A cloning strategy to obtain recombinant proteins with identical primary structure to the native forms

Margarida Santana, Maria Carmen Portillo and Juan Gonzalez*

Instituto de Recursos Naturales y Agrobiologia de Sevilla, IRNAS-CSIC

Avda. Reina Mercedes 10, 41012 Seville, Spain

* corresponding author

Email: jmgrau@irnase.csic.es

Telephone: +34 954624711

Fax number: +34 954624002

Key words: cloning strategy, expression vectors, native protein, recombinant protein, fusion tag, biocatalysts

Abstract

Recombinant gene strategies using fusion tags for purification are essential procedures to obtain large protein quantities. However, most cloning systems result in recombinant proteins with added amino acids inexistent in their native forms which can lead to significant changes in protein properties. An original and simple cloning strategy is proposed to obtain proteins identical in amino acid sequence to the native proteins.

At present, there is a wide range of cloning vectors available for the overexpression of recombinant proteins. Medicine, Pharmacy, food industry and most biotechnological industries are common users of recombinant gene technologies. Cloning and expression systems (CES) require the insertion of the gene sequence of interest fitting in the right reading frame (Sambrook et al., 1989). However, during this design, most CES incorporate several additional nucleotides, from 3 to over 30 bases, resulting in the addition of 1 to over 10 amino acids, respectively. The addition of amino acid residues to a protein of interest can lead to significant changes in their properties affecting, for instance, to their stability, antigenic characteristics, and reaction kinetics. Besides, most of the current commercial CES generate fusion recombinant proteins with a tag used for simplifying the purification of the recombinant protein. After digestion with specific proteases (factor Xa, enterokinase among others depending on each CES) the purified protein is liberated of that tag. Generally, the result is a recombinant protein carrying additional amino acids absent in its native form. Herein, a novel cloning strategy is proposed which accounts for the use of fusion tags and their removal after protein purification, resulting in recombinant proteins with a primary structure identical to the native proteins.

An example of the proposed strategy is described using the cloning, expression and purification system LYTAG Spin Plus Kit (Biomedal, S.L., Seville, Spain) with the vector pAlex2a as the cloning vector. The plasmid pAlex2a contains a 136 amino acid tag used for rapid protein purification which is removed by enterokinase digestion. The proposed procedure consists in introducing a cassette with two recognition sites for class IIS restriction enzymes each one oriented in opposite directions. Class IIS restriction enzymes recognize asymmetric sequences and cleave these sequences at a defined distance outside their recognition sites (Szybalski et al., 1991). Because the recognition sites for these restriction enzymes are asymmetric, they can be oriented in each of the two strains of the DNA so that the cleavage site is engineered as desired.

The insertion of two recognition sites for class IIS restriction enzymes oriented in opposite directions is a key aspect of the proposed methodology and it aims to external digestions off the cassette sequence which allows the removal of some additional nucleotides required during the insertion of the desired gene into the expression vector.

Figure 1 shows a strategy suggested to modify pAlex2a vector as a representative example to describe the proposed cloning method. First, a short sequence flanked by two *Btg*ZI recognition sites must be removed and the plasmid recircularized. This fragment is located downstream of the multiple cloning site (not shown in Figure 1) and does not affect the cloning and expression of recombinant proteins using this vector. Second, the resulting plasmid is double digested with *Psh*AI and *Eco*RV (which destroys the enterokinase recognition site) and religated. The plasmid is double digested with *Bam*HI and *Asc*I and the proposed cassette (with overhangs corresponding to digestions by *Bam*HI and *Asc*I) is ligated into the plasmid resulting in a construct ready to receive the gene of interest. The designed cassette, besides two *Btg*ZI sites, contains an additional restriction enzyme recognition site (*Nde*I) to easily evaluate if the right cloning vector is obtained. The critical points about the design of this cassette are the sites where the class IIS restriction enzyme cuts (beyond the 5’ and 3’ sides of the cassette). In this example, after the type IIS restriction digestion, the enterokinase cleavage site is reestablished and placed just before the start codon of the cloned gene. A cassette planned for this CES is described in Figure 1.

The gene of interest can be PCR amplified using a thermostable DNA polymerase with proofreading activity and two specially designed primers of general use (Figure 1). A forward primer containing the proposed sequence (5’-C TAG GCG CGC CAC AAG **ATG** + target gene sequence) followed by a target sequence corresponding to the beginning of the gene to be amplified and cloned; its start codon (ATG; bold) is already added to the primer

sequence. For those proteins lacking the ATG start codon (Met), this triplet can be removed. A restriction site for *Asc*I (underlined) is included for its insertion in the constructed vector. The reverse primer should contain the proposed sequence followed by the target sequence corresponding to the ending of the gene of interest (target gene sequence + **TGA** AGC TTA GT -3'). This reverse primer includes a *Hind*III restriction site (underlined) and a stop triplet (TGA; bold). This design allows cloning between the *Asc*I and *Hind*III sites of the vector constructed as described above. A double digestion of the constructed vector and the amplified gene is followed by a ligation leading to the insertion of the gene.

Next, the protocol involves a digestion with the type IIS restriction enzyme (*Btg*ZI in this example) which cuts the construct at two places, both beyond the inserted fragment (Figure 1b). After this digestion and its ligation, the enterokinase cleavage site is reestablished. Clone overexpression is followed by an easy purification using the LYTAG system included in this CES. After purification according to the manufacturer instructions, enterokinase digestion will remove the purification tag and the resulting primary structure of the recombinant protein will be identical to the native protein of interest (Figure 1b). This procedure has been successfully applied to produce a recombinant protein corresponding to a mutant *gnt*R gene from *Desulfovibrio vulgaris* Hildeborough (Heidelberg et al., 2004). Its overexpression was not affected by the proposed procedure (Supplementary data Figure S1).

Once the cloning vector contains the necessary cassette, the cloning is straightforward and represents a simple protocol which can be easily applied to the cloning and overexpression of genes from absolutely any source. Different class IIS restriction enzymes can be used to design appropriate cassettes for different CES as far as the right cutting points are calculated to achieve a lack of additional nucleotides and the right reading frame for the inserted gene sequence. A few

other examples (among many others) of class IIS restriction enzymes are suggested in Table 1 which can be used to design cloning procedures following the guidelines outlined in this study. These enzymes can be applied to other vectors and genes and to avoid those enzymes with recognition sites present in the desired gene sequence. Thus, a variety of CES and vectors can be used with the proposed procedure. We have designed protocols and cassettes for a variety of high-level expression vectors such as pAlex2a (the example provided in Figure 1)(Biomedal), pGEX family vectors using GST gene fusion systems (GE Healthcare, Piscataway, NJ, USA) (Example in Supplementary data Figure S2), and pIEx vector systems with S-tag and His-tag coding sequences (EMD Biosciences, Darmstadt, Germany). Enterokinase (example of Figure 1) and other proteases (e.g., Factor Xa in the example of supplementary data Figure S2) used to remove the tags of the fusion tag recombinant proteins might encounter difficulty to reach their recognition sites (Shahravan et al., 2008). In these cases protease digestion sites can get buried inside the recombinant protein structure. This inconvenient can be approached by partial denaturation (i.e., 1M urea) to produce a more open structure of the fusion tag recombinant protein (Shahravan et al. 2008) during protease digestion.

The outlined procedure results in a simple way to design amplification primers for cloning and recombinant gene expression leading to the production of recombinant proteins with an identical amino acid sequence to the native molecule. The procedure is highly useful for a broad range of applications in different fields including medicine, clinical analysis, pharmacy, food industry, and biotechnology. In this study, a method for the cloning and expression of recombinant proteins is described. The proposed strategy, besides utilizing the fusion tags for an easy and fast purification of the recombinant proteins, warrants final recombinant proteins containing identical primary structures than their native forms.

Acknowledgements. The authors acknowledge funding from the Government of Andalusia (Spain). M. Santana was recipient of grant SFRH/BPD/34404/2006 from the Fundação para a Ciência e a Tecnologia (Portugal).

References:

Heidelberg, J.F., Seshari, R., Haveman, S.A., Hemme, C.L., Paulsen, J.T., Kolonay, J.F., Eisen, J.A., Ward, N., Methe, B., Brinkac, L.M., Daugherty, S.C., Deboy, R.T., Dodson, R.J., Durkin, A.S., Madupu, R., Nelson, W.C., Sullivan, S.A., Fouts, D., Haft, D.H., Selengut, J., Peterson, J.D., Davidsen, T.M., Zafar, N., Zhou, L., Radune, D., Dimitrov, G., Hance, M., Tran, K., Khouri, H., Gill, J., Utterback, T.R., Feldblyum, T.V., Wall, J.D., Voordouw, G., Fraiser, C.M., 2004. The genome sequence of the anaerobic, sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough. Nature Biotechnol. 22, 554–559.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor laboratory, Cold Spring Harbor, N.Y.

Shahravan, S.H., Qu, X., Chan, I-S., Shin, J.A., 2008. Enhancing the specificity of the enterokinase cleavage reaction to promote efficient cleavage of a fusion tag. Protein Expr. Purif. 59, 314-319.

Szybalski, W., Kun, S.C., Hasan, N., Podhajska, A.J., 1991. Class-IIS restriction enzymes – a review. Gene 100, 13-26.

Figure legend

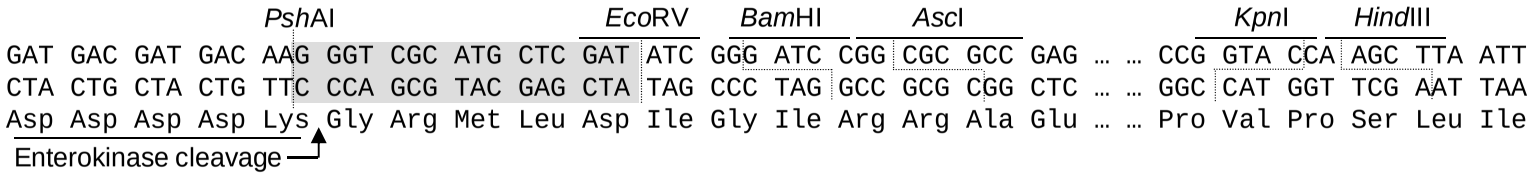
Figure 1. Schematic representation of the proposed cloning procedure applied to the plasmid pAlex2a using *Btg*ZI restriction sites in the designed cassette. In A the steps required to construct a vector containing the cloning cassette are shown. B presents the procedure to insert a gene of interest in that vector and produce the desired protein lacking any additional amino acid residue. At the final step, the product obtained after enterokinase digestion is the recombinant protein with primary structure identical to its native protein.

166

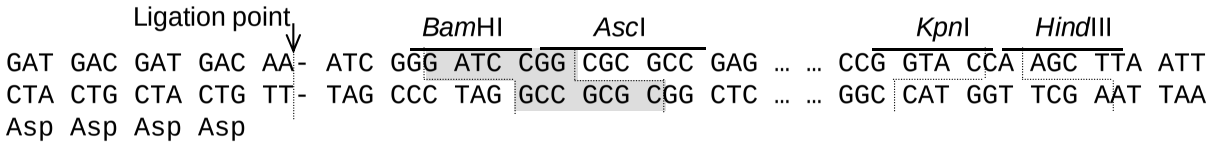
167 **Table 1**. A few examples of commercial class IIS restriction enzymes proposed as
168 candidates for their use in the design of a cloning procedure such as the one
169 outlined in this study. These restriction enzymes can be selected as needed
170 depending on the gene sequence to be cloned and the cloning vector to be used.
171

Class IIS restriction enzyme	**Recognition and cleavage sites**
AlwI	5′-GGATC(N)$_4$ CCTAG(N)$_5$
Alw26I	5′-GTCTC(N)$_1$ CAGAG(N)$_5$
BbsI	5′-GAAGAC(N)$_2$ CTTCTG(N)$_6$
BsgI	5′-GTGCAG(N)$_{16}$ CACGTC(N)$_{14}$
BtgZI	5′-GCGATG(N)$_{10}$ CGCTAC(N)$_{14}$
Eco57I	5′-CTGAAG(N)$_{16}$ GACTTC(N)$_{14}$
FokI	5′-GGATG(N)$_9$ CCTAC(N)$_{13}$
KspGUII	5′-CTCTTC(N)$_1$ GAGAAG(N)$_4$
Tth111II	5′-CAARCA(N)$_{11}$ GTTYGT(N)$_9$

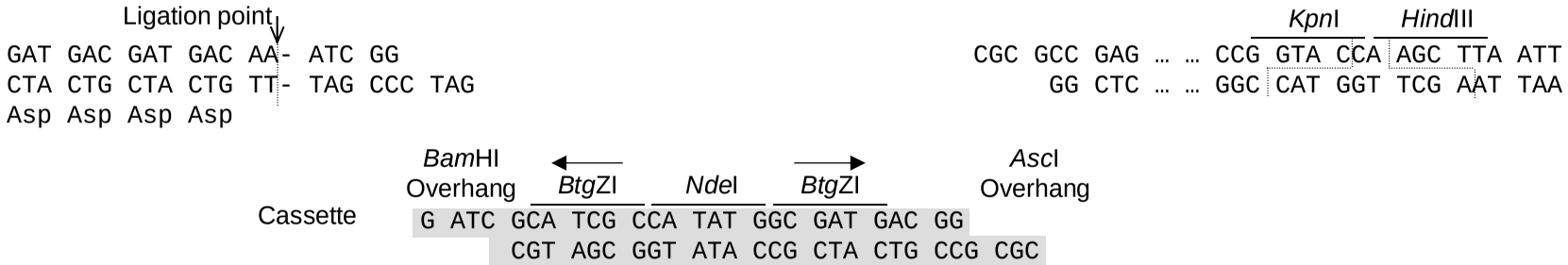
Figure 1 A

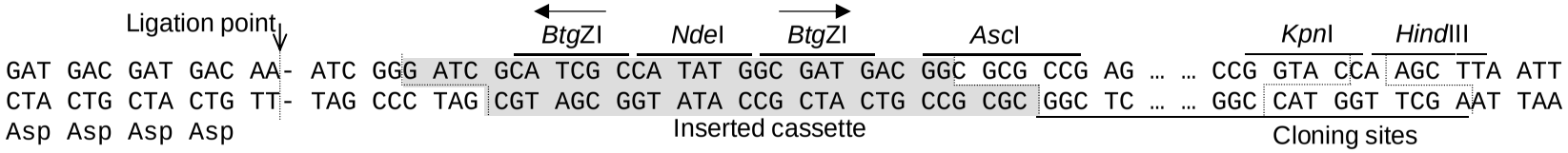


Digestion of the vector with *Psh*AI and *Eco*RV and religation



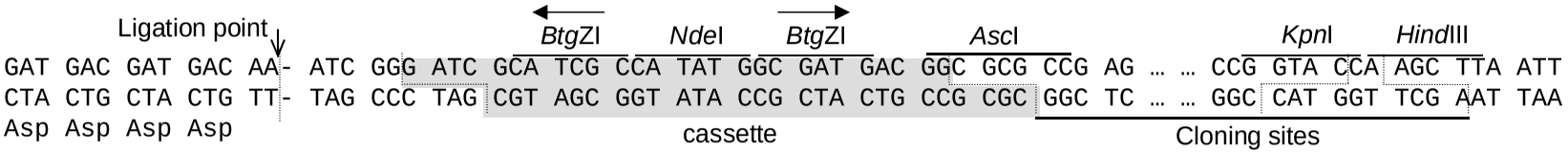
Digestion of the new construct with *Bam*HI and *Asc*I and insertion of the cassette.





The vector is ready to receive the gene of interest.

Figure 1 B



Amplification of the gene of interest using the forward primer: 5′ – C TAG GCG CGC CAC AAG **ATG** target gene sequence

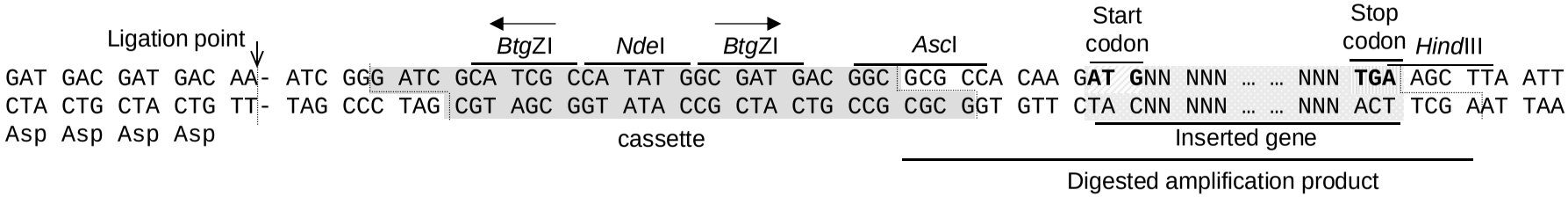
and the reverse primer: target gene sequence **TGA** AGC TTA GT – 3′

Digestion of the amplified product with *Asc*I and *Hind*III.

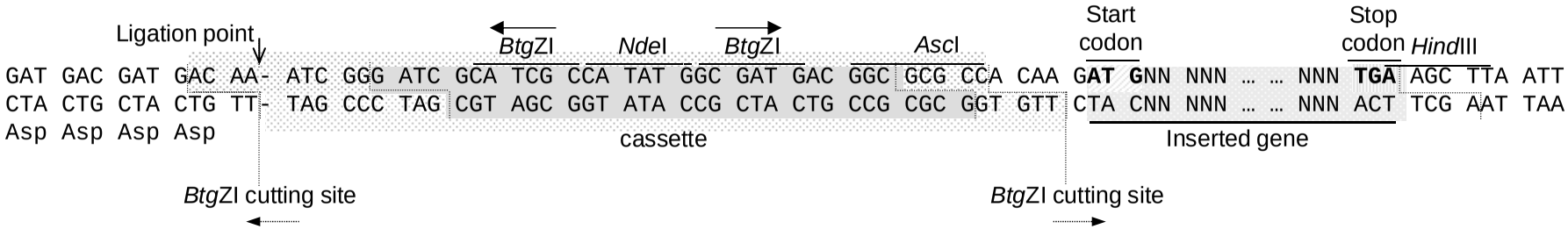
*Asc*I Start codon

Stop codon *Hind*III

Digestion of the constructed vector with *Asc*I and *Hind*III. Ligation of digested vector and amplification product



Digestion of the resulting construct with the type IIS restriction enzyme *Btg*ZI



Purification of recombinant protein using LYTAG system and digestion with enterokinase