Protein Expression and Purification 109 (2015) 47–54

Contents lists available at ScienceDirect

Protein Expression and Purification

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

Codon optimization of genes for efficient protein expression in mammalian cells by selection of only preferred human codons

Satoshi Inouye [a,*], Yuiko Sahara-Miura [a], Jun-ichi Sato [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 5 January 2015
and in revised form 31 January 2015
Available online 7 February 2015

This paper is dedicated to the late Professor J. Woodland Hastings (deceased on August 6, 2014).

Keywords:
Codon optimization
Synthetic gene
Luciferase
Photoprotein
Reporter gene

ABSTRACT

A simple design method for codon optimization of genes to express a heterologous protein in mammalian cells is described. Codon optimization was performed by choosing only codons preferentially used in humans and with over 60% GC content, and the method was named the "preferred human codon-optimized method." To test our simple rule for codon optimization, the preferred human codon-optimized genes for six proteins containing photoproteins (aequorin and clytin II) and luciferases (*Gaussia* luciferase, *Renilla* luciferase, and firefly luciferases from *Photinus pyralis* and *Luciola cruciata*) were chemically synthesized and transiently expressed in Chinese hamster ovary-K1 cells. All preferred human codon-optimized genes showed higher luminescence activity than the corresponding wild-type genes. Our simple design method could be used to improve protein expression in mammalian cells efficiently.

Introduction

Genetic codes comprise 64 nucleotide triplets (codons) encoding 20 amino acids and 3 stop codons. The frequencies of codon usage among organisms differ markedly. At present, codon optimization of heterologous protein genes has been applied to achieve optimum expression of proteins in specific host cells such as bacteria, fungi, yeasts, plants, and mammals [1–5]. The codon-optimized genes for proteins were chemically synthesized using the suitable codons for the host cells without altering the amino acid sequences. A method of codon optimization for proteins has been proposed using several software applications [4,5].

To express a heterologous protein in mammalian cells, the overall proportions of usage of each codon were altered to closely match human codon usage (http://www.kazusa.or.jp/codon/) [6]. Further, nucleotide sequences of rare codons, restriction enzyme sites, recognition sites of transcription factors, cryptic splice sites (GT-AG sequence), and the polyadenylation signal sequence (AATAA) were avoided in designing codon-optimized genes. In addition, a high GC content of genes was also proposed [4,5]. We have named the method of gene synthesis by these rules as the "human codon-optimized method" in this article.

Here we propose a simple design rule for preparing codon-optimized genes for expression it in mammalian cells by selection of only preferentially used human codons and named our design method as the "preferred human codon-optimized method". Based on our simple design rule for codon optimization, we have synthesized preferred human codon-optimized genes and confirmed the reliability of our method for protein expression in mammalian cells. To evaluate our design rule of codon optimization, we chose six proteins catalyzing the bioluminescence reactions as model proteins including photoproteins [7,8], coelenterazine-utilizing luciferases [9–14], and firefly luciferases [15,16]. These proteins showed no sequence homology, and the bioluminescence reactions allow a sensitive, quantitative, rapid and reproducible assay. Our simple design rule for codon optimization offers insight into gene expression in mammalian cells.

Materials and methods

Gene syntheses and construction of expression vectors in mammalian cells

The synthetic genes for a wild-type codon gene (*wLcLuc* [16]), a human codon-optimized gene (*hAQ*), and preferred human codon-optimized genes (*pAQ*, *pCLII*, *pGLuc*, *pRLuc*, *pPyLuc*, and *pLcLuc*) were obtained from Eurofins Genomics K.K. (Tokyo, Japan). The

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

http://dx.doi.org/10.1016/j.pep.2015.02.002
1046-5928/

Table 1
Gene lists of photoproteins, coelenterazine-utilizing luciferases, and firefly luciferases used in this study.

Photoprotein & luciferase	Gene codon type (abbreviation)	Expression plasmid	GenBank accession No.	Gene reference
(i) Photoprotein				
Aequorin (AQ)	Wild (*wAQ*)	pcDNA3-wAQ	L29571	8
	Humanized (*hAQ*)	pcDNA3-hAQ	LC006262	This work
	Preferred (*pAQ*)	pcDNA3-pAQ	LC006263	This work
Clytin II (CLII)	Wild (*wCLII*)	pcDNA3-wCLII	AB360785	9
	Preferred (*pCLII*)	pcDNA3-pCLII	HJ241347	This work
(ii) Coelenterazine-utilizing secretory luciferase				
Gaussia princeps luciferase (GLuc)	Wild (*wGLuc*)	pcDNA3-wGLuc	AY015993	10
	Humanized (*hGLuc*)	pcDNA3-hGLuc	–[a]	11
	Preferred (*pGuc*)	pcDNA3-pGLuc	LC006266	This work
Mutated 19 kDa protein of *Oplophorus gracilirostris* luciferase (KAZ)	Humanized (*nanoLuc*)	pcDNA3-GLsp-nanoLuc	JQ437370	13
	Preferred (*nanoKAZ*)	pcDNA3-GLsp-dnKAZ	AB823628	14
(iii) Coelenterazine-utilizing luciferase				
Renilla reniformis luciferase (RLuc)	Wild (*wRLuc*)	pcDNA3-wRLuc	M63501	9
	Humanized (*hRLuc*)	pcDNA3-hRLuc	AY738226	–
	Preferred (*pRLuc*)	pcDNA3-pRLuc	LC006267	This work
(iv) Firefly luciferase				
North American firefly (*Photinus pyralis*) luciferase (PyLuc)	Wild (*wPyLuc*)	pJN-wPyLuc-sv	M15077	15
	Humanized (*hPyLuc*)	pGL4.13(luc2/sv40)[b]	AY73825	–
	Preferred (*pPyLuc*)	pJN-pPyLuc-sv	LC006854	This work
Japanese firefly (*Luciola cruciata*) luciferase (LcLuc)	Wild (*wLcLuc*)	pJN-wLcLuc-sv	M26194	16
	Preferred (*pLcLuc*)	pJN-pLcLuc-sv	LC006265	This work

[a] Nucleotide sequence is obtained from http://nanolight.com/NanoLights.html.
[b] *hPyLuc* from pGL4.13(luc2/sv40) has the mutations at Asn50 to Asp50, Asn119 to Gly119, and Ser-Lys-Leu (SKL) to Ile-Ala-Val (IAV) at the carboxyl terminus.

cDNA fragments of wild-type genes for *wAQ*, *wCLII*, *wRLuc*, *wGLuc*, and *wPyLuc* were obtained from pAQ440 [7], pCL31 [8], pRL-CMV (Promega, Madison, WI) [9], pBlue-wGLuc (Prolume, Pinetop, AZ) [10], and pT3/T7-luc (Clontech, Palo Alto, CA) [15], respectively, by PCR using appropriate synthetic primers containing the Kozak consensus sequence. To construct the expression vectors, the *Hin*dIII–*Xba*I fragments of photoprotein genes and coelenterazine-utilizing luciferase genes were inserted into the *Hin*dIII–*Xba*I sites of a pcDNA3 vector (Invitrogen, Carlsbad, CA) to give the corresponded expression vectors in mammalian cells under the control of cytomegalovirus (CMV)[1] promoter (Table 1 and Fig. 1A). For expression of firefly luciferase, the DNA fragments of firefly luciferase genes were replaced with the *Hin*dIII–*Xba*I fragment of *hPyLuc* in pGL4.13 (luc2/sv40, Promega) to produce the expression vectors of pJN-wPyLuc-sv, pJN-pPyLuc-sv, pJN-wLcLuc-sv, and pJN-pLcLuc-sv, which were under the control of the simian virus 40 (SV40) promoter (Table 1 and Fig. 1). The nucleotide sequences of genes were confirmed by DNA sequencing. The expression vectors for the mutated 19 kDa protein of *Oplophorus gracilirostris* [12], named nanoKAZ and nanoLuc, and the human codon-optimized *Gaussia* luciferase gene (*hGLuc*) were described previously [14,17,18].

Cell culture

A Chinese hamster ovary-K1 (CHO-K1) cell line was cultured in Ham's F-12 medium (Wako Pure Chemicals, Osaka, Japan) supplemented with 10% (v/v) heat-inactivated fetal calf serum (HyClone, US or Biowest, France), 100 units/mL of penicillin G, and 0.1 mg/mL of streptomycin (Wako Pure Chemicals) at 37 °C in a humidified atmosphere of 5% CO_2. For culturing human embryonic kidney 293 (HEK293), human hepatocellular carcinoma (HepG2), SV40-transformed green monkey kidney (COS-1), HeLa, and human gingival fibroblast (Gin-1) cells, DMEM medium (Wako Pure Chemicals) was used.

Transient expression of photoprotein and luciferase genes in CHO-K1 cells

The *Escherichia coli* strain JM83 or DH5α carrying the expression vector was cultured in Luria–Bertani medium, and the plasmid was prepared using a QIAGEN HiSpeed Midi kit according to the manufacturer's protocol. The concentration of DNA was determined by absorbance at 260 nm (1 OD = 50 μg/mL).

(i) Expression of photoprotein: CHO-K1 cells (1×10^5 cells in 2 mL of culture medium) were seeded in a 6-well plate ($n = 3$, Nunc, Cat. No. 140675) and cultured for 24 h. Before transfection, the culture medium was replaced with 2 mL of Ham's F-12 medium containing 10% fetal calf serum without antibiotics and the cells were incubated for 3 h. The mixture of the expression vector (1 μg) and FuGENE HD (3 μL, Promega) in 100 μL of serum-free medium was added to the cells. After incubation for 24 h, the cells were washed three times with 2 mL of PBS (D-PBS (−), Wako Pure Chemicals), suspended in 1 mL of PBS, and collected with a scraper. Cell extracts were prepared by sonication using a Branson (Danbury, CT) model 250 sonifier for 3 s and used for the assay.

(ii) Expression of luciferase: CHO-K1 cells (1×10^5 cells in 0.5 mL of culture medium without antibiotics) were seeded in a 24-well plate ($n = 4$, Falcon, Cat. No. 353047) and cultured for 24 h. The mixture of the expression vector (0.5 μg) and FuGENE HD (1.5 μL) in 25 μL of serum-free medium was added to the cells. After transfection for 24 h, the culture medium was recovered and the cells were

[1] *Abbreviations used:* w, wild-type codon; h, human codon-optimized; p, preferred human codon-optimized; AQ, aequorin; CLII, clytin II; GLuc, *Gaussia princeps* luciferase; KAZ, the catalytic 19 kDa protein of *Oplophorus gracilirostris* luciferase; nanoKAZ and nanoLuc, a KAZ mutant with 16 amino acid residue substitutions; RLuc, *Renilla reniformis* luciferase; PyLuc, *Photinus pyralis* luciferase; LcLuc, *Luciola cruciata* luciferase; CMV, cytomegalovirus; SV40, simian virus 40; GL_{sp}, signal peptide sequence of *Gaussia* luciferase for secretion; CTZ, coelenterazine; I_{max}, maximum intensity of luminescence; *Int.*, integrated intensity of luminescence; rlu, relative light units.