Gene 511 (2012) 177–186

Contents lists available at SciVerse ScienceDirect

Gene

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

Intergenic region between TATA-box binding protein and proteasome subunit C3 genes of Medaka function as the bidirectional promoter *in vitro* and *in vivo*

Kazuhiro E. Fujimori *, Kumiko Hazama, Takashi Kawasaki, Tomonori Deguchi, Syunsuke Yuba

Research Institute for Cell Engineering, National Institute of Advanced Industrial Science and Technology (AIST), Nakoji 3-11-46, Amagasaki, Hyogo 661-0974, Japan

ARTICLE INFO

Article history:
Accepted 12 September 2012
Available online 28 September 2012

Keywords:
Bidirectional promoter
Fluorescent protein
Gene expression
ETS
Evolution
Medaka

ABSTRACT

In the genome of eukaryotic organisms, each protein-coding gene has the unique promoter in the 5′-flanking region, and the direction of the promoter is usually controlled unidirectional. In this study, we revealed that the intergenic region between TATA-box binding protein (*tbp*) and proteasome subunit C3 (*psmc3*) genes in Medaka functions as bidirectional promoter *in vitro* and *in vivo*. The *tbp* and *psmc3* genes were allocated as a head-to-head configuration with a 719 bp intergenic region. A comparative analysis of gene arrangement surrounding loci of *tbp* in vertebrates also illustrated that it was unique in Acanthopterygii lineage. The transcription activities were about 1.2 times for *tbp* direction and 0.7 times for *psmc3* direction against that of SV40 promoter in Medaka fibroblasts, respectively. A dual fluorescent reporter assay directly showed that the bidirectional promoter could express two divergent genes concurrently without disruption of RNA polymerase II elongation. In addition, an analysis of sequential deletion of this promoter suggested that the ETS binding site was necessary for maximum expression of downstream gene, and only the ETS binding site was shared from fish to mammals. In mammals, high correlation with CpG islands was observed in such bidirectional promoters, no association was found in the *tbp*/*psmc3* bidirectional promoter in Medaka. These results suggest that molecular machineries of fish bidirectional promoter may be somehow different from those of mammals but the *cis*-acting element for binding ETS transcription factors is essential for divergent gene expression.

1. Introduction

In eukaryotes, each protein-coding gene is regulated transcriptionally by core promoters located at just upstream of the transcription start sites (TSS) and surrounding regulatory regions called as enhancers and silencers. Typically, it is known that there are several sequence motifs, which include the TATA-box, initiator, TFIIB recognition element, and downstream core-promoter element, that are commonly found in core-promoters (Butler and Kadonaga, 2002; Kadonaga, 2002). In contrast with the simple and direct targeting of RNA polymerase in prokaryotes, two specific architectures, the "Mediator" which is a huge protein complex and which bridges promoter recognition and binding to appropriate position in genomic DNA, and the chromatin which is based on a histone octamer, control the initiation and escape of the transcription by RNA polymerase (RNAP) II machinery in eukaryotes (Flanagan et al., 1991; Kelleher et al., 1990; Kornberg, 2005, 2007). Such architectures of the eukaryote-specific transcription machinery are very complex and still unsolved. Challenging to understand them, recent genome-wide studies of TSS analyzed by full-length cDNA sequencing, expression sequence tag (EST), cap-analysis gene expression (CAGE), 5′-end serial analysis of gene expression (5′-SAGE), chromatin immunoprecipitation assay using antibodies against RNAP II and histones, and the other derivative methods could provide us prediction of the existence of the vast numbers of potential core-promoters (Barrera et al., 2008; Barski et al., 2007; Birney et al., 2007; Carninci et al., 2006; Hashimoto et al., 2004; Katayama et al., 2007; Suzuki et al., 2001). In the multi-cellular organisms, however, each promoter activity of each gene should be confirmed experimentally in various kinds of normal cells or in individuals because the interpretation of genome context depends on cell types.

Taking account of the direction (strand) of two adjacent genes in genome, their arrangement is thought to be three cases: (1) tail-to-head manner on the same strand, (2) tail-to-tail manner on the different strand, and (3) head-to-head manner on different strand (Fig. 1). In the first and second cases in general, each regulatory regions (boxes indicated as RR, Fig. 1) or at least core-promoters should exist independently upstream region in each. In the third case, however, the arrangement of regulatory elements is not so simple. If the distance between two adjacent genes is very long, it would be reasonable

Abbreviations: *Tbp*, TATA-box binding protein; *psmc3*, proteasome subunit C3; *eGFP*, enhanced green fluorescent protein; *DsRedExp*, DsRedExpress; *luc*, luciferase; SV40, Simian virus 40; *TF*, transcription factor; ETS, E twenty six; TSS, transcription start site; RNAP, RNA polymerase; *CAGE*, cap-analysis gene expression; *hpt*, hours post transfection; *hpi*, hours post injection.

* Corresponding author at: Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Higashi 1-1-1, Tsukuba, Ibaraki 305-8566, Japan. Tel.: +81 29 849 1228; fax: +81 29 861 6144.
E-mail address: k-fujimori@aist.go.jp (K.E. Fujimori).

0378-1119/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2012.09.077

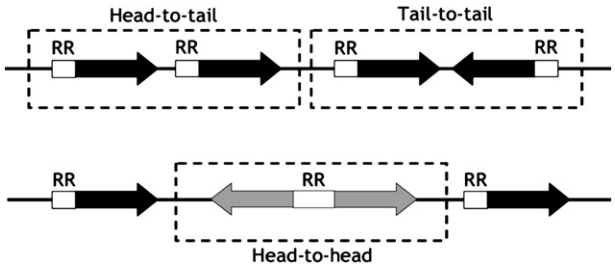

Fig. 1. Arrangement of adjacent genes in genome. Arrangement of adjacent genes and their regulatory regions framed by broken lines are able to be classified as (1) head-to-tail, (2) tail-to-tail and (3) head-to-head (divergent) configurations. Genes are indicated by arrows (black or gray) with direction of transcription and white boxes indicate their regulatory regions (RR).

to assume that there are two inherent promoters for each gene independently as well as the first and second cases. If it is very short, there is a possibility that the essential *cis*-acting elements for their gene expression are shared and overlapped. Recent studies showed that such bidirectional promoters are common and abundant architectures in mammalian genomes (Adachi and Lieber, 2002; Takai and Jones, 2004; Trinklein et al., 2004). In fact, the bidirectional promoters have been given concrete example, such as *sirt3* and *psmd13* (Bellizzi et al., 2007), *mrps12* and *sars2* (Zanotto et al., 2007), *nthl1* and *tsc2* (Ikeda et al., 2000), and *osgep* and *apex* (Ikeda et al., 2002). In addition, the Surfeit locus containing six house-keeping genes named as *surf-1* to *surf-6* has been determined (Duhig et al., 1998). Within six Surfeit genes, two pairs of neighboring Surfeit genes, *surf-5* and *surf-3*, and *surf-1* and *surf-2*, were located as divergent configuration with short intergenic regions, and were expressed ubiquitously and divergently (Lennard and Fried, 1991; Lennard et al., 1994). Furthermore, two genes correlated with aging, *sirt3* and proteasome 26S subunit D13 (*psmd13*) which were linked in head-to-head configuration in human genome had a bidirectional promoter activity in HeLa cells (Bellizzi et al., 2007). Thus, it seems that the bidirectional transcription is a fundamental mechanism to coordinate the expression of two genes in either a positive or negative manner in mammals.

As previously described above, the *tbp* is an essential component of the transcription machinery in eukaryote, and is also known as one of the most highly conserved genes across the known organisms with ubiquitous expression (Hoshiyama et al., 2001). Previous studies showed that *ras* cellular signaling and epidermal growth factor signaling pathways enhance human *tbp* expression *via* the ETS sites of the 5′-flanking region of *tbp* gene in the human cell lines (Foulds and Hawley, 1997; Johnson et al., 2000; Zhong et al., 2004). It has been reported that the juxtaposed genes proteasome subunit B1 (*psmb1*), *tbp* and programmed cell death 2 (*pcd2*) have syntenic relationship in mammals, birds, snake, fly, nematode and fission yeast (Antoniou et al., 2003; Trachtulec and Forejt, 2001). Interestingly, it is known that such relationship was broken, that is, the upstream region of *tbp* was frequently changed in fishes (Trachtulec et al., 2004). But the evolutional change or conservation of regulatory regions of *tbp* orthologs has been unsolved.

From such point of view, we were interested in the molecular machinery of the transcriptional regulation of *tbp* in a flesh water teleost Medaka (*Oryzias latipes*). Medaka has been used as a model animal for biological studies in the laboratory (Furutani-Seiki and Wittbrodt, 2004). Medaka genome believed to be about 800 Mb is smaller than most other vertebrates (Imai et al., 2007). The small genome offers an advantage for accurate analysis of the regulatory regions with high-resolution since core promoters, proximal and distal enhancer/silencer, are also smaller than most other vertebrates. Recently, whole genome shotgun methods have led to a draft sequence of the Medaka genome (Kasahara et al., 2007) which is now available on the Medaka Genome Database website (MGD, http://medaka.utgenome.org/). From our comparative genome analysis of Medaka *tbp* and neighbor genes to the other vertebrates, we found that the proteasome 26S subunit C3 (*psmc3*), which is an evolutionally conserved gene known as the same gene as the human immunodefective virus protein Tat-binding protein-1 (Corn et al., 2003; Sakao et al., 2000), was located at upstream of *tbp* as head-to-head manner with 719 bp intergenic region. The arrangement of these two genes in Medaka evoked us that the intergenic region should contain the functional bidirectional promoter of *tbp* and *psmc3* genes, and it evolved uniquely in fish lineage by dynamic re-arrangement of Medaka genome, for example, inversion, transversion, insertion and deletion. By luciferase and fluorescent gene reporter assays *in vitro* and *in vivo*, we clearly showed that it had the bidirectional promoter activity concurrently and that it shared the essential *cis*-regulatory elements and motifs in this region. The conservation of the amino acid sequence of protein-coding genes in head-to-head arrangement in genome and their evolution of regulatory region between them were also discussed.

2. Materials and methods

2.1. Fish maintenance

Adult fish (Hd-rR and SeeThrough strains provided by the National BioResource Project, http://www.nbrp.jp/index.jsp) were maintained at 27 °C on a 14:10 light:dark cycle. Fertilized eggs were produced by natural mass-mating of adult fish, collected daily, and raised at 28 °C in 0.3% artificial seawater (Rohtomarine II, Lei-Sea) until they reached the desired stage.

2.1.1. In silico prediction and cloning of Medaka tbp and psmc3 orthologs

For acquisition of the Medaka *tbp* and *psmc3* orthologs, we firstly searched the whole-genome shotgun sequence library produced by the Medaka Genome Project (http://dolphin.lab.nig.ac.jp/medaka/) with a TBLASTN program using the amino acid sequences of human and mouse *tbp* as queries. By hand-assembly of the obtained five shotgun reads, GOLWno8358_m19.g1, GOLWno1301_e17.g1, GOLWno966_m08.b1, GOLWno9969_n12.b1 and GOLWno9886_d14.g1, we obtained a partial cDNA sequence of Medaka *tbp*. Then, we surveyed again the Medaka Genome Database (MGD, assemble version 2004/06) using this sequence as a query. As a result, we ascertained that the Medaka *tbp* was allocated in scaffold 4677. As same as the prediction of *tbp*, we surveyed the MGD using human and mouse *psmc3* as queries, and found the Medaka *psmc3* located in the same scaffold. Based upon this prediction, we designed specific primers for *tbp* (TBP-S9: 5′-TAGTGG TTGGCAGTGTTGGAACGG-3′, TBP-AS1232: 5′-TGTAAGAAAACTTGCCAG CGAGG-3′) and *psmc3* (PSMC3-S: 5′-CGCAGTCAGACAGAGCAAAG-3′, PSMC3-AS: 5′-GGCAGTGGCCTTAAGCATAG-3′). To synthesize first strand cDNAs, total RNAs were purified from adult tissues or organs of Hd-rR fish: brain, heart, gill, liver, kidney, spleen, gut, skeletal muscle, testis and ovary. A QuickPrep *micro* mRNA Purification kit (GE Healthcare) was used according to the manufacturer's instructions. Subsequently, reverse transcription was performed with 200 ng of mRNA using an oligo-dT primer to generate single stranded cDNAs according to the manufacturer's instruction of First strand cDNA synthesis kit (GE Healthcare). The PCR amplified fragments were cloned into pGEM-T Easy vector (Promega). The plasmids were prepared by QIAprep Spin Miniprep kit (QIAGEN) and sequenced with the dye terminator method using a Genetic Analyzer 3130 (Applied Biosystems). The resulting DNA sequences were deposited as *tbp* (DDBJ/Accession No. AB362996) and *psmc3* (DDBJ/Accession No. AB362997).

2.2. Phylogenetic analyses of tbp and psmc3

To verify that the resultant genes were the true *tbp* and *psmc3* orthologs in Medaka, the amino acid sequences were aligned and analyzed with the CLUSTAL W program (Thompson et al., 1994). The sequence