Protein Expression and Purification 128 (2016) 93–100

Contents lists available at ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep

Protein expression of preferred human codon-optimized *Gaussia* luciferase genes with an artificial open-reading frame in mammalian and bacterial cells

Satoshi Inouye [a, *], Takahiro Suzuki [b]

[a] Yokohama Research Center, JNC Co., 5-1 Okawa, Kanazawa-ku, Yokohama 236-8605, Japan
[b] Department of Biochemistry, School of Dentistry, Aichi-Gakuin University, 1-100 Kusumoto-cho, Chikusa-ku, Nagoya 464-8650, Japan

ARTICLE INFO

Article history:
Received 6 July 2016
Received in revised form
1 August 2016
Accepted 3 August 2016
Available online 6 August 2016
This paper is dedicated to the late Dr. Frederick I. Tsuji (deceased on May 7, 2016).

Keywords:
Codon optimization
Synthetic gene
Luciferase
Photoprotein
Reporter gene

ABSTRACT

The protein expressions of three preferred human codon-optimized *Gaussia* luciferase genes (*pGLuc*, *EpGLuc*, and *KpGLuc*) were characterized in mammalian and bacterial cells by comparing them with those of wild-type *Gaussia* luciferase gene (*wGLuc*) and human codon-optimized *Gaussia* luciferase gene (*hGLuc*). Two synthetic genes of *EpGLuc* and *KpGLuc* containing the complete preferred human codons have an artificial open-reading frame; however, they had the similar protein expression levels to those of *pGLuc* and *hGLuc* in mammalian cells. In bacterial cells, the protein expressions of *pGLuc*, *EpGLuc*, and *KpGLuc* with approximately 65% GC content were the same and showed approximately 60% activities of *wGLuc* and *hGLuc*. The artificial open-reading frame in *EpGLuc* and *KpGLuc* did not affect the protein expression in mammalian and bacterial cells.

1. Introduction

Gaussia luciferase (GLase) is a secretory luciferase and consists of 185 amino acids including 17 amino acid residues of the signal peptide sequence for secretion [1–4]. GLase does not require any cofactors and uses only coelenterazine (a luciferin) and molecular oxygen for the luminescence reaction. Further, GLase is a thermostable luciferase with the highest specific activity among coelenterazine-utilizing luciferases including *Renilla* luciferase and the mutated 19 kDa protein of *Oplophorus* luciferase (nanoKAZ = nanoLuc) [4–6]. The small GLase with a rapid decay luminescence pattern is suitable as a reporter protein for video-rate bioluminescence imaging to detect the secretion of protein from living mammalian cells [7–10].

To express a heterologous protein in bacterial cells, the method of the codon optimization for a target protein was proposed by using computational tools [11,12]. Similarly, codon optimization was performed in mammalian cells and the improvement of protein expression was shown by immunoblot analyses [13].

Recently, we proposed a simple design method to prepare codon-optimized genes for expressing a heterologous protein efficiently in mammalian cells and named our method as the "preferred human codon-optimized method" [14]. By selecting only preferentially used codons from the human codon table (http://www.kazusa.or.jp/codon/) to synthesize the preferred human codon-optimized gene of several luciferases and photoproteins, we confirmed the efficient protein expression in mammalian cells [14]. When preferential codons in the human codon table were selected for gene synthesis, the genes obtained had the following characteristics: (i) High GC content with over 60% was observed in the nucleotide sequences of genes. (ii) The third position of codons corresponding to amino acids was for guanine (G) or cytosine (C), but not for adenine (A) and thymine (T), except for an arginine codon of AGA. Unexpectedly, stop codons (TAA, TAG, and TGA) were not present in the third frame of the nucleotide sequence, and the third frame showed an artificial unknown protein. (iii) The cryptic

Abbreviations: w, wild-type codon; h, human codon-optimized; p, preferred human codon-optimized; GLase, *Gaussia princeps* luciferase; *GLuc*, *Gaussia princeps* luciferase gene; GLsp, signal peptide sequence of *Gaussia* luciferase for secretion; CMV, cytomegalovirus; I_{max}, maximum intensity of luminescence; rlu, relative light units.

* Corresponding author.
E-mail address: sinouye@jnc-corp.co.jp (S. Inouye).

http://dx.doi.org/10.1016/j.pep.2016.08.006
1046-5928/

splice site of mRNA has a consensus sequence of GT/AG, but the combination frequencies of "GT" and "AG" may be very low. (iv) The restriction enzyme sites for recognizing the AT-rich sequence such as "GAATTC" for *Eco*RI were not present, and "AAGCTT" for *Hin*dIII will only be present in the amino acid sequence of Arg-Ser-Phe (AGA-AGC-TTC). On the other hand, the restriction enzyme sites for GC-rich sequences such as *Pst*I (CTGCAG) and *Sal*I (GTCGAC) were sometimes present. (v) There was no possibility of having AT-rich sequences such as the binding-site of the transcription factor for TATAAA or TATATAT, and the polyadenylation signal sequence of AATAA.

Previously, we synthesized a preferred human codon-optimized gene of *Gaussia* luciferase gene (*pGLuc*) [14], which has 64.1% GC content, and compared the expression level of GLase in mammalian cells with that of wild-type gene (*wGLuc*) [1] and human codon-optimized gene (*hGLuc*) [2]. The GC contents of *wGLuc* and *hGLuc* were 43.6% and 58.9%, respectively (Fig. 1 and Table 1). As a result, the luminescence activities of *pGLuc* and *hGLuc* expressed in Chinese hamster ovary-K1 (CHO-K1) cells were over 10-fold higher than those of *wGLuc*. Interestingly, the GLase expression level with *pGLuc* in HEK293, HepG2, and Gin-1 cells showed 1.5–2-fold higher activity than that with *hGLuc* [14]. The reasons for the differences in the luminescence activity between *pGLuc* and *hGLuc* in mammalian cells are still unclear.

In the *pGLuc* sequence [4], three amino acid codons of GCT for alanine, GGT for glycine, and AGT for serine were used to introduce the stop codons (TGA, TAG, and TGA, respectively) in the third frame (Fig. 2). Further, GTC for valine was chosen as a codon, but GTG for valine was not use as a preferred codon in the human codon usage table (Fig. 2). To understand the effects of these differences on the protein expression of GLase in mammalian cells, we prepared the complete preferred human codon-optimized genes of *EpGLuc* and *KpGLuc* and the expressions of these GLase genes in mammalian cells were compared with those of *pGLuc, hGLuc,* and *wGLuc*. In addition, we characterized the protein expression of these genes in *Escherichia coli* cells.

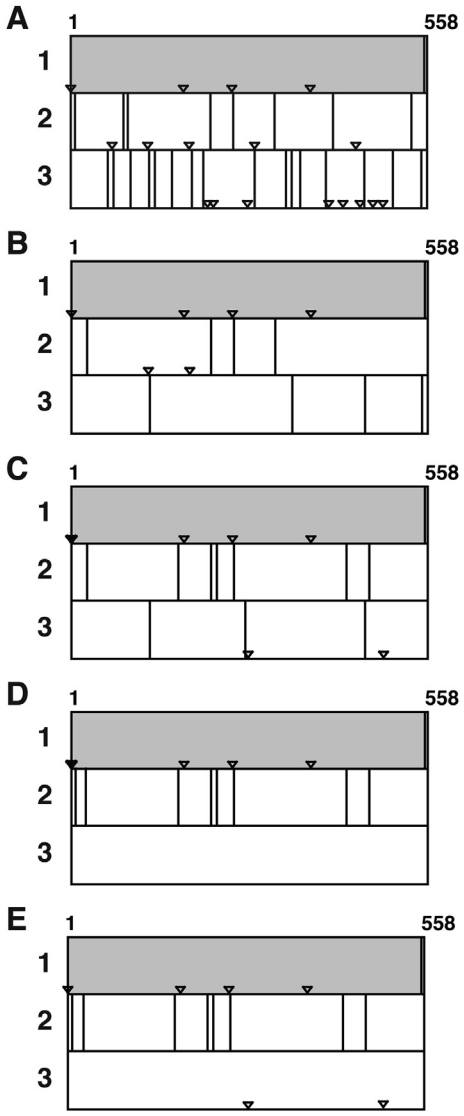

Fig. 1. Open-reading frames in various synthetic GLase genes. The numbers on the left margin indicate the reading frame 1, 2, and 3, respectively. The shaded boxes show the open-reading frames for GLase (frame 1). Arrowheads and vertical lines show the positions of methionine codon and stop codons, respectively. A, *wGLuc*; B, *hGLuc*; C, *pGLuc*; D, *EpGLuc*; E, *KpGLuc*.

2. Materials and methods

2.1. Gene design

Gene design for the preferred human codon gene derived from the primary sequence of protein was performed using the "Reverse Translate" software from Sequence Manipulation Suite (http://www.bioinformatics.org/sms2/rev_trans.html) with the human codon usage table from Kazusa or. jp. (http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=9606&aa=1&style=GCG). Two complete preferred human codon genes for GLase were designed and named *EpGLuc* and *KpGLuc*. In *EpGLuc*, the codon usage of AGG for arginine was chosen and AGA for arginine was used in *KpGLuc* (Table 1 and Fig. 2).

2.2. Construction of expression vectors for mammalian cells and bacterial cells

i) Expression vectors in mammalian cells: The expression vectors for wild-type *GLuc* (*wGLuc*), human codon-optimized *GLuc* (*hGLuc*), and preferred human codon-optimized *GLuc* (*pGLuc*) were used as pcDNA3-wGLuc, pcDNA3-hGLuc and pcDNA3-pGLuc, respectively, as described in our previous report [14]. The genes of *EpGLuc and KpGLuc* were obtained from Eurofins Genomics K.K. (Tokyo, Japan). To construct the expression vectors in mammalian cells, the *Hin*dIII-*Xba*I fragments of *EpGLuc and KpGLuc* were inserted into the *Hin*dIII-*Xba*I sites of a pcDNA3 vector (Invitrogen, Carlsbad, CA) to give pcDNA3-EpGLuc and pcDNA3-KpGLuc, respectively. These vectors had an identical sequence between the promoters and the initial methionine codon in the same vector (Fig. 3).

ii) Expression vectors in bacterial cells: The cold-inducible expression vector of pCold II [15] or pCold-ZZ-P [16,17] was used to produce GLase in *E. coli* cells. In a pCold-ZZ-P vector, GLase was expressed as a fused protein to the ZZ domain. The ZZ domain is the synthetic IgG-binding domain of protein A and serves as a soluble partner of GLase [16]. The *Eco*RI-*Xba*I fragments of GLase genes (*wGLuc*, *hGLuc*, *pGLuc*, *EpGLuc*, and *KpGLuc*) were obtained by PCR procedures using the amino-terminal primers of wGLuc/N-EcoRI (5′ gccGAATTC AAA CCA ACT GAA AAC AAT GA 3′, *Eco*RI site underlined) for *wGLuc* or hpEKGLuc/N-EcoRI (5′ gccGAATTC AAG CCC ACC GAG AAC AAC GA 3′) for *hGLuc*, *pGLuc*, *EpGLuc*, and *KpGLuc*, with the primer of pcDNA3.1/BGH-R (3′ GGAGCTGACACGGAAGAT 5′) at the carboxyl-terminal region in the vector. The fragments were