358.
6. Li T, Zhang J (2004) Stable expression of three genes from a tricistronic retroviral vector containing a picornavirus and 9-nt cellular internal ribosome entry site elements. J Virol Methods. 115: 137-44.
7. Szymczak AL, Workman CJ, Wang Y, Vignali KM, Dilioglou S, et al. (2004) Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. Nat Biotechnol 22: 589-594.
8. Oh SC, Nam SY, Kwon HC, Kim CM, Seo JS, et al. (2001) Generation of fusion genes carrying drug resistance, green fluorescent protein, and herpes simplex virus thymidine kinase genes in a single cistron. Mol Cells. 11: 192-7.
9. Hansen SG, Cope TA, Hruby DE (2002) BiZyme: a novel fusion protein-mediating selection of vaccinia virus recombinants by fluorescence and antibiotic resistance. Biotechniques 32: 1178, 1180, 1182-1187.
10. Bennett RP, Cox CA, Hoeffler JP (1998) Fusion of green fluorescent protein with the Zeocin-resistance marker allows visual screening and drug selection of transfected eukaryotic cells. Biotechniques. 4: 478-82.
11. Abbate J, Lacayo JC, Prichard M, Pari G, McVoy MA (2001) Bifunctional protein conferring enhanced green fluorescence and puromycin resistance. Biotechniques 31: 336-340.

12. Green MR, Sambrook J (2012) Molecular Cloning: A Laboratory Manual (4th edn.) Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
13. Andrades JA, Han B, Becerra J, Sorgente N, Hall FL, et al. (1999) A recombinant human TGF-beta1 fusion protein with collagen-binding domain promotes migration, growth, and differentiation of bone marrow mesenchymal cells. Exp Cell Res. 250:485-98.
14. Roudkenari MH, Bouzari S, Oloomi M, Jafari A, Shahrokhi N, et al. (2005) Expression of a Chimeric Protein Containing the Catalytic Domain of Shiga-Like Toxin and Human Granulocyte Macrophage Colony-Stimulating Factor (hGM-CSF) in Escherichia coli and Its Recognition by Reciprocal Antibodies. IBJ 9: 143-8.
15. Miller AD, Buttimore C (1986) Redesign of retrovirus packaging cell lines to avoid recombination leading to helper virus production. Mol Cell Biol 6: 2895-2902.
16. Guillen KP, Kurkjian C, Harrison RG (2014) Targeted enzyme prodrug therapy for metastatic prostate cancer - a comparative study of L-methioninase, purine nucleoside phosphorylase, and cytosine deaminase. J Biomed Sci 21: 65.
17. Sapp CM, Li T, Zhang J (1999) Systematic comparison of a color reporter gene and drug resistance genes for the determination of retroviral titers. J Biomed Sci 6: 342-348.
18. Zhang J, Temin HM (1993) Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication. Science 259: 234-238.