Transcriptional regulation of a horizontally transferred gene from bacterium to chordate

Yasunori Sasakura1, Yosuke Ogura1,†, Nicholas Treen1,‡, Rui Yokomori2, Sung-Joon Park2, Kenta Nakai2, Hidetoshi Saiga3, Tetsushi Sakuma4, Takashi Yamamoto5, Shigeki Fujiwara5 and Keita Yoshida1

1Shimoda Marine Research Center, University of Tsukuba, Shimoda, Shizuoka 415-0025, Japan
2The Institute of Medical Sciences, University of Tokyo, Tokyo 108-8639, Japan
3Department of Biological Sciences, Graduate School of Science and Engineering, Tokyo Metropolitan University, 1-1 Minamiohshima, Hachioji, Tokyo 192-0397, Japan
4Department of Mathematical and Life Sciences, Graduate School of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan
5Department of Applied Science, Kochi University, 2-5-1, Akebono-cho, Kochi-shi, Kochi 780-8520, Japan

The horizontal transfer of genes between distantly related organisms is undoubtedly a major factor in the evolution of novel traits. Because genes are functionless without expression, horizontally transferred genes must acquire appropriate transcriptional regulations in their recipient organisms, although the evolutionary mechanism is not known well. The defining characteristic of tunicates is the presence of a cellulose containing tunic covering the adult and larval body surface. Cellulose synthase was acquired by horizontal gene transfer from Actinobacteria. We found that acquisition of the binding site of AP-2 transcription factor was essential for tunicate cellulose synthase to gain epidermal-specific expression. Actinobacteria have very GC-rich genomes, regions of which are capable of inducing specific expression in the tunicate epidermis as the AP-2 binds to a GC-rich region. Therefore, the actinobacterial cellulose synthase could have been potentiated to evolve its new function in the ancestor of tunicates with a higher probability than the evolution depending solely on a spontaneous event.

1. Introduction

Horizontal gene transfer has had a great impact on evolution because newly acquired genes can add new genetic functions to host organisms that can lead to changes in their biological characters [1]. Because gene donor and recipient organisms have different systems of transcriptional control, genes acquired by horizontal transfer must be further evolved in recipient organisms to achieve transcription in appropriate timings and regions. In particular, the evolution of transcriptional control would be extremely difficult in the horizontal transfer events across kingdoms such as between bacteria and multicellular animals. In addition to the different status of genomic DNAs of these groups, the bodies of animals are orderly organized as tissues and organs but these organizations are not present in unicellular bacteria. In animals, genes are expressed in the specific manner according to information of tissues and organs. Hence, genes coming from bacteria need to be incorporated into the tissue- or organ-specific gene expression system of animals for exerting their functions in animal bodies. Taking the complexity of transcriptional regulations of animals into consideration, the evolution of the appropriate transcriptional manner of horizontally transferred bacterial genes would be an improbable event to achieve. Although an accidental event could be a factor propelling...
such an improbable evolution to occur, the improbability might suggest an alternative scenario that an intrinsic compatibility between the transferred gene and recipient organism further promoted the successful evolution of transcriptional regulation.

The extant tunicates possess a gene encoding cellulose synthase [2–4]. Tunicate cellulose synthase is thought to have been acquired by horizontal gene transfer from a bacterial genome [3,4]. Tunicates contain cellulose in their tunic, the mantle layer surrounding their body [5,6]. A major group of tunicates, ascidians, use the tunic to protect their body from predators during their sessile lifestyle. A mutation of the cellulose synthase gene of the ascidian Ciona intestinalis results in the loss of cellulose from the tunic [7], suggesting that this gene is essential for cellulose production in ascidians. Moreover, the mutants of cellulose synthase exhibit defects in starting and maintaining a sessile lifestyle, suggesting acquisition of cellulose synthesizing ability promoted ascidians to evolve their sessile lifestyle. Cellulose synthase of tunicates including C. intestinalis is expressed in the epidermis [2–4], and the ancestor of tunicates must have evolved a system to express this gene in the epidermis.

To understand how tunicates evolved epidermal expression of cellulose synthase, we investigated transcriptional control of C. intestinalis cellulose synthase (Ci-CesA) [3]. We found that AP-2 transcription factor is the key factor responsible for the epidermal expression of Ci-CesA. Because AP-2 can preferentially recognize GC-rich DNA elements, and due to the GC-richness of actinobacterial genome from which tunicate CesA gene derives, we propose a novel mechanism to explain how the tissue-specific expression system of tunicate cellulose synthase successfully evolved that could be an alternative to the hypothesis that tunicate cellulose synthase overcame an improbable event of spontaneous acquisition of a GC-rich AP-2 binding site in the AT-rich genome of the ancestor of tunicates.

2. Material and methods

(a) Constructs

Upstream regions of Ci-CesA and Cs-CesA were amplified by polymerase chain reaction (PCR) and were inserted at the BamHI site of pPD1.27 [8]. The DNA fragment containing the AP-2 binding site of Ci-CesA was amplified by PCR, and was inserted at the PstI sites of pSPPlCiBraK [9], pPDCIKbbasAL [10] and pSPDITPOG [11] for enhancer analyses. The DNA fragments were inserted at the PstI site of the corresponding vectors. TALENs that target the AP-2 binding site of Ci-CesA were assembled using the four-module golden gate method [12–14]. Knockout of the AP-2 binding site was done by introducing in vitro synthesized TALEN mRNAs according to the previous method [15]. The full-name of the plasmid constructs according to the nomenclature rules [16] are shown in electronic supplementary material, table S1.

(b) Introduction of exogenous nucleotides

Plasmid DNAs were electroporated to one cell embryos according to the previous reports [9,17]. RNAs were synthesized in vitro with Megascript T3 kit (Ambion), Poly A tailing kit (Ambion) and Cap structure analogue (New England Biolabs).

The antisense morpholino oligonucleotide (MO) that can disrupt splicing of Ci-AP2L2 was previously described [18]. The RNAs and MOs were microinjected into unfertilized eggs according to the previous study [19]. The concentration of RNA and MO in the injection medium was 100 ng ul⁻¹ and 0.5 mM, respectively.

(c) Fluorescence observation, histochemical staining and whole-mount in situ hybridization

Fluorescent images were taken with a Zeiss fluorescent microscope AxioImager.Z1 and a black and white camera AxioCam MRm. Images were pseudocoloured. Histochemical staining of β-galactosidase was done according to the previous study [19]. Whole-mount in situ hybridization of Ci-AP2L2 was done basically according to the previous study [20]. The modified points are as follows: tailbud embryos were fixed with 1/10 formalin diluted by seawater for a few days at 4°C, and final washing step was done with 0.5× SSC, 50% formamide and 0.1% Tween20 instead of 1× SSC, 50% formamide and 0.1% Tween20.

(d) Gel retardation assay

The entire translated region of Ci-AP2L2 cDNA was inserted into the pCMX vector [21]. The Ci-AP2L2 protein was produced using the TNT-coupled Reticulocyte Lysate System (Promega). The double-stranded DNA probes, named Ci and Cs, contained the putative AP-2 binding sites found within Ci-CesA and CesA upstream regions, respectively. The probes, named mCi and mCs, contained specific point mutations within the binding sites. The Gel retardation assay was carried out, according to the protocol described [22].

(e) Statistical analysis

The GC content of every 100 base pair (bp) window of the genome sequences of C. intestinalis (ftp://ghost.zool.kyoto-u.ac.jp/cgi-bin/gb2/gbrowse/C.intestinalis) and Ciona savignyi (ftp://ftpensembl.org/pub/release-70/fasta/ciona_savignyi/dna/) were scored. When unidentified nucleotides (as represented by n in the genome database) were included in the windows, the windows were omitted from the analyses because we could not specify their GC content. The GC content of the 100 bp windows including the AP-2 binding site of Ci-CesA and CesA at the centre (GCSAP in electronic supplementary material, figure S9) were then scored. The number of 100 bp windows that have a GC content equal to or higher than those of the GCSAPs were divided by the total number of 100 bp windows, to calculate the probability of appearance.

3. Results

(a) AP-2 is responsible for epidermal expression of cellulose synthase

Our previous study [7] has shown that an approximately 2000 bp upstream region from the initiation codon of Ci-CesA is sufficient for epidermal-specific gene expression (electronic supplementary material, figure S1). A series of deletions of the cis element from its 5’ end showed that the region between 450 and 431 bp upstream is necessary for epidermal expression of this gene (electronic supplementary material, figure S1). A search for known transcription factor-binding sites in this region suggested that the region contains a putative binding site of AP-2 (figure 1a), a conserved transcription factor that functions in the epidermis.
of chordates including _C. intestinalis_ [18,23,24]. When the AP-2 binding site in the reporter construct was mutated, epidermal expression of the reporter gene was completely abolished (figure 1b–d; electronic supplementary material, figure S2), suggesting that the AP-2 binding site is essential for epidermal expression of _Ci-CesA_. Furthermore, when the AP-2 binding site in the genome was mutated with a pair of transcription activator-like effector nucleases (TALENs) designed to specifically target the binding site, epidermal expression of _Ci-CesA_ was dramatically reduced (figure 1e,f; electronic supplementary material, figure S3). These data suggest that a single AP-2 binding site is the essential element for driving _Ci-CesA_ in the epidermis whose function could not be compensated by any other elements in the _Ciona_ genome. To show the conservation of the AP-2 binding site among ascidian species, we analysed the _cis_ element of _Cs-CesA_, the orthologous cellulose synthase gene of another ascidian _Ciona savignyi_ [2]. The 1383 upstream region of _Cs-CesA_ drove _LacZ_ reporter gene expression in the epidermis when introduced in _C. intestinalis_ (electronic supplementary material, figure S4a,b), suggesting the conservation of the _cis_ element among these ascidians. A series of deletion and mutation analyses has shown that the _cis_ element of _Cs-CesA_ also possesses one putative AP-2 binding site that is essential for epidermal expression (electronic supplementary material, figure S4b–f).

(figure 1. AP-2 binding site is necessary for expressing _C. intestinalis_ Ci-CesA in the epidermis. (a) The consensus sequence of AP-2 binding sites [23], and the predicted AP-2 binding sites (highlighted in red) found in the _cis_ element of _Ci-CesA_ and _Cs-CesA_. μAP2: mutated AP-2 binding site. (b) A schematic diagram of the _cis_ element analysis. The left side numbers indicate the length of the DNAs from the initiation codon of _Ci-CesA_. The red box represents the AP-2 binding site. The scores of larvae showing epidermal expression of β-galactosidase are described at the right side. NE, not examined, because we could not see the signal in the inner tissues due to staining in the epidermis. Expression in the mesoderm (probably due to vector backbone sequence) assures us to see that the constructs were successfully introduced. (c,d) Expression of β-galactosidase reporter protein in larvae introduced with the reporter construct, as examined by histochemical staining. Epi, epidermis. Mes, mesoderm. (e,f) Disrupting the AP-2 binding site of _Ci-CesA_ in the _C. intestinalis_ genome with a TALEN pair. (e) Design of the TALEN pair for mutating the AP-2 binding site of _Ci-CesA_. The orange, blue, green and red boxes represent modules of the DNA-binding domain of TALENs that, respectively, recognize C, A, T and G nucleotides. The AP-2 binding site (AP2BS) is highlighted in red. (f) Expression of _Ci-CesA_ in TALEN-introduced tailbud embryos, as revealed by whole-mount _in situ_ hybridization. Left is an embryo introduced with the left side TALEN described in panel e as a control. Right is an embryo introduced with left and right TALEN pair described in panel e. The numbers at the right side of the panels illustrate the number of embryos expressing _Ci-CesA_ throughout the epidermis. (Online version in colour.)
Ciona intestinalis has two genes encoding homologues of AP-2, and it was shown that one of them, Ci-AP2L2, is expressed in the epidermis [18,25]. A gel retardation assay showed that Ci-AP2L2 has the binding activity to the putative AP-2 binding sites of both Ci-CesA and Cs-CesA (figure 2a). We next examined the necessity of Ci-AP2L2 for epidermal expression of Ci-CesA. When Ci-AP2L2 was knocked down by an antisense MO [18], expression of Ci-CesA was completely abolished from the epidermis (figure 2b–d). This effect of MO was ameliorated by simultaneously introducing mRNA of Ci-AP2L2 with the MO (figure 2d), suggesting that the effect of MO is specifically caused by disrupting Ci-AP2L2. Taken together, Ci-AP2L2 is essential for epidermal expression of Ci-CesA.

(b) Acquisition of the AP-2 binding site can promote epidermal expression

The above results showed that AP-2 is the transcription factor responsible for epidermal expression of Ci-CesA. This suggests that the tunicate cellulose synthase gene regulatory region acquired a binding site of AP-2 during evolution and this resulted in the expression of cellulose in the epidermis. We investigated whether acquisition of the AP-2 binding site is sufficient for epidermal expression. When an approximately 100 bp-long genomic fragment including the AP-2 binding site was fused to minimal promoters of three C. intestinalis cis element was fused to minimal promoters of three C. intestinalis genes that are not expressed in the epidermis [9–11], the AP-2 binding site and promoter fusions expressed the downstream reporter genes in the epidermis (electronic supplementary material, figure S5–S7). When the AP-2 binding site of the fragment was mutated, the fragment lost the epidermal enhancer activity (electronic supplementary material, figure S8). Therefore, the acquisition of the DNA region including the AP-2 binding site of Ci-CesA is sufficient for inducing epidermal expression. The addition of a minimal AP-2 binding site could not induce epidermal expression (electronic supplementary material, figure S6 and S7), suggesting that the DNA element surrounding the AP-2 binding site has a supportive role for epidermal expression.

(c) Actinobacterial DNA is potent to act as an epidermal enhancer in Ciona

Tunicate cellulose synthases are unusual in having two domains with similarity to cellulose synthase and cellulase of the glycoside hydrolase (GH)-6 family [2–4]. Recent phylogenetic analyses have supported that the tunicate cellulose synthase could have originated from the gene unit of an actinobacterial group. In the actinobacterial genome, genes encoding cellulose synthase and GH-6 cellulase are very close to each other (electronic supplementary material, figure S9a). The cellulose synthase and cellulase genes may have been united to be a single gene during evolution of tunicates.

Actinobacteria are a common bacterial group in the ocean that can be found in the ascidian digestive tube [26]. Actinobacteria are known to possess very GC-rich genomes (over 70% G and C in many species) [27,28], suggesting that the original cellulose synthase of tunicates may have been embedded within a very GC-rich sequence. Indeed, we found that a 100-bp-long stretch around the AP-2 binding sites of Ci-CesA has a significantly higher GC content in contrast with the AT-rich CesA (electronic supplementary material, figure S9b) [29]. We named this region GCSAP, after the GC- rich element surrounding the AP-2 binding site. Similar AT:GC ratios were seen in the DNA sequence surrounding the AP-2 binding site of Cs-CesA (electronic supplementary material, figure S9b). Because animal genomes are generally AT-rich [30], the common ancestor of chordates should have had an AT-rich genome, suggesting that GC-richness of GCSAP might not be the legacy from the genome of a chordate ancestor. Rather, it is
more probable that the GC-rich feature originated from the actinobacterial genome.

Although it is possible to assume that transcriptional regulation of tunicate cellulose synthase was acquired solely by an accidental event, we could suggest an alternative scenario in which the evolution has a higher chance of success, and in which AP-2 could have been selected with higher probability than other transcription factors responsible for tunicate epidermal expression [18,31,32]. AP-2 is known to recognize GC-rich DNA sequences [23]. Indeed, the AP-2 binding sites essential for epidermal expression of Ci-CesA and Cs-CesA are very rich in G and C (figure 1a). As mentioned above, the original cellulose synthase gene was rich in GC. This GC-rich DNA sequence has a high probability of coincidently containing GC-rich AP-2 binding sites, or very small modifications of the DNA sequence could easily make a functional binding site of AP-2. Indeed, we found the identical sequence of Ciona AP-2 binding site around cellulose synthase/cellulase genes of three actinobacterial genomes (electronic supplementary material, figure S9a).

Some bacterial DNA fragments including the AP-2 binding sequence, when fused with promoters of C. intestinalis genes, induced epidermal expression of reporter genes in the Ci-AP2L2-dependent manner (figure 3; electronic supplementary material, figure S10). These results support that actinobacterial genome sequences surrounding cellulose synthase and cellulase genes have the potential to be an enhancer for epidermal expression in tunicates preceding their horizontal transfer to the tunicate genome.

4. Discussion

In this study, we showed that AP-2 is the key transcription factor responsible for epidermal expression of tunicate cellulose synthase. The common ancestor of chordates was thought to have had epidermal AP-2, as was suggested by the conservation of epidermal expression of AP-2 among extant chordates [24]. Therefore, the incorporation of this transcription factor in the regulatory system of cellulose synthase is reasonable. Actinobacterial cellulose synthase may have needed to acquire epidermal expression immediately after the horizontal transfer, because genes that are not being expressed will almost certainly be dispensable for the ancestor of chordates; the bacterial cellulose synthase may have mutated to become functionless soon after horizontal gene transfer if it failed to acquire expression in the ancestor of tunicates. At the same time, it would have been difficult for actinobacterial cellulose synthase to be incorporated into AP-2-dependent transcriptional activation immediately after the horizontal transfer if the evolution depended solely on spontaneous events, such as the acquisition of de novo AP-2 binding sites by accumulation of mutations on the cis element of the cellulose synthase gene. We cannot exclude the possibility that a cellulose synthase gene accidentally overcame these improbable events, such as the coincidental insertion of a bacterial cellulose synthase gene downstream of an AP-2 binding site that was already present in the tunicate genome (figure 4); however, we assume that there was an additional factor that promoted
the evolution of epidermal expression of cellulose synthase in the ancestor of tunicates. Both AP-2 and the actinobacterial genome have the shared characteristic summarized as ‘GC-richness’: AP-2 recognizes GC-rich elements for its DNA-binding, and cellulose synthase of actinobacteria is generally GC-rich due to the conserved genomic feature of this bacterial group. This suggests that the ancestor of tunicates could have evolved epidermal expression of cellulose synthase with higher probability than a spontaneous event. Indeed, some genomic DNA fragments of actinobacteria contain sequences identical to the AP-2 binding site, and these fragments can act as an epidermal enhancer in *Ciona*, indicating that the cellulose synthase gene of this bacterial group could have simultaneously transferred both the gene body and epidermal enhancer during the horizontal transfer event.

AP-2 started epidermal expression of tunicate cellulose synthase by binding to the *cis* element of cellulose synthase after the GC-rich gene was acquired from actinobacteria (figure 4). Even if the GC-rich genome had a good potential to include an element to accept AP-2 binding (or a few mutations in the GC-rich DNA easily make an AP-2 binding site), the probability may not be high enough to create multiple binding sites during evolution. The epidermal enhancer of *Ciona CesA* has only a single AP-2 binding site, unlike transcriptional regulations of many *Ciona* genes that require multiple transcription factor-binding sites [9,18,33]. The simple transcriptional regulation of *Ci-CesA* may be a simple solution to allow an improbable event to occur. After the horizontal transfer event it is likely that both the *cis* element and the open reading frame of tunicate cellulose synthase became richer in AT by accumulating neutral mutations, acquiring introns and optimizing the codons (figure 4). However, GCSAP was likely to accumulate less mutations due to its supportive role for epidermal expression. Finally, our hypothesis may not be restricted to tunicate cellulose synthase: this kind of mutuality of gene donor and recipient organisms could be the key to explain the successful horizontal gene transfers between organisms that are phylogenetically far from each other [34,35].

Figure 4. Evolutionary processes of tunicate cellulose synthase gene. The red and green bars, respectively, represent GC-rich and AT-rich DNA. Organism names are shown in blue colour. BS, binding site. (a) A possible process in which GC richness of actinobacterial genome facilitated the acquisition of an AP-2 dependent epidermal enhancer. (b) A process in which *Ci-CesA* coincidently acquired AP-2 binding site by being introduced downstream of a pre-existing AP-2 binding site in the *Ciona* genome. (Online version in colour.)
References


