Marine Genomics 2 (2009) 103–111

Contents lists available at ScienceDirect

Marine Genomics

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

The occurrence of eukaryotic type III glutamine synthetase in the marine diatom *Chaetoceros compressum*

Shigeharu Kinoshita [a,1], Sayaka Isu [a,1], Gen Kaneko [a,1], Hiroshi Yamada [b], Takeya Hara [b], Yasuo Itoh [b], Shugo Watabe [a,*]

[a] *Department of Aquatic Bioscience, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan*
[b] *Central Laboratory Marine Ecology Research Institute, Iwawada, Onjuku, Isumi-gun, Chiba 299-5105, Japan*

ARTICLE INFO

Article history:
Received 14 March 2009
Received in revised form 26 May 2009
Accepted 3 June 2009

Keywords:
GS
GSIII
Molecular evolution
Plankton

ABSTRACT

Glutamine synthetase (GS) has been described as one of the oldest functioning genes and thus a good molecular clock protein. GS is diverged into three distinct forms, type I (GSI), type II (GSII) and type III (GSIII), the last type of which is a member of the most recently discovered family among GSs and thus has been reported from a limited number of prokaryotes. In the present study, we determined the full-length sequence of GSIII from the marine diatom *Chaetoceros compressum*. The 3′ untranslated region of the diatom GSIII gene was composed of a polyadenylation signal followed by a poly $(A)^+$ tail, clearly demonstrating that its mRNA is transcribed from the eukaryotic genome. We also screened available genome databases and identified full-length GSIII sequences from 5 eukaryotic species. These eukaryotic GSIIIs specifically contained regions A–D and a long additional sequence flanking region V toward the C-terminal site, both being specific to GSIII. Phylogenic analysis revealed that eukaryotic GSIIIs are not within a monophyletic relationship with the possible occurrence of lateral gene transfer in GSIII during evolution.

1. Introduction

Glutamine synthetase (GS; EC 6.3.1.2) is an enzyme involved in glutamine synthesis from glutamate and ammonia at the expenditure of ATP. GS also acts in conjugation with glutamate synthase (GOGAT) to incorporate nitrogen into various organic compounds such as purines, pyrimidines and amino sugars. Plants and microorganisms assimilate ammonia mainly through the GS-GOGAT pathway (Forde and Cullimore, 1989). These lines of information lead us to suppose that corresponding genes may exist in all forms of phototroph. Actually, the GS gene (*GS*) has been described as one of the oldest functioning genes (Kimura, 1983) and thus an excellent molecular clock (Pesole et al., 1991).

GS is diverged into three distinct forms, type I (GSI), type II (GSII) and type III (GSIII). Most prokaryotes have GSI composed of 12 identical subunits each of Mr about 50,000 (Valentine et al., 1968; Yamashita et al., 1990). Eukaryotes and certain bacteria contain GSII with 8 identical subunits each of Mr about 40,000 (Meister, 1974), whereas GSI and GSII share a similar primary structure. Although GSI had been thought to distribute only to prokaryotic cells such as bacteria and archaea, Mathis et al. (2000) have recently described the occurrence of GSI in vascular plants. Vertebrates also have GSI named lengsin which was discovered by EST analysis as a major protein in eye lens (Wistow et al., 2002). Lengsin belongs to GSI family but has no enzyme activity (Grassi et al., 2006; Wyatt et al., 2006).

On the other hand, GSIII, initially identified in an anaerobic bacterium *Bacteroides fragilis* that grows in mammalian intestines (Hill et al., 1989), is composed of six identical subunits each of Mr about 75,000 with little homology to GSI and GSII. GSIII was the most recently discovered family among GSs and thus has been described from a limited number of bacteria including a few cyanobacteria, but not from eukaryotes. We previously screened heat stress-responsive genes in the marine diatom *Chaetoceros compressum* (Kinoshita et al., 1998, 2001, 2002) and isolated a cDNA fragment, *HI-9*, that encodes a peptide homologous to a *B. fragilis* GSIII sequence (Kinoshita et al., 1998). Besides our finding, Robertson et al. (1999) isolated *sgsA* which encodes GSII from the other marine diatom *Skeletonema costatum*. However, *sgsA* contained no homologous nucleotide sequence in its complete cDNA with *HI-9*. On the other hand, Robertson and Alberte (1996) purified GS2 from *Skeletonema costatum* and showed its GSIII-like characteristics in molecular size and subunit composition. However, GS2 was described only for internal 28 and N-terminal 25 amino acid residues, but not yet for its complete primary structure. Such limited information about eukaryotic GSIII raises questions about evolutional relationships among three types of GS and further about their functional differences between prokaryotes and eukaryotes.

In this study, we determined the full-length cDNA sequence of *GSIII* from a eukaryote, the marine diatom *Chaetoceros compressum*. The deduced amino acid sequence showed a high homology with

* Corresponding author. Tel.: +81 3 5841 7520; fax: +81 3 5841 8166.
E-mail address: awatabe@mail.ecc.u-tokyo.ac.jp (S. Watabe).
[1] These authors contributed equally to this work.

1874-7787/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.margen.2009.06.003

bacterial GSIIIs, especially in regions which are highly conserved in distinct GS types. Furthermore, we screened *GSIII* nucleotide sequences by using diatom *GSIII* sequence as a probe from available eukaryotic genome databases.

2. Materials and methods

2.1. Cells and culture conditions

The marine diatom *Chaetoceros compressum* collected in the coastal area off Kyushu, Japan, was cultured according to Kinoshita et al. (1998). Briefly, a unialgal culture was carried out in 5 l flasks supplemented with the f/2 medium of Guillard and Ryther (1962). Cells were grown at 20 °C under a 12:12 light:dark cycle illumination by about 5000 lx with cool-white fluorescent lamps. Once a population of 10^5 cells/ml was attained, culture dilution commenced. Culture lots each containing a volume of 0.5 l were transferred to new flasks and added with a fresh medium of 4.5 l.

2.2. RNA preparation

Cells were exposed to heat stress as described before (Kinoshita et al., 2001). Briefly, cells were grown at 20 °C, collected by filtration through a 20 μm mesh, and transferred to flasks for heat treatment. The heat treatment was accomplished in a water bath, where flasks containing algal cultures in a logarithmic growth phase (5–10×10^4 cells/ml) were incubated at 30 °C for 15 min. Following the heat treatment, cells were transferred to the flasks preincubated at 20 °C, grown for 2 h, and collected through a gentle filtration. Cells were quickly frozen in liquid nitrogen and stored at −80 °C until used. Cells stored at −80 °C were powdered in a mortar and pestle containing liquid nitrogen, and total RNAs were extracted using an ISOGEN solution (Nippon Gene, Tokyo, Japan). Poly $(A)^+$ RNA was fractionated using an Oligotex-dT30 super (Takara, Otsu, Japan).

2.3. Rapid amplification of cDNA ends

5′ Rapid amplification of cDNA ends (RACE) was carried out by the oligo-capping method (Maruyama and Sugano, 1994). First strand cDNA was synthesized using the GeneRacer Kit (Invitrogen, Carlsbad, CA, USA). PCR was performed with the first strand cDNA as a template using a primer set of oligonucleotide GeneRacer 5′ Nested Primer (Invitrogen) and HI-9_5RACE4 (5′-AGGAGCGACTTCTTCGATAAT-3′). PCR was carried out with denaturation at 94 °C for 5 min, followed by 25 cycles of denaturation at 94 °C for 30 s, annealing at 60 °C for 30 s, and extension at 72 °C for 2 min. The final extension step was performed at 72 °C for 10 min. Subsequently, nested PCR was performed using a primer set of GeneRacer 5′ Nested Primer (Invitrogen) and HI-9_5RACE5 (5′-TTCAAACTGG TTTGGTGCTAC-3′). The gene-specific oligonucleotide primers HI-9_5RACE4 and HI-9_5RACE5 were designed from a partial cDNA sequence obtained in our previous study (Kinoshita et al., 1998).

To amplify a full-length cDNA clone, two gene specific primers, LHI-9F_0 (5′-TTTCCGAATAGAAAATGCTTACTTC-3′) and LHI-9F_1 (5′-GAA-TAGAAAATGCTTACTTCACCG-3′), were newly designed from a cDNA sequence obtained by 5′RACE. PCR was performed with the first strand cDNA constructed by 5′RACE as a template using a primer set of LHI-9F_0 and GeneRacer 3′ Primer (Invitrogen). Subsequently, nested PCR was performed with a primer set of LHI-9F_1 and GeneRacer 3′ Nested Primer (Invitrogen). PCR was carried out by the same method as described for 5′RACE. PCR products obtained were subcloned into pGEM-T Easy vector (Promega, Madison, WI, USA).

2.4. DNA nucleotide sequencing and sequence alignment

DNA nucleotide sequencing was performed for 5′ and 3′ strands of subclones labeled with Big Dye Terminator Cycle Sequencing Kit using a DNA sequencer model 3100 (Applied Biosystems, Foster City, CA, USA).

Sequence homology search was performed using computer-aided BLAST search and Ensemble genome database. Sequence alignment and phylogenetic analysis were carried out using CLUSTAL W available at the DNA Data Bank of Japan (DDBJ). The program employs the neighbor-joining method (Saitou and Nei, 1987) and Kimura's two-parameter model for phylogenetic tests (Kimura, 1983). One thousand bootstrapped data sets were generated to estimate the statistical significance of the branching. The tree was drawn using the TreeView program (Page, 1996).

3. Results

3.1. Full-length cDNA cloning and sequence analysis

The full-length HI-9 cDNA from the diatom *Chaetoceros compressum* consisted of 2321 nucleotides (nt) and contained the sequence of the HI-9 cDNA fragment previously obtained (Kinoshita et al., 1998) (Fig. 1). An open reading frame consisted of 2151 nt encoding 717 amino acid residues with Mr 78,606. The 3′ untranslated region contained 108 nt followed by a poly $(A)^+$ tail. A putative polyadenylation signal located 26–30 nt upstream from the poly $(A)^+$ tail. These results indicate that HI-9 mRNA is transcribed from the eukaryotic nuclear gene.

Comparison of the deduced amino acid sequence of the HI-9 cDNA with available data by using the BLAST search program revealed 32–33% sequence identity with the sequences of GSIII from bacteria so far reported (Tables 1 and 2), but no significant homology (7–15%) with those of GSI and GSII from prokaryotic and eukaryotic organisms including sgsA (GSII) of diatom *Skeletonema costatum*.

Robertson et al. (1999) determined an N-terminal amino acid sequence of a GSIII-like protein (GS2) purified from *Skeletonema costatum*. The determined sequence contained 25 amino acid residues which overlapped with 28–52 residues of HI-9, but no methionine in contrast to HI-9 (Fig. 1), suggesting that the N-terminal 27 residues of HI-9 would be cleaved during processing to form a mature protein. An apparent molecular weight (Mr) of the putative mature HI-9 was calculated to be 75,720, well coinciding with Mr about 75,000 reported for GSIII from prokaryotes (Hill et al., 1989). Based on these results, HI-9 was defined as diatom GSIII (dGSIII).

3.2. Structural characteristics of eukaryotic GSIII

DDBJ and Ensemble genomic databases were screened for eukaryotic GSIII by using the full-length amino acid sequence of dGSIII as a probe. The full-length GSIII sequences were newly found in 5 unicellular eukaryotic speies: namely, green alga *Ostreococcus lucimarinus* and *O. tauri*, parasitic protozoan *Trichomonas vaginalis*, social amoebae *Dictyostelium discoideum*, and parasitic amoeba *Entamoeba histolytica* (Tables 1 and 2, Fig. 2). All these eukaryotic GSIIIs were predicted from the databases constructed from whole-genome shotgun sequences.

While five major regions I–V are conserved between GSI and GSII (Rawlings et al., 1987) (Fig. 2), corresponding regions of prokaryotic GSIIIs are slightly different from those of GSI and GSII (Pesole et al., 1991). These five regions bear key structural features of the active site and form α/β barrel (Almassy et al., 1986). In eukaryotic GSIIIs also contained regions I–V which were much more similar to prokaryotic GSIIIs than GSIs and GSIIs (Fig. 2).

GSI and GSII contain 19 catalytic residues (Eisenberg et al., 2000) and all these residues were conserved in prokaryotic and eukaryotic GSIIIs (Fig. 2). Prokaryotic GSIIIs have been known to have a long C terminal sequence of about 250 amino acid residues extending from region V (Hill et al., 1989). Such additional sequences are not contained in GSI and GSII and thus the subunit molecular weight of prokaryotic GSIII is larger than those of GSI and GSII. Eukaryotic GSIIIs