Minireview

Insertional gene fusion technology

Nobuhide Doi, Hiroshi Yanagawa*

Mitsubishi Kasei Institute of Life Sciences, 11 Minamiooya, Machida, Tokyo 194-8511, Japan

Received 2 June 1999

Abstract The classical 'end to end' gene fusion technique has widely been used for monitoring gene expression, biological screening and purification of recombinant proteins. Recent progress with the 'insertional' gene fusion approach, on the other hand, has demonstrated that this technique can be utilized for membrane protein topology analysis, display of randomized protein libraries and design of biosensor proteins. In this review, we describe examples of insertional gene fusion and compare the old and new gene fusion techniques.

Key words: Domain insertion; Protein engineering; Membrane protein topology; Directed evolution; Biosensor; Green fluorescent protein

1. Introduction

Since the development of recombinant DNA techniques in the late 1970s, gene fusion technology has been widely used for biological screening, recovery and purification of recombinant proteins [1,2]. Further, green fluorescent protein (GFP) has recently been applied as a marker for gene expression and protein localization by fusing its gene to a target gene of interest [3,4]. In order to construct such fusion proteins, two types of connection are possible (Fig. 1). One is 'end to end' fusion in which the N-terminus of one domain is linked to the C-terminus of the other domain. The second is 'insertional' fusion in which one domain is inserted in-frame into the middle of the other parent domain. So far, almost all fusion proteins have been constructed by the end to end fusion approach, but several examples of insertional gene fusion have recently been reported. In this review, we focus on insertional fusion proteins and discuss the advantages and disadvantages of the insertional fusion technique.

2. Construction of insertional fusion proteins

Table 1 lists insertional fusion proteins that have been prepared so far by inserting a segment of more than 100 amino acid (aa) residues into intact proteins. These insertional fusion proteins can be classified into three categories according to their experimental purposes (Fig. 2).

2.1. Analysis of membrane protein topology

The insertional (or 'sandwich') gene fusion technique was first described by Ehrmann et al. [5] to determine membrane protein topologies more precisely than was possible with the classical end to end fusion approach. They inserted alkaline phosphatase (AP) as a reporter protein into various cytoplasmic and periplasmic loops of a multispanning membrane protein, MalF. AP exhibited a high activity when it was fused into a periplasmic loop of MalF, whereas the activity was very low when AP was inserted into a cytoplasmic loop of the membrane protein (Fig. 2A). Differing from the classical end to end fusions of AP with truncated membrane proteins [6], the insertional fusion proteins contain the whole sequence of the membrane protein and hence, this technique is a more sensitive monitor of membrane protein topology. To date, this AP insertional fusion approach has been applied to many integral membrane proteins, such as the human β2-adrenergic receptor [7], the N-terminal domain of the multidrug resistance protein Mdr1 [8], the hydrophobic component of the ABC transporter KpsM [9] and aromatic aa permeases [10–12].

2.2. Display of randomized protein libraries

The parent domain (Fig. 1) can serve as a scaffold to display random peptide or protein libraries (reviewed in [13]). So far, such a protein scaffold has been used to display relatively short peptides, but we found that random sequence proteins of more than 100 aa residues could be also inserted into a surface loop region of an enzyme, *Escherichia coli* RNAse HI [14]. Conformationally constrained random sequence polypeptides on the scaffold are expected to be stabilized and protected against proteolytic degradation. Further characterization of the random sequence proteins displayed on the surface of RNase HI indicated that the scaffold enzyme activity was correlated with the degree of order in the structure of the random sequence polypeptide [15]. Thus, the parent domain can be used as a structural probe for the insert domain. As shown in Fig. 2B, by monitoring the scaffold enzymatic activity, the hill-climbing of random sequence proteins on a foldability landscape can be analyzed [13]. Not only RNAse HI, but also kanamycin nucleotidyltransferase and GFP can be used as scaffolds for displaying random sequence proteins (N. Doi, T. Yomo and H. Yanagawa, unpublished data).

Natural intact domains can also be displayed on the cell surface and on virus particles. For example, Lu et al. inserted a thioredoxin domain into a major structural component of the *E. coli* flagellin, FliC, to display the protein on the *E. coli* cell surface [16]. Though a filamentous phage display system has been used to display several enzymes (reviewed in [13]), these enzymes were fused to the N-terminus of the gene III capsid protein. An exception was quite recently reported:

*Corresponding author. Fax: (81) (42) 724-6317.
E-mail: hyana@libra.ls.m-kagaku.co.jp

Abbreviations: aa, amino acid; AP, alkaline phosphatase; BLIP, β-lactamase inhibitory protein; GFP, green fluorescent protein

PII: S0014-5793(99)00991-6

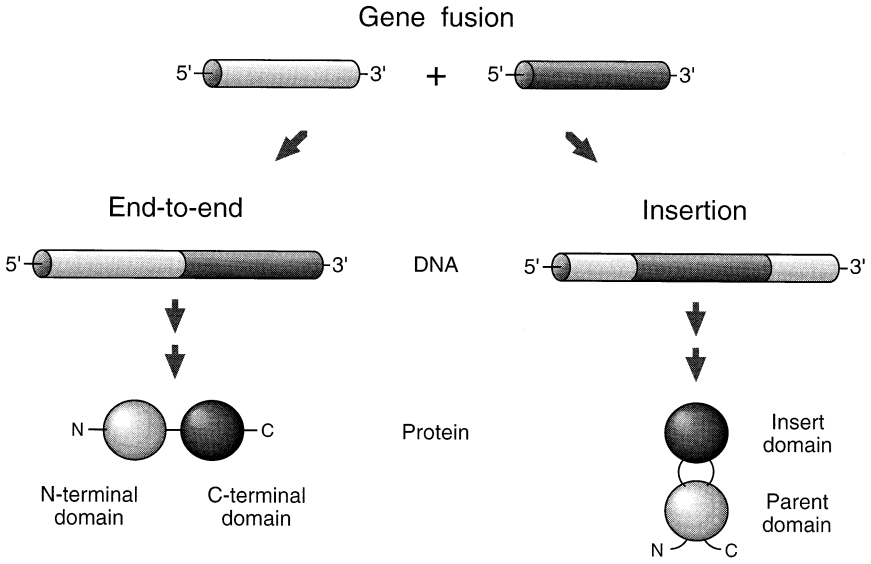

Fig. 1. Schematic representation of the two types of gene fusion approaches.

Kratz et al. inserted GFP into the middle of a core protein of hepatitis B virus capsid [17].

2.3. Design of biosensor proteins

Since our finding that a soluble domain accommodates insertions of large domain sequences with an unexpectedly high frequency [14], several fusion proteins have been produced by insertion of a globular domain into another soluble domain [18–21]. Betton et al. [18] inserted TEM1 β-lactamase into the maltose binding protein MalE. The purified fusion proteins retained both penicillinase and maltose binding activities. Interestingly, the penicillinase activity of the insert domain was modulated by a conformational change of MalE upon binding of maltose. Such phenomena have also been observed in experiments involving small peptide insertions [22–24]. These experiments suggest that the insertional fusion of a binding domain and a reporter domain can be used to develop a new generation of molecular biosensors.

One of the most useful reporter domains is the autofluorescent protein GFP from the jellyfish *Aequorea victoria* [4]. GFP has been fused to a large number of proteins by the end to end fusion approach [3]. A few examples of insertional fusion based on GFP were recently also reported.

First, GFP can be used as the insert domain (one example was noted in Section 2.2). Siegel and Isacoff [25] inserted GFP into a Shaker potassium channel from *Drosophila melanogaster*. The voltage-dependent rearrangements in the potassium channel induced changes in the fluorescence of GFP, i.e. this fusion protein acts as an optical sensor for membrane voltage. Strictly speaking, this fusion may be an end to end fusion, not an insertional fusion, because the insertion site of GFP is located between the last transmembrane domain and the large C-terminal extramembrane segment of the potassium channel. As an example of undoubted insertional fusion, Biondi et al. [20] performed random insertions of GFP into the cAMP-dependent protein kinase regulatory subunit to screen for functionally important regions in the protein. The goal of their work is to construct a cAMP sensor by using the two-component fluorescence resonance energy transfer with the catalytic subunit.

Alternatively, GFP can also be used as the parent domain. We have proposed a method of constructing generic GFP-based biosensors in which a desired molecular recognition domain is inserted into a loop of GFP [21]. We chose β-lactamase as a model protein to be inserted into GFP, attempting to construct an optical sensor for β-lactamase inhibitory protein (BLIP). By using random mutagenesis of the insertional fusion protein and screening for BLIP sensitivity, we obtained a sensor protein in which the GFP fluorescence increased upon binding of BLIP (Fig. 2C) [21].

Table 1
Engineered insertional fusion proteins

Insert domain (aa length)	Parent domain (aa length)	References
E. coli AP (450)	Various integral membrane proteins	[5,7–12]
E. coli thioredoxin (109)	*E. coli* flagellin, FliC (497)	[16]
Artificial random proteins (120)	*E. coli* RNAse HI (155)	[14]
A. victoria GFPΔC (232)	Shaker K^+ channel (616)	[25]
TEM1 β-lactamase (263)	*E. coli* MalE (369)	[18]
B. subtilis β-xylanase (185)	*B. macerans* β-glucanase mutant (212)	[19]
A. victoria GFP (238)	*Dictyostelium discoideum* protein kinase A R subunit (327)	[20]
A. victoria GFP (238)	Hepatitis B virus core protein (183)	[17]
TEM1 β-lactamase (263)	*A. victoria* GFP (238)	[21]

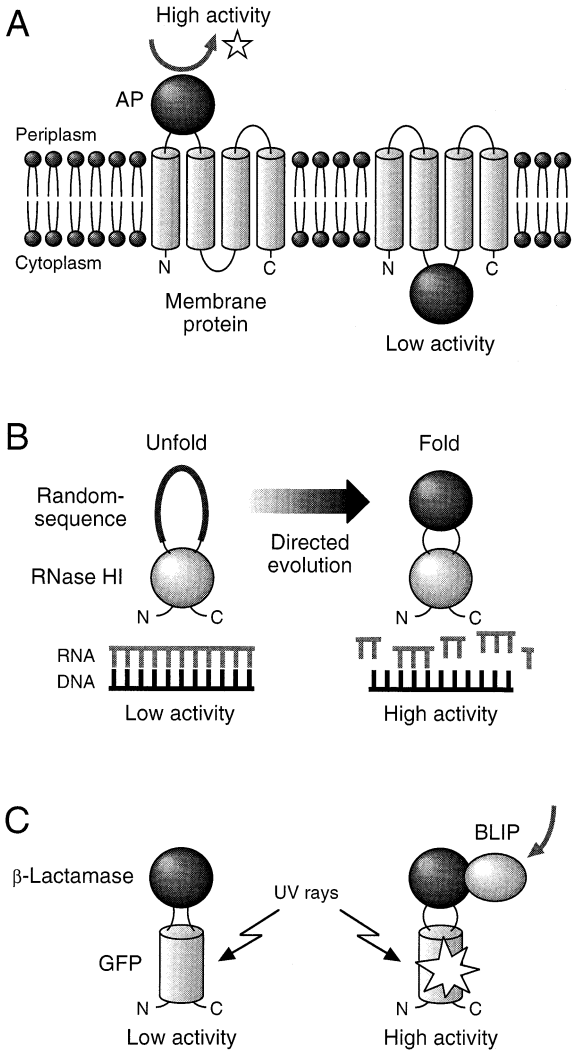

Fig. 2. Applications of domain insertion for functional switching.

3. Advantages and disadvantages of insertional fusion

The main difference between insertional fusion and end to end fusion is whether the two domains are connected with two linkers or one linker. Since a double connection allows fewer degrees of freedom than any single connection, insertional fusion proteins are expected to form more rigid and stable structures than end to end fusion proteins [26]. Indeed, Aÿ et al. [19] constructed a stable fusion protein, GluXyn-1, by insertion of the 1,4-β-xylanase from *Bacillus subtilis* into a circularly permuted mutant of the 1,3-1,4-β-glucanase from *Bacillus macerans*. GluXyn-1 is a very stable bifunctional protein and is inaccessible to proteolytic attack in vivo. This property is advantageous for the design of bifunctional proteins with a high stability, but is unfavorable when one domain needs to be detached from the other, e.g. after purification of the fusion proteins.

Gene construction for an end to end fusion protein is simple and easy. Points to be considered are, at most, whether one domain should be linked to the N- or C-terminus of the other and the length and sequence of the linker peptide. Most end to end fusion proteins maintain the intact function of both proteins. On the other hand, construction of an insertional fusion protein is rather complicated, because it requires precise information on the parent domain structure to identify a suitable insertion site. Inappropriate design often leads to destabilization and inactivation of the insertional fusion protein. However, as described in Section 2, this sensitivity can be utilized as a structural probe for membrane proteins and random polypeptides and can be applied to the design of molecular sensors or switches.

The advantages and disadvantages of insertional fusion compared to end to end fusion are summarized in Table 2. The requirement of N- and C-terminal proximity of the insert domain may not be a serious demerit, because terminal proximity is observed in the large majority of globular proteins (see [27] and references cited therein). In conclusion, the major advantages of insertional fusion lie in the ambivalent potential for a high stability and sensitivity of the product protein. The insertional fusion technique should thus be combined with screening methods [13,21] to select stable bifunctional proteins or sensor proteins as required.

Table 2
Advantages and disadvantages of the insertional fusion protein

Advantages
Have a rigid structure with high stability
Less susceptible to proteolytic degradation in vivo
Can be used as probes for proteins whose structures are little understood
Applicable to the design of biosensors
Disadvantages
Need proximity of the terminals of the insert domain
Must find 'permissive sites' of the parent domain
Two linkers are not easily cleaved

References

[1] Slauch, J.M. and Silhavy, T.J. (1991) Methods Enzymol. 204, 213–248.
[2] Ford, C.F., Suominen, I. and Glatz, C.E. (1991) Protein Expr. Purif. 2, 95–107.
[3] Gerdes, H.-H. and Kaether, C. (1996) FEBS Lett. 389, 44–47.
[4] Tsien, R.Y. (1998) Annu. Rev. Biochem. 67, 509–544.
[5] Ehrmann, M., Boyd, D. and Beckwith, J. (1990) Proc. Natl. Acad. Sci. USA 87, 7574–7578.
[6] Hennessey, E.S. and Broome-Smith, J.K. (1993) Curr. Opin. Struct. Biol. 3, 524–531.
[7] Lacatena, R.M., Cellini, A., Scavizzi, F. and Tocchini-Valentini, G.P. (1994) Proc. Natl. Acad. Sci. USA 91, 10521–10525.
[8] Bibi, E. and Béjà, O. (1994) J. Biol. Chem. 269, 19910–19915.
[9] Pigeon, R.P. and Silver, R.P. (1994) Mol. Microbiol. 14, 871–881.
[10] Sarsero, J.P. and Pittard, A.J. (1995) J. Bacteriol. 177, 297–306.
[11] Pi, J. and Pittard, A.J. (1996) J. Bacteriol. 178, 2650–2655.
[12] Cosgriff, A.J. and Pittard, A.J. (1997) J. Bacteriol. 179, 3317–3323.
[13] Doi, N. and Yanagawa, H. (1998) Cell. Mol. Life Sci. 54, 394–404.
[14] Doi, N., Itaya, M., Yomo, T., Tokura, S. and Yanagawa, H. (1997) FEBS Lett. 402, 177–180.
[15] Doi, N., Yomo, T., Itaya, M. and Yanagawa, H. (1998) FEBS Lett. 427, 51–54.
[16] Lu, Z., Murray, K.S., Cleave, V.V., LaVallie, E.R., Stahl, M.L. and McCoy, J.M. (1995) BioTechnology 13, 366–372.
[17] Kratz, P.A., Böttcher, B. and Nassal, M. (1999) Proc. Natl. Acad. Sci. USA 96, 1915–1920.

[18] Betton, J.-M., Jacob, J.P., Hofnung, M. and Broome-Smith, J.K. (1997) Nat. Biotechnol. 15, 1276–1279.
[19] Aÿ, J., Götz, F., Borriss, R. and Heinemann, U. (1998) Proc. Natl. Acad. Sci. USA 95, 6613–6618.
[20] Biondi, R.M., Baehler, P.J., Reymond, C.D. and Véron, M. (1998) Nucleic Acids Res. 26, 4946–4952.
[21] Doi, N. and Yanagawa, H. (1999) FEBS Lett. (in press).
[22] Brennan, C.A., Christianson, K., La Fleur, M.A. and Mandecki, W. (1995) Proc. Natl. Acad. Sci. USA 92, 5783–5787.
[23] Benito, A., Feliu, J.X. and Villaverde, A. (1996) J. Biol. Chem. 271, 21251–21256.
[24] Feliu, J.X. and Villaverde, A. (1998) FEBS Lett. 434, 23–27.
[25] Siegel, M.S. and Isacoff, E.Y. (1997) Neuron 19, 735–741.
[26] Russell, R.B. (1994) Protein Eng. 7, 1407–1410.
[27] Doi, N. and Yanagawa, H. (1998) FEBS Lett. 430, 150–153.