Biochimica et Biophysica Acta xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

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

Review

Bacterial expression and re-engineering of *Gaussia princeps* luciferase and its use as a reporter protein

Nan Wu, Tharangani Rathnayaka, Yutaka Kuroda *

Department of Biotechnology and Life Science, Graduate School of Engineering, Tokyo University of Agriculture and Technology, 2-24-16 Nakamachi, Koganei-shi, Tokyo 184-8588, Japan

ARTICLE INFO

Article history:
Received 31 March 2015
Received in revised form 24 April 2015
Accepted 15 May 2015
Available online xxxx

Keywords:
Bioluminescence imaging
Protein folding
Protein engineering
Protein domain
Disulfide bonds
Bioinformatics
Structure prediction

ABSTRACT

Bioluminescence, the generation of visible light in a living organism, is widely observed in nature, and a large variety of bioluminescent proteins have been discovered and characterized. Luciferase is a generic term for bioluminescent enzymes that catalyze the emission of light through the oxidization of a luciferin (also a generic term). Luciferase are not necessarily evolutionary related and do not share sequence or structural similarities. Some luciferases, such as those from fireflies and *Renilla*, have been thoroughly characterized and are being used in a wide range of applications in bio-imaging. *Gaussia* luciferase (GLuc) from the marine copepod *Gaussia princeps* is the smallest known luciferase, and it is attracting much attention as a potential reporter protein. GLuc identification is relatively recent, and its structure and its biophysical properties remain to be fully characterized. Here, we review the bacterial production of natively folded GLuc with special emphasis on its disulfide bond formation and the re-engineering of its bioluminescence properties. We also compare the bioluminescent properties under a strictly controlled in vitro condition of selected GLuc's variants using extensively purified proteins with native disulfide bonds. Furthermore, we discuss and predict the domain structure and location of the catalytic core based on literature and on bioinformatics analysis. Finally, we review some examples of GLuc's emerging use in biomolecular imaging and biochemical assay systems.

1. Introduction

Luciferase is a generic name for enzymes that emit light by catalyzing substrates referred to as luciferin ([1], pages xix–xxi). A large variety of luciferases from different biological and evolutionary origins have been identified, however, few have been thoroughly characterized and fewer still have been cloned and used in analytical applications. The most commonly used luciferase in experimental imaging applications originates from the North American firefly (*Photinus pyralis*) [2]. Firefly luciferase (FLuc)'s molecular weight is 62 kDa (Table 1) [3]. It catalyzes the substrate benzothiazole (D-luciferin) in the presence of oxygen, ATP and Mg^{+2}, and emits green light with a peak emission of 562 nm [4]. Its structure has been solved by X-ray crystallography [5]. Several FLuc variants with different luminescence characteristics have been designed, which has enlarged its application range [6].

Renilla luciferase (RLuc) derived from sea pansy (*Renilla reniformis*) is also well characterized. It has a molecular weight of 36 kDa [3,7], and it catalyzes coelenterazine to produce the excited coelenteramide state (Table 1) emitting a blue light with a peak emission of 480 nm. Unlike FLuc, RLuc does not require ATP for activity, and it is becoming an essential bioluminescence marker for in vivo imaging [8–10]. A physically and chemically stabilized RLuc variant with improved bioluminescence properties [11] and its X-ray crystal structure was reported [12]. In addition to the well known FLuc and RLuc, luciferases from various organisms having distinct sequences, structures, and biochemical properties have also been reported [13–16] (Table 1).

Gaussia luciferase (GLuc) is a coelenterazine-dependent luciferase (Table 1) identified from the marine copepod *Gaussia princeps* [15] (Fig. 1), which lives in temperate and tropical waters worldwide [17]. GLuc, which catalyzes the production of a bright blue light, is secreted by *Gaussia princeps* in combination with rapid swimming probably as a defense mechanism [18]. GLuc's sequence (mRNA, complete cds) was first reported in 1999 [15]. It is presently attracting much attention for bio-imaging applications because of its small size and strong luminescence activity, as demonstrated by over two hundred recently reported examples. Here, we review literature on GLuc, with a special emphasis on the formation of disulfide bonds in bacterially expressed GLuc and its functional re-designing. In addition, we discuss its structural and functional properties in the light of the literature and of bioinformatics analysis of its amino acid sequence.

2. Biochemical properties of GLuc

With 185 amino acids and a molecular mass of 19.9 kDa, GLuc is the smallest luciferase reported so far (Table 1, Fig. 4A). GLuc's

Abbreviations: ATP, adenosine triphosphate; HPLC, high performance liquid chromatography

* Corresponding author. Tel./fax: +81 42 388 7794.
E-mail address: ykuroda@cc.tuat.ac.jp (Y. Kuroda).

http://dx.doi.org/10.1016/j.bbapap.2015.05.008
1570-9639/

t1.1 **Table 1**
t1.2 Comparison of GLuc with selected luciferases.

t1.3 Species	Size (aa)	Molecular Wt (kDa)	Cofactors	Substrate	Isoelectric point	Cysteine number (frequency, %)	Reference
t1.4 *Gaussia princeps* (Copepod)	185	20	None	Coelenterazine	6.8	11 (5.9)	[3]
t1.5 *Photynus pyralis* (Firefly)	550	62	ATP, Mg^{+2}	D-Luciferin (benzothiazole)	6.59	4 (0.7)	[3]
t1.6 *Renilla reniformis* (Sea pansy)	311	36	None	Coelenterazine	5.8	3 (1.0)	[3]
t1.7 *Oplophorus*	196	21	None	Coelenterazine	5.0	1 (0.5)	[16]
t1.8 *Pleuromamma*	198	22	N/A	N/A	6.6	17 (8.6)	[15]
t1.9 *Vibrio harveyi* (LuxA)	355	37	Long-chain aldehyde	Reduced riboflavin phosphate (FMNH2)	4.77	8 (2.2)	[13,64]
t1.10 *Lingulodinium polyedrum*	1241	137	None	Dinoflagellate luciferin	7.31	29 (2.3)	[14]

t1.11 FLuc contains four free cysteines. Little influence on its luminescence was observed by changing them to serine [65]. RLuc contains three free cysteines, and two of them affect its lumi-
t1.12 nescence property when mutated to alanine [66].

bioluminescence intensity is significantly stronger (over 200 fold) than that of FLuc and RLuc [3], which are commonly used as reporter proteins. Alike RLuc, GLuc catalyzes the substrate, coelenterazine, to produce the excited state of coelenteramide by emitting blue light (Fig. 2) with a peak emission at 480 nm [3]. Unlike FLuc, GLuc does not require ATP or other co-factors for activity. The bioluminescence activity of GLuc is highly stable under a wide range of temperatures (90% and 65% of the luminescence intensity is retained after a 30-minute incubation at 60 °C and 90 °C, respectively) [19,20] and in an acidic environment (down to pH 1.5) [21]. GLuc's bioluminescence is salt concentration dependent [19,22] and activity reaches a maximum at 50 mM NaCl [19]. GLuc's bioluminescence is strongly inhibited by heavy metal ions (Cu^{2+}) and stimulated by monovalent ions (Cl^-, I^-, Br^-) [22]. GLuc is highly specific for coelenterazine, and shows narrow substrate specificity unlike RLuc, which catalyzes coelenterazine, and also several analogs as well [23].

3. Expression of recombinant GLuc

Various expression systems for GLuc have been reported. GLuc was cloned and expressed in mammalian [3,24], bacterial [21], insect [25], and green algae [26] cells. Expression in *Escherichia coli* is an important step toward re-designing a protein, not only because it provides a handy and inexpensive method for producing recombinant proteins, but also and foremost because genetic modifications are most easily introduced and assessed with this system [27]. To our knowledge, the first quantitative analysis of GLuc expression in *E. coli* was reported by Verhaegen M and Christopoulos T.K. [18]. They designed an *E. coli* overexpression system that produces a biotinylated bap (biotin acceptor peptide)-GLuc fusion protein. The final yield of GLuc purified by Avidin affinity chromatography from the soluble fraction of the *E. coli* culture was 0.55–0.69 mg/L, but nearly 60% of the luminescence was lost during affinity chromatography purification [18].

Fig. 1. Picture of *Gaussia princeps* provided by Professor Russel Hopcroft, University of Alaska — Fairbanks (UAF), and reproduced with the author's permission.

4. Multiple disulfide bonds and expression of native GLuc in *E. coli*

GLuc contains eleven cysteines corresponding to an amino acid frequency of 5.95%, which is in sharp contrast to FLuc and RLuc, containing respectively four and three non-disulfide bonded cysteines (Table 1). Ten of the GLuc's cysteines form disulfide bonds [22,28], whereas the 11th cysteine is located in the pro-sequence, which presumably helps GLuc's folding by reshuffling non-native disulfide bonds in vivo [29,30]. Disulfide bonded proteins are usually difficult to express in *E. coli* in a native form, because *E. coli* 's cytoplasm provides a reducing environment, and cysteines are air-oxidized upon cell breakage and during purification, often resulting into non-native disulfide bonds [27]. Even if the cysteines were oxidized in the cytoplasm, *E. coli* does not have the molecular machinery to reshuffle non-native disulfide bonds into the correct ones, and most multi-disulfide proteins thus turn into functionally inactive aggregates [27].

To overcome the problem of non-native disulfide bonds, GLuc was expressed in the *E. coli* periplasm using a prokaryotic secretion signal peptide, the pelB leader sequence [31,32]. However, the yield and specific activity of the secreted GLuc were relatively low (250 μg/200 mL culture), and the pelB cleavage during protein expression in *E. coli* was inefficient [31]. Cell free systems, where the protein folding environment can be controlled, have been used in an attempt to assist native disulfide bonds formation [28,33]. In principle, some proteins can be recovered from insoluble aggregates by optimizing the refolding conditions [34]. However, with as much as five disulfide bonds, the refolding of GLuc appears to be a difficult task [19]. Finally, low expression temperature (15 °C) increases the fraction of natively folded GLuc, but also decreases the final protein yield [19].

Formation of non-native disulfide bonds in recombinant proteins is closely related to the low solubility of miss-folded proteins. Aggregation often competes with the folding reaction [20], but GLuc, like most small proteins, is intrinsically able to fold into a native structure [29]. Thus, a GLuc with enhanced solubility is anticipated to increase the yield of well folded protein with native-like disulfide bonds. Recently, various attempts for increasing the solubility of GLuc have been reported. In one of them, GLuc was fused to a highly soluble IgG-binding domain of protein A (ZZ-domain, 116 amino acid residues) [35], but it remains unclear whether the fused GLuc was solubilized in a native, fully functional form with the correct disulfide bonds. On the other hand, we constructed a GLuc variant with a Solubility Enhancement Peptide tag (SEP tag [36–38]), consisting of nine aspartic acid tags, attached to its C terminus, which can significantly increase protein solubility. As a result, 98% of the tagged GLuc was expressed in the soluble fraction at 25 °C, which was fully functional and formed native-like disulfide bonds [20].

Please cite this article as: N. Wu, et al., Bacterial expression and re-engineering of *Gaussia princeps* luciferase and its use as a reporter protein, Biochim. Biophys. Acta (2015), http://dx.doi.org/10.1016/j.bbapap.2015.05.008