Available online at www.sciencedirect.com
ScienceDirect
Journal of Biotechnology 131 (2007) 223–230
Journal of BIOTECHNOLOGY
www.elsevier.com/locate/jbiotec

Protein expression from synthetic genes: Selection of clones using GFP

Sandrine Denis-Quanquin, Lucille Lamouroux, Andrée Lougarre, Sabrina Mahéo, Isabelle Saves, Laurent Paquereau, Pascal Demange, Didier Fournier*

IPBS, 205 route de Narbonne, 31000 Toulouse, France

Received 22 March 2007; received in revised form 25 June 2007; accepted 13 July 2007

Abstract

Construction of synthetic genes is today the most elegant way to optimize the heterologous expression of a recombinant protein. However, the selection of positive clones that incorporate the correct synthetic DNA fragments is a bottleneck as current methods of gene synthesis introduce 3.5 nucleotide deletions per kb. Furthermore, even when all predictable optimizations for protein production have been introduced into the synthetic gene, production of the protein is often disappointing: protein is produced in too low amounts or end up in inclusion bodies. We propose a strategy to overcome these two problems simultaneously by cloning the synthetic gene upstream of a reporter gene. This permits the selection of clones devoid of frame-shift mutations. In addition, beside nucleotide deletion, an average of three non-neutral mutations per kb are introduced during gene synthesis. Using a reporter protein downstream of the synthetic gene, allows the selection of clones with random mutations improving the expression or the folding of the protein of interest. The problem of errors found in synthetic genes is then turned into an advantage since it provides polymorphism useful for molecular evolution. The use of synthetic genes appears as an alternative to the error-prone PCR strategy to generate the variations necessary in protein engineering experiments.

Keywords: Expression; GFP; Mutation; Synthetic gene

1. Introduction

Synthetic genes made by oligonucleotide assembly have been used since the late 1970s (Gupta et al., 1968; Crea et al., 1978; Goeddel et al., 1979) and today several methods of synthesis are currently used (Sandhu et al., 1992; Stemmer et al., 1995; Young and Dong, 2004). These techniques draw considerable attention in protein engineering when the DNA sequence encoding the protein of interest is not available or is difficult to clone (Scheibel, 2004; Chang et al., 2002) when the species utilizes a non-universal genetic code (Chang et al., 2006) or when the sequence needs to be optimized for a suitable host for heterologous expression. Optimization may be obtained by choosing an appropriate codon table (Feng et al., 2000; Milek et al., 2000; Wheeler et al., 1996; Hale and Thompson, 1998; Baedeker and Schulz, 1999), by improving the initiation of translation (Milek et al., 2000) and by elimination of polyadenylation signals (Withers-Martinez et al., 1999; Gurkan and Ellar, 2003b), false priming events (Wheeler et al., 1996), cryptic promoters (Bertorello et al., 1990), strong secondary RNA structures (Wheeler et al., 1996), sequences that contribute to RNA instability (Kang et al., 2004) or signals of post-translational modifications (Milek et al., 2000). The use of a synthetic gene appears to be the most valid solution because the corrections that would be necessary to optimize the sequence are too numerous to be considered by site-directed mutagenesis. In addition, gene synthesis allows the introduction of many unique restriction sites along the sequence in order to facilitate subsequent cloning steps (Del Vecchio Blanco et al., 1998; Wheeler et al., 1996; Baedeker and Schulz, 1999). However, the assembly of synthetic oligonucleotides generally results in frequent DNA sequence errors (Smith et al., 2003; Hoover and Lubkowski, 2002; Xiong et al., 2004; Delroisse et al., 2005). Indeed, the majority of the sequence errors come from the inac-

* Corresponding author. Tel.: +33 5 61 17 54 45; fax: +33 5 61 17 59 94.
E-mail address: fournier@ipbs.fr (D. Fournier).

0168-1656/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbiotec.2007.07.725

curacy of the oligonucleotide sequences used as starting material for the construction of the synthetic gene: despite the excellent yield of the coupling reaction between two nucleotides (yield ~99.99%), the length of the sequence of the entire synthetic gene will necessarily lead to by-products containing internal deletions or point mutations. Moreover, the subsequent PCR amplification of the full-length sequence may introduce additional mutations into the final DNA fragment. Consequently, the assembly of synthetic oligonucleotides generally necessitates either screening of a large number of clones to find the correct desired sequence or subsequent correction of the cloned sequences (Subramanian et al., 2002; Gurkan and Ellar, 2003a,b; Cherepanov et al., 2000; Barber et al., 2002; Lin et al., 2002; Carr et al., 2004; Hale and Thompson, 1998; Baedeker and Schulz, 1999; Binkowski et al., 2005; Smith et al., 2003; Delroisse et al., 2005; Popovic et al., 2006; Hu et al., 2006; Wu et al., 2006).

The current way of limiting the risk of introducing deletions in the synthetic gene is to previously purify the oligonucleotides by electrophoresis (Lehmann et al., 2000). This purification step selects molecules of the correct size and hence devoid of deletions; its efficiency decreases with the oligonucleotide's length leading to the use of oligonucleotides shorter than 60 nucleotides (Xiong et al., 2004). Another way is to build small fragments of 500–800 bp, to select the accurate ones and finally to join them to obtain longer sequences (Kodumal et al., 2004; Xiong et al., 2004). A final but time consuming possibility consists in the removal of error-containing sequences by selective digestion using mismatch repair proteins or by binding on MutS (Carr et al., 2004; Binkowski et al., 2005; Smith and Modrich, 1997). Even with all possible optimizations, proteins are often not efficiently produced or folded. Unpredictable mutations may be then selected by combinatorial directed evolution techniques such as random mutagenesis (Cadwell and Joyce, 1992) or DNA recombination (Stemmer, 1994).

In this study, we propose a strategy involving the expression of a reporter protein fused to the C-terminal extremity of the protein of interest to simultaneously select synthetic genes devoid of frame-shift mutations and with mutations providing satisfactory expression and folding of the protein. We used the green fluorescent protein (GFP) as a reporter to select clones expressing the complete synthetic gene with useful mutations for expression.

2. Materials and methods

2.1. Materials

Commercially available oligonucleotides were obtained from Sigma Genosys or Eurobio. They were deprotected and desalted but not size-purified. All oligonucleotides constituting the genes of interest were designed for optimal translation according to codon optimization using DNAsmac software. All the reagents used were biology molecular grade. Vectors pET17b and pGEX2T were from Novagen and GE Healthcare, respectively.

Fig. 1. Annotated DNA sequence of the promoter, polylinker and GFP fusion region of the four vectors.

2.2. Vector construction

The vector pTG (Fig. 1) was constructed in such a way that the synthetic gene has to carry the sequences required for the initiation of translation allowing expression of the GFP fusion. The GFP encoding gene was obtained by PCR amplification from pEGFP-N1 (Clontech) and inserted into the pBluescript SK$^-$ vector downstream of the LacUV5 promoter. This promoter was exchanged for the tac promoter to increase the reporter gene transcription (de Boer et al., 1983) by insertion of a 46 bp long double strand oligonucleotide between the SapI and EcoRI restriction sites. Colonies expressing the higher amount of GFP were selected, i.e. the greenest fluorescent colonies under illumination at 388 nm. Surprisingly, it appeared that the expression of GFP was very high, GFP accounted for more than 50% of total proteins and expression was not stable after successive subcultures. We then selected among these colonies, the three expressing the highest amounts of GFP after making five replicas. They all harbored mutated tac promoters (Fig. 2) illustrating errors originating from oligonucleotide synthesis, which in this case was useful to obtain optimized transcrip-