

Research Article

Evolution of the B-Block Binding Subunit of TFIIC That Binds to the Internal Promoter for RNA Polymerase III

Sachiko Matsutani

Division of Microbiology, National Institute of Health Sciences, Tokyo 158-8501, Japan

Correspondence should be addressed to Sachiko Matsutani; sachiko@nihs.go.jp

Received 11 October 2013; Revised 17 December 2013; Accepted 19 December 2013; Published 12 February 2014

Academic Editor: Andres Moya

Copyright © 2014 Sachiko Matsutani. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Eukaryotic RNA polymerase III transcribes tRNA genes, and this requires the transcription factor TFIIC. Promoters are within genes, with which the B-block binding subunit of TFIIC associates to initiate transcription. The binding subunits are more than 1000 amino acids in length in various eukaryotic species. There are four regions with conserved sequence similarities in the subunits. The helix-turn-helix motif is included in one of these regions and has been characterized as the *B-block_TFIIC* family in the Pfam database. In the NCBI and EMBL translated protein databases, there are archaeal proteins (approximately 100 amino acids in length) referred to as B-block binding subunits. Most of them contain a *B-block_TFIIC* motif. DELTA-BLAST searches using these archaeal proteins as queries showed significant multiple blast hits for many eukaryotic B-block binding subunits on the same proteins. This result suggests that eukaryotic B-block binding subunits were constituted by repeating a small unit of *B-block_TFIIC* over a long evolutionary period. Bacterial proteins have also been annotated as B-block binding subunits in the databases. Here, some of them were confirmed to have significant similarities to *B-block_TFIIC*. These results may imply that part of the RNAP III transcription machinery existed in the common ancestry of prokaryotes and eukaryotes.

1. Introduction

While bacteria and Archaea have their inherent single RNA polymerases, eukaryotes have multiple types of RNA polymerase. Eukaryotic RNA polymerase III (RNAP III) is one of them and synthesizes tRNA, 5S ribosomal RNA, and other small RNAs (for review [1, 2]). Mammalian short interspersed elements (SINEs), which are retrotransposons, are also transcribed by RNAP III [2]. Most RNAP III promoters are inside the sequences expressed as RNAs, and these internal promoters can be divided into three categories based on their organization and transcription factor dependence [3]. In the category of tRNA genes and SINEs, there are internal promoters (type II promoters) consisting of the A- and B-blocks of short nucleotide sequences. A- and B-block sequences are well conserved in various eukaryotes. These promoters are recognized directly by the transcription initiation factor TFIIC, which is a six-subunit protein [4, 5]. Investigations of TFIIC assembly on DNA in yeasts and human have demonstrated that B-block binding subunit of TFIIC first associates with the B-block of the internal

promoter (see also Table 1) [1]. TFIIC bound to DNA recruits another transcription factor TFIIB, and then TFIIB assembles RNAP III at the start site of transcription.

When the cDNAs for the B-block binding subunits of the yeast, rat, and human TFIICs were cloned in previous studies, their amino acid sequences were compared for similarities (Table 1) [6–9]. However, these sequences are diverse, and no homology was detected between the yeast and mammalian subunits [7, 8]. Four conserved regions were subsequently identified in the B-block binding subunits of animals, plants, and fungi using the accumulated nucleotide and amino acid sequence data of genomes: three conserved regions are located in the N-terminal one-third regions of the subunits and one is in the C-terminal regions (Figure 1) [10]. However, no DNA binding motifs were detected in any of the four regions. Recently, improved programs for motif detection, such as CD-search, revealed that one of the four regions has a helix-turn-helix (HTH) motif, which forms a typical DNA binding structure (region II in Figure 1). This region is shown as a family of *B-block_TFIIC* (PF04182) belonging to the clan *HTH* (CL0123) in the Pfam database, which is a large

TABLE I: B-block binding subunits of TFIICs in yeasts and vertebrates.

Subunit name (other name)	Organism	Length (aa)	GI number	Reference
TFC3p (τ 138)	<i>Saccharomyces cerevisiae</i>	1160	6319317	[6]
Sfc3p	<i>Schizosaccharomyces pombe</i>	1339	19112919	[9]
hTFIIIC220 (TFIIIC α , GTF3C1)	<i>Homo sapiens</i>	2109	101943240	[8]
rTFIIIC220	<i>Rattus norvegicus</i>	2148	19424204	[7]

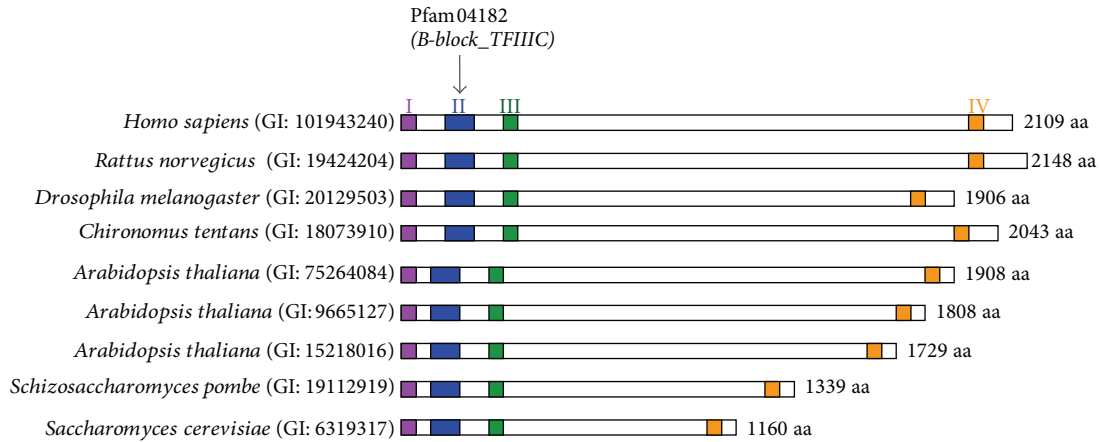


FIGURE 1: Four regions with conserved sequence similarities in the eukaryotic B-block binding subunits. The N-terminal region (shown as region I in the figure), near the N-terminal region (regions II and III), and near the C-terminal region (region IV) [10]. In the Pfam database, which is a large collection of protein families, the eukaryotic B-block binding subunits are shown to contain the specific HTH motif classified as the *B-block_TFIIC* family (PF04182) (<http://pfam.sanger.ac.uk/family/PF04182#tabview=tab0>). The motif is in region II in most eukaryotic subunits. GI: 101943240 and GI: 75264084 are the numbers updated from GI: 4753161 and GI: 25402830, respectively [10].

collection of protein families (<http://pfam.sanger.ac.uk/>): there are currently 324 sequences from 262 species in this family. This domain is considered to directly associate with the B-block sequence of the internal promoter for RNAP III.

RNAP III is generally known to be present in eukaryotes, but not in prokaryotes. However, archaeal and bacterial hypothetical proteins which have been defined or annotated as B-block binding subunits can be found in the translated protein databases of NCBI and EBI. Although some of the information in these databases has not yet been reviewed or confirmed, several proteins have been shown to be members of the *B-block_TFIIC* family in the Pfam database (<http://pfam.sanger.ac.uk/family/PF04182#tabview=tab7>). Interestingly, the promoter sequences of the A- and B-blocks are conserved in bacterial tRNA genes, and bacterial tRNA genes can serve as templates for eukaryotic RNAP III [11].

In this study, the amino acid sequences of the eukaryotic B-block binding subunits and their possible prokaryotic relatives were investigated *in silico* for similarities, and their structural relationships are reported.

2. Methods

2.1. In Silico Analysis. The programs used to compare the primary structures of proteins were as follows: DELTA-BLAST on the NCBI website at http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&BLAST_PROGRAMS=deltaBlast&PAGE_TYPE=BlastSearch&LINK_LOC=BlastHomeAd [12]; Clustal Omega [13] in the EBI website

at <http://www.ebi.ac.uk/Tools/msa/clustalo/>; and Pfam sequence search in the website at <http://pfam.sanger.ac.uk/search?tab=searchProteinBlock> [14]. The search set chosen and algorithm parameters used in each of the searches have been described in Section 3. The alignment parameters in Clustal Omega were used at the default settings. CD-search is the NCBI's interface and this is used to search the Conserved Domain Database for protein or nucleotide query sequences (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) [15]. DELTA-BLAST performs CD-search to construct a position-specific score matrix (PSSM) and then searches a sequence database using the PSSM [12]. The results of CD-search, which were displayed together with those of DELTA-BLAST, were used in this study. DELTA-BLAST was always performed with one iteration. Neighbor-joining phylogenetic tree was constructed by using the Clustal W program in the DDBJ website of <http://clustalw.ddbj.nig.ac.jp/index.php?lang=ja> and NJplot [16–18]. The alignment parameters were used at the default settings.

The databases used were the NCBI protein database (<http://www.ncbi.nlm.nih.gov/protein>) and UniProtKB/TrEMBL (<http://www.uniprot.org/>).

3. Results

3.1. B-Block Binding Subunit-Like Proteins in Prokaryotes. As described in Section 1, several archaeal and bacterial proteins are shown to belong to the *B-block_TFIIC* family

in the Pfam database (see <http://pfam.sanger.ac.uk/family/PF04182#tabview=tab7>). To date, many archaeal sequences have been defined or annotated as B-block binding subunits in the protein database of NCBI. Table 2 shows the features of representative sequences. CD-search showed that these archaeal proteins significantly matched *B-block_TFIIC* (Table 2). Similar results were obtained in the Pfam sequence searches (Table 2). These results confirmed that the archaeal proteins examined here were related to the eukaryotic B-block binding subunit at the amino acid sequence level. However, they were between 100 and 200 amino acids (aa) in length, while eukaryotic subunits are more than 1000 aa in length (Table 2; Figure 1).

Bacterial sequences are also defined or annotated as B-block binding subunits in the NCBI protein database (Section 1). However, much of the information on these proteins has not yet been reviewed and CD-search frequently did not hit *B-block_TFIIC* (data not shown). In the Pfam database of *B-block_TFIIC*, there are two bacterial proteins (<http://pfam.sanger.ac.uk/family/PF04182#tabview=tab7>). The representative proteins in which *B-block_TFIIC* was detected by the CD-search or Pfam sequence search are shown in Table 2. These searches were performed under default conditions. The bacterial proteins were short in length (between 100 and 200 amino acids) (Table 2). *B-block_TFIIC* was not detected in any protein by both programs, and the *E*-values obtained by these searches were mostly higher than those of archaeal proteins (Table 2). Archaeal proteins appeared to be more similar than bacterial proteins to the eukaryotic *B-block_TFIIC* sequences. The archaeal and bacterial proteins referred to here were predicted from coding DNA sequences, and therefore, whether these proteins are actually present in cells and have some functions in vivo remains unknown.

3.2. Structural Relationship between the Eukaryotic B-Block Binding Subunit and Archaeal Protein. As described above, archaeal B-block binding subunits are approximately 100–200 aa in length, while eukaryotic subunits are more than 1000 aa in length. Region II in the eukaryotic subunit, which is approximately 100 aa in length, contains the *B-block_TFIIC* motif (Figure 1; Section 1). DELTA-BLAST was used to examine sequence similarities between eukaryotic and archaeal B-block binding subunits. Searches were conducted on the nonredundant protein sequences of eukaryotes (taxid: 2759) using the archaeal proteins shown in Table 2 as queries and the phrase “B-block binding” as an Entrez query. This Entrez query was used to restrict the search to a subset of proteins referred to as “B-block binding” in the database. Other conditions were set as default. Authentic *B-block_TFIIC* regions in most of organisms, which are annotated in the NCBI protein database, were primarily identified with the lowest *E*-values (data not shown). However, there were several cases in which the archaeal sequence did not hit the *B-block_TFIIC* regions annotated in the NCBI database but showed significant hits for other regions in the same eukaryotic B-block binding subunits. For example, when the *Metallosphaera yellowstonensis* sequence

(GI: 496365863 in Table 2) was used as a query, in the sequences of *Pediculus humanus corporis* (GI: 242025343), *Drosophila willistoni* (GI: 195434252), *Anopheles gambiae* (GI: 347968303), and *Oryzias latipes* (GI: 432847756), the regions of aa positions 375–466, 429–516, 374–429, and 381–475 had significant *E*-values of $2e^{-6}$, $3e^{-6}$, $3e^{-5}$, and $5e^{-4}$, respectively (Figure 2). On the other hand, the *B-block_TFIIC* regions of the four proteins, which are annotated in the NCBI protein database (aa positions 173–242, 176–249, 166–242, and 178–252), were not detected (Figure 2). CD-search was performed and confirmed that the annotations of the four proteins are correct (data not shown). These results are shown in Figure 2 and Supplementary Figure 1 (see Supplementary Material available online at <http://dx.doi.org/10.1155/2014/609865>): a combination of the *Pyrolobus fumarii* sequence (GI: 347523111 in Table 2) and the *Anopheles gambiae* sequences (GI: 347968303) or *Drosophila ananassae* sequence (GI: 194765831); a combination of the *Methanoplanus petrolearius* sequence (GI: 307353829 in Table 2) and the *D. willistoni* sequence (GI: 195434252) or *Nasonia vitripennis* sequence (GI: 345495267); a combination of the *Methanosalsum zhilinae* sequence (GI: 335930125 in Table 2) and the *Drosophila yakuba* sequence (GI: 195472611) or *Drosophila melanogaster* sequence (GI: 20129503); a combination of the *Methanofollis liminatans* sequence (GI: 490137988) and *A. gambiae* sequence (GI: 347968303); a combination of the *Methanoregula formicicum* sequence (GI: 432331009 in Table 2) and the *Pediculus humanus corporis* sequence (GI: 242025343) or *A. gambiae* sequence (GI: 347968303).

As described above, the authentic *B-block_TFIIC* region of the *A. gambiae* sequence (GI: 347968303) was not detected in DELTA-BLAST searches using *M. yellowstonensis*, *P. fumarii*, *M. liminatans*, and *M. formicicum* sequences as queries. However, in the search using the *Methanocella conradii* sequence (GI: 383320206 in Table 2) as a query, the authentic region in the *A. gambiae* sequence was hit at a significant *E*-value ($7e^{-8}$) (Figure 2). There was another significant hit (*E*-value of $5e^{-6}$) in this case and the mosquito hit region mostly overlapped with regions that were detected with the *M. yellowstonensis*, *P. fumarii*, *M. liminatans*, and *M. formicicum* sequences (Figure 2; Supplementary Figure 1). Similar results were obtained, for example, with respect to the *D. melanogaster* and *D. yakuba* sequences (GI: 20129503 and GI: 195472611) when DELTA-BLAST searches were performed using the *M. zhilinae* and *M. conradii* sequences as queries (Figure 2; Supplementary Figure 1). Not a few eukaryotic B-block binding subunits showed significant blast hits for both of their authentic *B-block_TFIIC*s regions and one or more regions different from the authentic region in the same sequences (Figure 2; Supplementary Figure 1): for example, the *M. yellowstonensis* sequence hit the fungal *Punctularia strigosozonata* sequence (GI: 390604017) in four regions, and the *P. fumarii* sequence hit the yeast *Candida tropicalis* sequence (GI: 255729444) in two regions.

These results showed that many B-block binding subunits from yeasts to vertebrates have one or more *B-block_TFIIC*-like regions in addition to the authentic

TABLE 2: Prokaryotic proteins annotated as B-block binding subunits of TFIICs in the NCBI protein database and UniProtKB/TrEMBL.

Protein name	Source organism	Length (aa)	GI number or entry name	aa positions	Alignment with <i>B-block-TFIIC</i> ^a (query/Pfam04182)	E-value
Archaea						
B-block binding subunit of TFIIC	<i>Metallosphaera yellowstonensis</i>	123	496365863	16-63/11-58		$2.87e^{-4}$ ($3.2e^{-7}$)
B-block binding subunit of TFIIC	<i>Pyrobobus fumarii</i> 1A	138	347523111	12-78/1-70		$7.93e^{-8}$ ($7.5e^{-9}$)
B-block binding subunit of TFIIC	<i>Methanocella conradii</i> HZ254	169	383320206	9-68/6-69		$7.18e^{-7}$ ($3.3e^{-6}$)
AsnC family transcriptional regulator	<i>Methanoplanus limicola</i>	115	490179377	16-66/17-70		$1.11e^{-3}$ ($1.5e^{-5}$)
Putative AsnC family transcriptional regulator	<i>Methanoplanus petrolearius</i> DSM 11571	117	307353829	5-70/1-70		$1.07e^{-4}$ ($4.3e^{-5}$)
Conserved hypothetical protein	<i>Methanosalsum zhiilinae</i> DSM 4017	115	335930125	8-62/8-66		$1.31e^{-4}$ ($4.2e^{-6}$)
AsnC family transcriptional regulator	<i>Methanofollis liminatans</i>	113	490137988	11-66/11-69		$1.37e^{-7}$ ($1.7e^{-8}$)
B-block binding subunit of TFIIC	<i>Methanoregula formicicum</i> SMSP	135	432331009	34-87/13-69		$7.32e^{-9}$ ($2.7e^{-9}$)
B-block binding subunit of TFIIC	<i>Methanomethylovorans hollandica</i> DSM15978	120	435851699	9-62/9-66		$3.70e^{-5}$ ($3.3e^{-5}$)
Bacteria						
B-block binding subunit of TFIIC	<i>Streptococcus dysgalactiae</i> subsp. <i>equisimilis</i> SK1250	144	340766721	31-87/1-57		0.02 (ND)
B-block binding subunit of TFIIC	<i>Streptococcus urinalis</i>	151	493791539	31-92/1-62		$7.25e^{-3}$ (ND)
B-block binding subunit of TFIIC	<i>Streptococcus icatuli</i> 707-05	160	356753269	43-107/1-69		$3.40e^{-5}$ (ND)
AsnC family transcriptional regulator	<i>Sulfurimonas autotrophica</i> DSM 16294	102	307721429	20-70/21-71		$2.71e^{-3}$ (ND)
Putative uncharacterized protein	<i>Lactobacillus gasserii</i> 224-1	120	DIYLG5	(10-64/2-55)		ND (0.00023)
SaR protein	<i>Staphylococcus haemolyticus</i> JCSC1435	114	Q4L8F0	37-81/7-51		0.04 (ND)

^aData obtained from the CD-search using the prokaryotic protein sequences as queries. The E-values shown in parentheses are data from the Pfam sequence searches. "ND" means not detected. The aa positions of alignments obtained in the Pfam sequence searches are not shown except the case of DIYLG5. Both searches were performed under the default conditions.

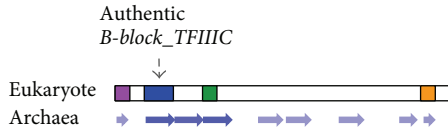


FIGURE 3: Highly schematized cartoon showing distribution of the archaeal B-block binding subunit sequence in the eukaryotic subunit. Colored regions in the eukaryotic subunit correspond to those in Figure 1. Based on all of the results shown in Figure 2 and Supplementary Figure 1, the archaeal subunits (shown as arrows) are placed below the eukaryotic subunit. Arrows are colored more intensely where BLAST hits were frequently detected with lower E -values.

III [10]. In this study, these two sequences were hit with DELTA-BLAST using the *M. conradii* sequence (GI: 383320206) as a query: two regions were detected in each of these proteins (*A. thaliana* aa positions 105–179 and 318–393 and *D. melanogaster* aa positions 180–244 and 371–431) (Figure 2). When the amino acid sequences of the detected regions were searched in the previous alignments shown in Figure 3 of Matsutani [10] by eye, the former sequence corresponded to regions II and the latter corresponded to region III in each of the proteins. The combined alignment of regions II and III in the *A. thaliana* and *D. melanogaster* sequences via the *M. conradii* sequence is shown at the top of Figure 4(b): the amino acid residues that are conserved well in the *B-block_TFIIC* family (<http://pfam.sanger.ac.uk/family/PF04182#tabview=tab3>) were conserved in this alignment. They were also conserved in regions II and III alignments of Figures 3B and 3C in [10]. These results suggest that the *B-block_TFIIC*-like sequence commonly exists as region III in eukaryotic B-block binding subunits (Figure 3).

The C-terminal regions of several fungal B-block binding subunits were hit with E -values lower than threshold in DELTA-BLAST searches using the archaeal sequences as queries (Figure 2): when the *M. yellowstonensis* sequence (GI: 496365863) was used as a query, in the sequences of *Fibroporia radiculosa* (GI: 403413618, 2083 aa), *Stereum hirsutum* (GI: 389742219, 2162 aa), *Punctularia strigosozonata* (GI: 390604017, 2353 aa), and *Trametes versicolor* (GI: 392571338, 2227 aa), the regions of aa positions 2013–2053, 2091–2133, 2284–2321, and 2154–2191 had significant E -values of $4e^{-4}$, 0.002, 0.002, and 0.001, respectively. When the *P. fumarii* sequence (GI: 347523111) was used as a query, the C-terminal region (aa positions 2285–2322) of the *P. strigosozonata* sequence (GI: 390604017, 2353 aa) was hit at an E -value of 0.002 (Figure 2). These fungal C-terminal regions were aligned with the *M. yellowstonensis* and *P. fumarii* sequences using Clustal Omega. The alignments are shown in the middle of Figure 4(b): the amino acid residues conserved in the archaeal sequences and regions II and III were conserved in these fungal C-terminal regions, which corresponded to the latter half of region IV alignment in Figure 3C of [10].

There were no hits in the N-terminal regions of the eukaryotic B-block binding subunits at E -values lower than threshold in DELTA-BLAST searches using the archaeal

sequences as queries. The N-terminal regions in several subunits, were hit at E -values higher than threshold together with the authentic *B-block_TFIIC* regions showing significant E -values. For example, aa positions 1–36 of the *M. yellowstonensis* sequence (GI: 496365863) were similar to aa positions 1–33 of the *Exophiala dermatitidis* sequence (GI: 378726632) (E -value of 0.65), and aa positions 1–27 of the *M. conradii* sequence (GI: 383320206) were similar to aa positions 7–33 of the *Kluyveromyces lactis* sequence (GI: 50302891) (E -value of 0.27). All such N-terminal regions are shown in the lower part of Figure 4(b) as alignments with their relevant archaeal sequences. The amino acid residues conserved in the archaeal sequences and regions II, III, and IV in the eukaryotic B-block binding subunits were conserved in these alignments, although their lengths were short (Figure 4(b)). Furthermore, these alignments were compared with the region I alignment shown in Figure 3A of [10]. These alignments appeared to correspond to the former half of the region I alignment (see the bottom of Figure 4(b)).

Sequences similar to the archaeal B-block binding subunits were sometimes detected out of the four conserved regions in the eukaryotic B-block binding subunits (Figure 2; Supplementary Figure 1): for example, aa positions 16–63 of the *M. yellowstonensis* sequence (GI: 496365863) were similar to aa positions 1539–1584 of the *Selaginella moellendorffii* sequence (GI: 302788556, 1772 aa) (E -value of $7e^{-4}$), and aa positions 3–86 of the *P. fumarii* sequence (GI: 347523111) were similar to aa positions 934–1019 of the *Batrachochytrium dendrobatidis* sequence (GI: 328768215, 2346 aa) (E -value of 0.010). All these results suggest that the eukaryotic B-block binding subunit was mainly constructed by repeated duplication of the *B-block_TFIIC* sequence (Figure 3).

3.4. Investigation of the Primary Structures of the Bacterial B-Block Binding Subunits. As already described and shown in Table 2, the NCBI and EBI protein databases include bacterial proteins defined or annotated as B-block binding subunits. Similarities between the bacterial proteins shown in Table 2 and the *B-block_TFIIC* motif seemed to be unclear, because the E -values were mostly higher than those of archaeal subunits to the *B-block_TFIIC* motif, and both of the CD-search and Pfam search did not detect a *B-block_TFIIC* motif for each of the queries (Table 2). Therefore, possible similarities were investigated using Clustal Omega. The six bacterial B-block binding subunits in Table 2 were first aligned with themselves. The six bacterial proteins were then aligned with the *B-block_TFIIC* cdd sequence (conserved domain's consensus sequence) which was shown in CD-search. As shown in Figure 5(a), the amino acid residues conserved in the alignment consisting only of bacterial proteins corresponded well to those conserved in the alignment of the bacterial proteins and the *B-block_TFIIC* sequence. CD-search of the bacterial proteins showed that they have similarities also to other HTH motifs such as *MarR*, *MarR_2*, and *HTH27*, and the E -values were frequently lower than those of *B-block_TFIIC* (Table 2 and Figure 5(b)). Like *B-block_TFIIC*, *MarR* (PF01047), *MarR_2* (PF12802), and *HTH_27* (PF13463) are members of the clan *HTH* (CL0123)

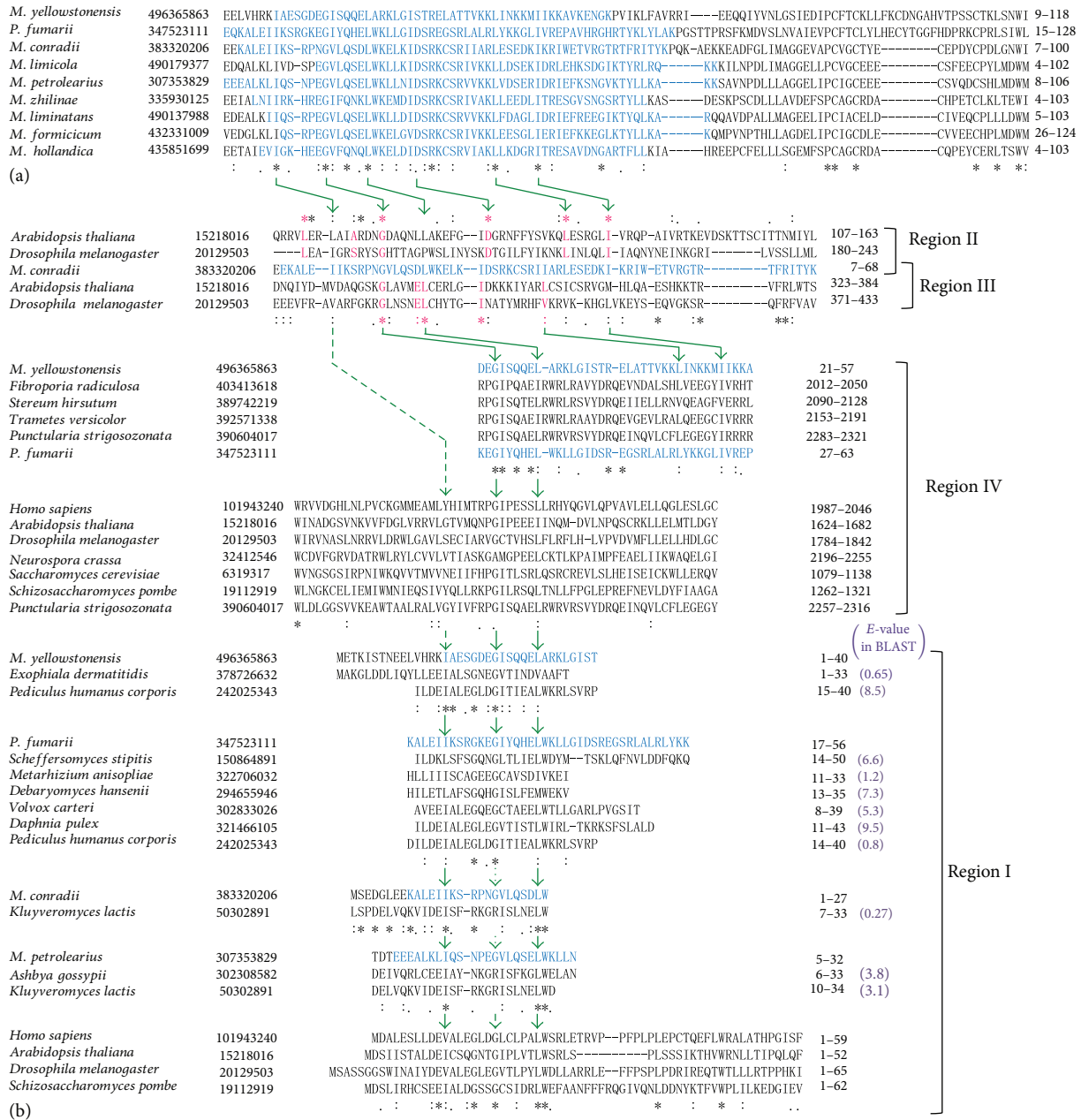


FIGURE 4: Comparison of the primary structures of the archaeal B-block binding proteins with the regions conserved in the eukaryotic B-block binding subunits. Each of the scientific names is followed by the GI number, aa sequence, and aa positions. The *B-block_TFIIC* regions in the archaeal proteins are colored in blue (see Table 2). Green arrows indicate common residues both in Archaea and eukaryotic conserved regions. Green dotted arrows mean no conservation between the two alignments just above and just below the arrows. (a) Clustal Omega alignment of archaeal proteins defined or annotated as B-block binding subunits in the NCBI protein database. Amino acid residues of the N-terminal and C-terminal ends which were not conserved in the nine protein sequences were cropped in this alignment. (b) Clustal Omega alignments of the archaeal sequences with each of regions I, II, III, and IV in the eukaryotic B-block binding subunits. Regions II and III of the *A. thaliana* and *D. melanogaster* proteins (GI: 15218016 and GI: 20129503) were used for comparison with archaeal proteins. This is because these sequences are used in the previous alignments of regions II and III (Figure 3 in [10]) and were detected in DELTA-BLAST searches using the *M. conradii* sequence as a query in this study (Figure 2). These two proteins could precisely link the previous regions II and III alignments to the present alignment with the archaeal sequence. The alignments of regions II and III are shown in a combined form via the *M. conradii* sequence for clarity. Amino acid residues shown in magenta were conserved also in the alignments of Figures 3B and 3C in [10]. For comparison of the archaeal sequences with region IV, the C-terminal regions of the fungal proteins which were detected at significant E-values in Figure 2 were aligned with their related archaeal sequences. Just below the alignment, the C-terminal regions of several eukaryotic B-block binding subunits (region IV) were aligned (see also Figure 3D in [10]). To examine whether the archaeal proteins are related to region I, the N-terminal regions of the eukaryotic B-block binding subunits which were simultaneously detected with the authentic *B-block_TFIIC* regions were visually searched for from the results of DELTA-BLAST. Clustal Omega was then performed. E-values of the matches to the N-terminal regions in DELTA-BLAST were higher than threshold (shown in parentheses next to the aa positions of the Clustal alignments). An alignment of the N-terminal regions of several eukaryotic B-block binding subunits (region I) is shown at the bottom of the Figure (see also Figure 3A in [10]).

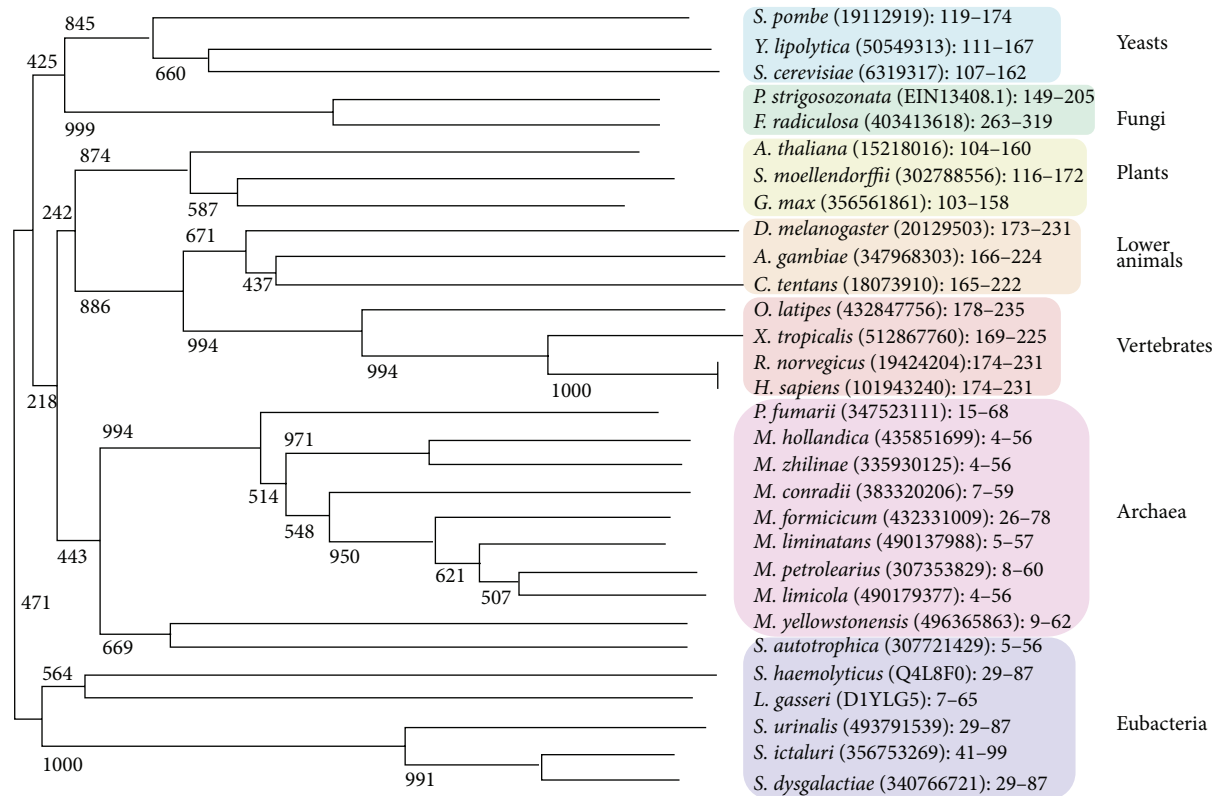


FIGURE 6: Phylogenetic tree for the alignment of *B-block_TFIIC* family. The neighbor-joining method was used and bootstrap values are shown in the tree [16–18]. The aa positions of sequences used for Clustal W are indicated next to the subunit GI numbers.

B-block_TFIIC sequence was common in various eukaryotic B-block binding subunits. Therefore, the *B-block_TFIIC* sequence had possibly begun to be repeated in the first primitive eukaryotes. Alternatively, the sequence might have been originally repeated in the ancient B-block binding subunit. It may be imagined that the repetition was lost in the evolutionary process from the ancient organisms to prokaryotes to leave single copies on the subunits. However, this is opposite to the established view that DNA duplications have contributed to the evolution of organisms [19, 20].

Archaea, eubacteria, and eukaryotes have been thought to possess the transcription machinery specific to each of them [21]. Although only one RNA polymerase in Archaea corresponds to the eukaryotic RNAP II, orthologs of the eukaryotic RNAP III subunit Rpc34 are present there [22]. Rpc34, which is a specific and essential subunit of RNAP III, interacts with the transcription factor TFIIB to participate in RNAP III recruitment [23]. It is suggested that the functional separation of RNAP predates the origin of eukaryotes [22]. Rpc34 contains two domains, which are the N-terminal HTH and the C-terminal Zn-finger domains [22]. Interestingly, HTH regions in several of the archaeal Rpc34s showed the significant similarity to the *B-block_TFIIC* motif in CD-searches (data not shown). Additionally, the *M. yellowstonensis* sequence (GI: 496365863) used in this study

showed similarity to Rpc34: when CD-search was performed, its HTH region was aligned with the HTH region of the Rpc34 motif (PF05158) at an *E*-value of 0.03 (data not shown), although the value was much higher than that of the alignment with *B-block_TFIIC* (Table 2).

Molecular phylogenetic studies using small subunit rRNA and the proteins like actin and α -tubulin place fungi as more closely related to animals than either is to plants [24, 25]. However, it is reported that with respect to the B-block binding subunits of TFIICs, animals appear to be evolutionarily closer to plants than to fungi [10]. This was shown by the results of PSI-BLAST searches using B-block binding subunits as queries. The Pfam website provides a phylogenetic tree of the family of *B-block_TFIIC*, also where animals are more closely related to plants than to fungi (<http://pfam.sanger.ac.uk/family/PF04182#tabview=tab5>). In this tree, prokaryotic *B-block_TFIIC* sequences are not contained. Therefore, the phylogenetic tree was constructed together with the prokaryotic *B-block_TFIIC* sequences. The tree also indicated that animals are evolutionarily closer to plants than to fungi (Figure 6). Although Archaea was placed as more closely related to animals and plants than to fungi (Figure 6), this result seems to be less reliable. When DELTA-BLAST searches were performed in the eukaryotic protein database using archaeal B-block binding subunit sequences as

queries and the phrase “B-block binding” as an Entrez query, many fungal B-block binding subunits were hit with lower *E*-values than those of animals and plants (data not shown).

Some bacterial proteins in databases were confirmed to contain a *B-block_TFIIC* motif in this study. The host species belong only to four genera although whole genome sequencing has been completed in many bacterial genera. Why is the B-block binding subunit absent in the other bacteria? Three possibilities are there. The first one is that in the other bacteria genes encoding the B-block binding subunit proteins had been lost. The protein might have been nonessential for bacterial survival. The second possibility is that the B-block binding subunit had highly diverged in the other bacteria, and those protein sequences cannot be detected with the similarity search programs that are currently used. This possibility is extensively discussed in the following paragraph. The third possibility is that horizontal gene transfers between bacteria and Archaea or eukaryotes. A few cases of horizontal gene transfer from Archaea to bacteria have been reported [26]. However, this possibility seems less plausible than the other two possibilities because A- and B-block sequences of the internal promoter are generally conserved in bacterial tRNA genes.

Bacterial IS1 is a mobile DNA (for review, [27]) and appears to possess the RNAP III promoter sequence in the internal region, like bacterial tRNA genes [28]. The RNAP III promoter-like sequence in IS1 acts as a cis-element to stimulate RNA synthesis from promoters located upstream of the cis-element [28, 29]. The RNAP III promoter sequence of Alu, which is a human SINE, also stimulates RNA synthesis in *E. coli* [28]. The product of the *E. coli artA* gene is shown to bind to the internal region of IS1 and stimulate transcription [29, 30]. Although the primary structure of the bacterial ArtA protein was compared with those of the prokaryotic B-block binding subunits in this study, clear similarities were not found. More improved programs to analyze protein structure may clarify these points in the future. When the structures of the eukaryotic B-block binding subunits were previously investigated in silico, the HTH motif was not detected in any program [10].

The relatives of the RNAP III transcription machinery may have existed in the common ancestry of eukaryotes and prokaryotes due to the presence of the *B-block_TFIIC* motif in archaeal and bacterial proteins and type II promoter sequences in prokaryotic genomes.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

References

- [1] L. Schramm and N. Hernandez, “Recruitment of RNA polymerase III to its target promoters,” *Genes and Development*, vol. 16, no. 20, pp. 2593–2620, 2002.
- [2] R. J. White, “Transcription by RNA polymerase III: more complex than we thought,” *Nature Reviews Genetics*, vol. 12, no. 7, pp. 459–463, 2011.
- [3] E. P. Geiduschek and G. A. Kassavetis, “The RNA polymerase III transcription apparatus,” *Journal of Molecular Biology*, vol. 310, no. 1, pp. 1–26, 2001.
- [4] C. Ducrot, O. Lefebvre, E. Landrieux, J. Guirouilh-Barbat, A. Sentenac, and J. Acker, “Reconstitution of the yeast RNA polymerase III transcription system with all recombinant factors,” *The Journal of Biological Chemistry*, vol. 281, no. 17, pp. 11685–11692, 2006.
- [5] H. Dumay-Odelot, C. Marck, S. Durrieu-Gaillard et al., “Identification, molecular cloning, and characterization of the sixth subunit of human transcription factor TFIIC,” *The Journal of Biological Chemistry*, vol. 282, no. 23, pp. 17179–17189, 2007.
- [6] O. Lefebvre, C. Carles, C. Conesa et al., “TFC3: gene encoding the B-block binding subunit of the yeast transcription factor IIC,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 21, pp. 10512–10516, 1992.
- [7] G. Lagna, R. Kovelman, J. Sukegawa, and R. G. Roeder, “Cloning and characterization of an evolutionarily divergent DNA-binding subunit of mammalian TFIIC,” *Molecular and Cellular Biology*, vol. 14, no. 5, pp. 3053–3064, 1994.
- [8] N. D. L’Etoile, M. L. Fahnestock, Y. Shen, R. Aebersold, and A. J. Berk, “Human transcription factor IIC box B binding subunit,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, no. 5, pp. 1652–1656, 1994.
- [9] Y. Huang, M. Hamada, and R. J. Maraia, “Isolation and cloning of four subunits of a fission yeast TFIIC complex that includes an ortholog of the human regulatory protein TFIIC β ,” *The Journal of Biological Chemistry*, vol. 275, no. 40, pp. 31480–31487, 2000.
- [10] S. Matsutani, “Similarities in transcription factor IIC subunits that bind to the posterior regions of internal promoters for RNA polymerase III,” *BMC Evolutionary Biology*, vol. 4, article 26, 2004.
- [11] W. R. Folk, H. Hofstetter, and M. L. Birnstiel, “Some bacterial tRNA genes are transcribed by eukaryotic RNA polymerase III,” *Nucleic Acids Research*, vol. 10, no. 22, pp. 7153–7162, 1982.
- [12] G. M. Boratyn, A. A. Schäffer, R. Agarwala, S. F. Altschul, D. J. Lipman, and T. L. Madden, “Domain enhanced lookup time accelerated BLAST,” *Biology Direct*, vol. 7, article 12, 2012.
- [13] F. Sievers, A. Wilm, D. Dineen et al., “Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega,” *Molecular Systems Biology*, vol. 7, article 539, 2011.
- [14] S. R. Eddy, “Accelerated profile HMM searches,” *PLoS Computational Biology*, vol. 7, no. 10, Article ID e1002195, 2011.
- [15] A. Marchler-Bauer, S. Lu, J. B. Anderson et al., “CDD: a conserved domain database for the functional annotation of proteins,” *Nucleic Acids Research*, vol. 39, no. 1, pp. D225–D229, 2011.
- [16] M. A. Larkin, G. Blackshields, N. P. Brown et al., “Clustal W and Clustal X version 2.0,” *Bioinformatics*, vol. 23, no. 21, pp. 2947–2948, 2007.
- [17] N. Saitou and M. Nei, “The neighbor-joining method: a new method for reconstructing phylogenetic trees,” *Molecular Biology and Evolution*, vol. 4, no. 4, pp. 406–425, 1987.
- [18] G. Perrière and M. Gouy, “WWW-query: an on-line retrieval system for biological sequence banks,” *Biochimie*, vol. 78, no. 5, pp. 364–369, 1996.
- [19] S. Ohno, *Evolution by Gene Duplication*, Springer, New York, NY, USA, 1970.

- [20] J. Zhang, "Evolution by gene duplication: an update," *Trends in Ecology and Evolution*, vol. 18, no. 6, pp. 292–298, 2003.
- [21] J. S. Fassler and G. N. Gussin, "Promoters and basal transcription machinery in eubacteria and eukaryotes: concepts, definitions, and analogies," *Methods in Enzymology*, vol. 273, pp. 3–29, 1996.
- [22] F. Blombach, K. S. Makarova, J. Marrero, B. Siebers, E. V. Koonin, and J. van der Oost, "Identification of an ortholog of the eukaryotic RNA polymerase III subunit RPC34 in Crenarchaeota and Thaumarchaeota suggests specialization of RNA polymerases for coding and non-coding RNAs in Archaea," *Biology Direct*, vol. 4, article 39, 2009.
- [23] I. Brun, A. Sentenac, and M. Werner, "Dual role of the C34 subunit of RNA polymerase III in transcription initiation," *The EMBO Journal*, vol. 16, no. 18, pp. 5730–5741, 1997.
- [24] P. O. Wainright, G. Hinkle, M. L. Sogin, and S. K. Stickel, "Monophyletic origins of the metazoa: an evolutionary link with fungi," *Science*, vol. 260, no. 5106, pp. 340–342, 1993.
- [25] S. L. Baldauf and J. D. Palmer, "Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 90, no. 24, pp. 11558–11562, 1993.
- [26] E. V. Koonin and Y. I. Wolf, "Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world," *Nucleic Acids Research*, vol. 36, no. 21, pp. 6688–6719, 2008.
- [27] J. Mahillon and M. Chandler, "Insertion sequences," *Microbiology and Molecular Biology Reviews*, vol. 62, no. 3, pp. 725–774, 1998.
- [28] S. Matsutani, "Possible presence and role of the promoter sequence for eukaryotic RNA polymerase III in bacteria," *Genetica*, vol. 131, no. 2, pp. 127–134, 2007.
- [29] S. Matsutani, "The internal sequence of IS1 stimulates RNA synthesis from the IS1 own and exogenous promoters," *Journal of Biological Systems*, vol. 13, no. 3, pp. 313–329, 2005.
- [30] S. Matsutani, "Bacterial ArtA protein specifically binds to the internal region of IS1 in vitro," *Advances in Bioscience and Biotechnology*, vol. 3, no. 7, pp. 869–875, 2012.