The Whole Set of Constitutive Promoters Recognized by RNA Polymerase RpoD Holoenzyme of Escherichia coli

Tomohiro Shimada1,2,3, Yukiko Yamazaki4, Kan Tanaka3, Akira Ishihama1,2*

1Department of Frontier Bioscience, Hosei University, Koganei, Tokyo, Japan, 2Micro-Nano Technology Research Center, Hosei University, Koganei, Tokyo, Japan, 3Chemical Resources Laboratory, Tokyo Institute of Technology, Nagatsuda, Yokohama, Japan, 4Genetics Strains Research Institute, National Institute of Genetics, Mishima, Shizuoka, Japan

Abstract
The promoter selectivity of Escherichia coli RNA polymerase is determined by the sigma subunit with promoter recognition activity. The model prokaryote Escherichia coli contains seven species of the sigma subunit, each recognizing a specific set of promoters. The major sigma subunit, sigma-70 encoded by rpoD, plays a major role in transcription of growth-related genes. Concomitant with the increase in detection of promoters functioning in vivo under various stressful conditions, the variation is expanding in the consensus sequence of RpoD promoters. In order to identify the canonical sequence of "constitutive promoters" that are recognized by the RNA polymerase holoenzyme containing RpoD sigma in the absence of supporting transcription factors, an in vitro mixed transcription assay was carried out using a whole set of variant promoters, each harboring one base replacement, within the model promoter with the conserved -35 and -10 sequences of RpoD promoters. The consensus sequences, TTGACA(-35) and TATAAT(-10), were identified to be ideal for the maximum level of open complex formation and the highest rate of promoter opening, respectively. For identification of the full range of constitutive promoters on the E. coli genome, a total of 2,701 RpoD holoenzyme-binding sites were identified by Genomic SELEX screening, and using the reconfirmed consensus promoter sequence, a total of maximum 669 constitutive promoters were identified, implying that the majority of hitherto identified promoters represents the TF-dependent "inducible promoters". One unique feature of the constitutive promoters is the high level of promoter sequence conservation, about 85% carrying five-out-of-six agreements with -35 or -10 consensus sequence. The list of constitutive promoters provides the community resource toward estimation of the inducible promoters that operate under various stressful conditions in nature.

Introduction
The bacterial RNA polymerase core enzyme with the subunit structure αIIβ′β′⋅ωε is fully active in catalysis of RNA polymerization but is unable to initiate transcription from promoters. Transcription initiation from gene promoters requires an additional dissociable sigma subunit, which reversibly associates with the core enzyme to form the holoenzyme, and directs the core enzyme to recognize promoters for transcription initiation. Most bacteria encode multiple species of the sigma factor [1,2]. In Escherichia coli, seven species of the sigma subunit exist, each recognizing a specific set of promoters [1,3]. The intracellular levels of seven sigma subunits vary depending on cell growth conditions [3,4]. Sigma replacement is a simple mechanism of switching of the pattern of genome transcription [4,5], and the intracellular concentration of seven sigma factors is therefore a primary determinant of the pattern of genome transcription.

The sigma-70, encoded by the rpoD gene, is the primary and major sigma, which is responsible for transcription of most of the genes that are expressed in exponentially growing phase of E. coli [1,2,6,7]. The holoenzyme containing RpoD sigma recognizes in vitro a pair of hexanucleotide sequence elements, TTGACA (-35) and TATAAT (-10) which are situated at 10 and 35 bp upstream, respectively, of transcription initiation sites [2, 6, 8, 9], and a spacer DNA of approximately 17 bp in length separates these two hexanucleotide sequences. This consensus sequence of RpoD promoters was originally proposed based on in vitro transcription assays of some model templates by purified RNA polymerase [10,11]. Later the promoter sequences were determined for a variety of functioning promoters in vivo, one by one, by using ordinary molecular genetic approaches [12,13]. More recently high-throughput experimental systems such as ChiP-chip analysis of RNA polymerase-bound DNA sequences [14,15] and RNA-Seq analysis of whole sets of transcripts [16–18] have been employed for searching functioning promoters in vivo. In parallel, computational approaches have been employed to identify promoters relying on the consensus sequences predicted based on the known promoters. Up to the present time, more than 2,000 experimentally defined or computational predicted promoters have been identified, which are under the direct control of RpoD sigma (see the promoter databases such as RegulonDB [19] and EcoCyc [20]). As a result, the striking diversity appeared in the canonical
promoter sequence, mainly arising from the detection of weak promoters, which are activated only in the presence of positive transcription factors (TFs).

The constitutive promoters have been defined as those that are active in vivo in all circumstances, but it is practically impossible to identify the whole set of constitutive promoters under various environmental conditions. Instead we propose in this study to define the “constitutive promoter” as the promoters that are recognized in vitro by RNA polymerase RpoD holoenzyme alone in the absence of additional supporting proteins with regulatory functions. For the identification of the whole set of constitutive promoters on the E. coli genome, we performed the Genomic SELEX screening system of genome DNA sequences recognized by the RNA polymerase holoenzyme containing RpoD sigma. The Genomic SELEX system was developed to identify DNA sequences recognized in vitro by DNA-binding transcription factors [21] and successfully applied for the identification of regulation targets of many TFs [22]. After SELEX-chip screening, a total of 2,071 sites were identified on the E. coli genome, which are recognized by the RpoD holoenzyme alone. The location of constitutive promoters within these RpoD holoenzyme-binding regions was then computationally identified using the consensus sequence, which was experimentally determined in this study using the in vitro mixed transcription assay [23–25]. The total number of constitutive promoters on the entire E. coli genome was thus estimated to be between minimum 492 and maximum 669. This number of constitutive promoters represents only about one-fourth of the hitherto identified promoters on the E. coli genome. By mapping the holoenzyme-binding peaks were identified, of which 1,075 (40%) are located within spacers, detailed search for the promoters are located within spacers, downstream of transcription units but downstream of another transcription unit located within 315 type-B spacers located upstream of one transcription unit but downstream of another transcription unit (Tables 2 and 3 for the whole list; see also Table S2); and 87 peaks are located within 543 intergenic spacers (average 1.98 sites for each spacer) (Fig. 1A). Since the majority of hitherto identified promoters are located within spacers, detailed search for the constitutive promoters was focused on the total of 1,075 peaks within 543 spacers.

**Results**

**Genomic SELEX screening of RpoD holoenzyme-binding sequences on the E. coli genome**

The constitutive promoters are transcribed in vitro by the RNA polymerase RpoD holoenzyme alone in the absence of supporting TFs. In order to identify the whole set of constitutive promoters on the entire genome of E. coli K-12 M3110, we first performed a mass-screening in vitro of the whole set of sequences that are recognized by the reconstituted RpoD holoenzyme. For this purpose, we prepared sigma-free core enzyme by passing the purified RNA polymerase three times through phosphocellulose column chromatography in the presence of 5% glycerol (note that sigma-core interaction becomes stronger in the presence of increasing glycerol concentration as used for prolonged storage of the holoenzyme [26]). The level of remaining sigma subunits was less than 0.1%, if any, as detected by both protein staining and immuno-staining with antibodies against each of all seven species of E. coli sigma subunits (RpoD, RpoN, RpoS, RpoH, RpoF, RpoE and FecI) (data not shown). The stoichiometry between core enzyme subunits was also checked by immuno-staining with antibodies against the core subunits, RpoA, RpoB, RpoB and RpoZ. The RpoD holoenzyme fully saturated with RpoD sigma was reconstituted by mixing this sigma-free core enzyme and 4-fold molar excess of purified RpoD sigma, which alone does not bind to DNA.

For the identification of DNA sequences that are recognized by RpoD holoenzyme, we employed the Genomic SELEX screening system [21], in which a library of E. coli genome DNA fragments of 200–300 bp in length was used instead of synthetic oligonucleotides with all possible sequences used in the original SELEX method [27–29]. The multi-copy plasmid library of 200–300 bp-long random DNA fragments was constructed from the E. coli K-12 W3110 genome [21]. The library used in this study contained 7-fold molar excess of the entire genome, and thus a single and the same sequence might be included in 7 different overlapping segments on average, thereby increasing the resolution of mapping of SELEX fragments. In each experiment of Genomic SELEX screening, the mixture of genome DNA fragments, which was regenerated by PCR from the genome DNA library, was mixed with 2-fold molar excess of the reconstituted RpoD DNA polymerase holoenzyme, and subjected to Genomic SELEX screening. DNA-holoenzyme complexes formed were recovered using the anti-RpoC antibody, which gave the highest level of RNA polymerase recovery among all the anti-core subunits. RNA polymerase-associated DNA was isolated from the antibody precipitates, amplified by PCR, and subjected to next cycles of SELEX. After two-cycles of SELEX screening, the final products of RpoD holoenzyme-bound DNA fragments were subjected to mapping on the genome using a DNA tilling microarray (Oxford Gene Technology, Oxford, UK) [30–32]. On the DNA tilling array used, the 60 b-long DNA probes are aligned at 105 bp-intervals in the order of E. coli genome sequence, and therefore approximately 300 bp-long SELEX fragments should bind to two or more consecutive probes. This criterion was employed to avoid the background noise of non-specific binding of RpoD holoenzyme-bound DNA fragments to the tilling array.

The sequences with binding affinity to the RpoD holoenzyme formed a number of peaks along the entire E. coli genome. By setting the cut-off level of 2.0% relative to the highest peak located within a spacer upstream of slaI (SsrA smRNA) and downstream of smpB (SmpB trans-translation factor), a total of 2,701 RpoD holoenzyme-binding peaks were identified, of which 1,075 (40%) are located within 543 intergenic spacers (average 1.98 sites for each spacer) (Fig. 1A). On the other hand, a total of 1,626 (60%) peaks are located inside of 777 open reading frames (average 2.09 peaks per gene) (Fig. 1A). Since the majority of hitherto identified promoters are located within spacers, detailed search for the constitutive promoters was focused on the total of 1,075 peaks within 543 spacers.

**Location of the constitutive promoters within the E. coli genome**

The spacers containing RpoD holoenzyme-binding sites can be classified into three types (Fig. 1A): 404 peaks are located within 177 type-A spacer between bidirectional transcription units (Table 1 for the whole list; see also Table S1); 583 peaks are located within 315 type-B spacers located upstream of one transcription unit but downstream of another transcription unit (Table 2 and 3 for the whole list; see also Table S2); and 87 peaks are located within type-C spacers downstream of both transcription units of type-C spacers. Based on the transcription direction of flanking genes, the total number of constitutive promoters was predicted to range between minimum 492 (177 A-type plus 315 B-type) and maximum 669 (177x2 A-type plus 315 B-type).

Type-A spacers should contain at least two promoters for bidirectional transcription. The RpoD holoenzyme-binding sites
Distribution of RpoD holoenzyme-binding sites

A

| Spacers | Genes |
|---------|-------|
| 1,075 sites (40%) (543 spacers) | 1,626 sites (60%) (777 genes) |

Type-A 403
Type-B 585
Type-C 87

(Compared, 2,701)

Identification of the consensus sequence of constitutive promoters recognized by the RpoD holoenzyme

For identification of the constitutive promoters within type-A and type-B spacers with the binding sites of RpoD holoenzyme, we performed in vitro search using the consensus sequence of constitutive promoters. The current databases of E. coli promoters include both experimentally identified and computationally predicted promoters. The prediction of promoters has been performed using the canonical promoter, TTGACA-17 bp-TATAAT, which was originally identified using the in vitro transcription studies [10, 11, 38]. We then tried to identify the consensus sequence of constitutive promoters recognized in vitro by RpoD holoenzyme alone.

In order to experimentally confirm the consensus sequence recognized by RNA polymerase RpoD holoenzyme, we first constructed an ideal promoter with the complete set of consensus TTGACA (-35) and TATAAT (-10) sequences, separated by a spacer of 17 bp in length, starting from the lacUV5 promoter [23]. To identify the best RpoD promoter giving the highest activity and a two RNA polymerase subunits (RpoB and RpoC).

The total number of RpoD promoters (or the transcription initiation sites) listed in the current databases (RegulonDB and EcoCyc) are as many as 2,082, indicating that the majority of known promoters represent TF-dependent “inducible promoters” that are expressed only under the support of positive regulatory proteins (Fig. 1B). Among the total of 2,082 RpoD promoters (or transcription initiation sites) listed in the current version of promoter databases, the promoter sequences are described for a total of 582, of which 434 (75%) are located within the same spacers that contain the constitutive promoters (255 in type-A spacers and 179 in type-B spacers) (shown under black background of P columns in Tables 1, 2 and 3).
### Table 1. Constitutive Promoters (Type-A Spacers).

| Map Promoter sequence | Right Operon | Left Gene | D | RpoD | D | Right Gene | Left Operon | Promoter sequence |
|-----------------------|--------------|-----------|---|------|---|------------|-------------|-------------------|
| TTAACG-16-AAATAT | caiTABCDE | caiT | < | HNS | > | fixA | fixABCX | GTGACA-17-TAAAT |
| TTGATT-18-TGAAT | leuLABCD | leuL | < | > | leuO* | leuO | TTPGCA-17-TATTAT |
| TTGCT-18-CATAAA | coaED | coaE | < | > | guaC | guaC | TTTATA-18-GATTAT |
| ATGACG-18-TATAAT | gcd | gcd | < | HNS | > | hpt | hpt | TTAATA-18-TATAAG |
| TTTAAA-17-TATTAT | can | can | < | > | yadG | yadGH | |
| ATACA-17-GATTAT | yafV | yafV | < | > | ivy | ivy | TTTGAA-17-TACGT |
| TTGCTA-17-TACAT | fadE | fadE | < | > | lpcA | lpcA | CTGACC-17-TGATG |
| TTGTGC-17-TACAT | dinJ-yafQ | dinJ | < | HNS | > | yafL | yafL | TTTACA-17-TACGT |
| TTGAAC-17-TATTAT | phoE | phoE | < | HNS | > | prob | probA | TTGTTA-17-TAAAT |
| CTGCCA-16-TATTAT | insAB-2 | insA2 | < | > | inst-1 | inst-1 | |
| TCTACA-17-TATTAT | yagA-ykgN | yagA* | < | > | yagE | yagEF | TTTACC-16-CATAGT |
| TTGTA-16-AATATG | argF | argF | < | > | ykpS | ykpS | TTGAAT-18-TGTAAT |
| CTGCCA-16-TATTAT | yafL | yafL | < | > | yafJ | yafJ | |
| TCTACA-16-TATTAT | yafV | yafV | < | > | ivy | ivy | TTTGAA-17-TACGT |
| TTGCTA-17-TTACAT | betI | betI* | < | HNS | > | betT | betT | TGGACG-17-TTACAT |
| TTYAACG-17-TATAT | yaiS | yaiS | < | HNS | > | tauA | tauABCD | TTTGGA-16-TAAAT |
| TTYAACG-16-TATTAT | qmcA-yybJ | qmcA | < | > | ybbL | ybbLM | |
| TTGATT-17-TAACAC | ybcY | ybcY | < | HNS | > | tflX | tflX | ATGGCA-15-TTTAAA |
| TTGCT-17-TACCAT | ybdK | ybdK | < | > | soxE | soxE | ATGACG-16-TAAAT |
| TTGATT-18-TATTAT | fepD | fepD | < | > | entS | entS | AAGACA-17-AATAT |
| ATGACG-17-TATTAT | yapV | yapV | < | > | yahN | yahN | ATGACG-17-TTTAAA |
| TTYAACG-17-TATAT | yaiS | yaiS | < | HNS | > | tauA | tauABCD | TTTGGA-16-TAAAT |
| TTYAACG-16-TATTAT | qmcA-yybJ | qmcA | < | > | ybbL | ybbLM | |
| TTGATT-17-TTAACAC | ybcY | ybcY | < | HNS | > | tflX | tflX | ATGGCA-15-TTTAAA |
| TTGCT-17-TACCAT | ybdK | ybdK | < | > | soxE | soxE | ATGACG-16-TAAAT |
| TTGATT-18-TATTAT | fepD | fepD | < | > | entS | entS | AAGACA-17-AATAT |
| ATGACG-17-TATTAT | yapV | yapV | < | > | yahN | yahN | ATGACG-17-TTTAAA |
| TTYAACG-17-TATAT | yaiS | yaiS | < | HNS | > | tauA | tauABCD | TTTGGA-16-TAAAT |

Constitutive Promoters of RpoD Sigma Factor
Table 1. Cont.

| Map | Promoter sequence | Right Operon | Left Gene | D | RpoD | D | Right Gene | Left Operon | Promoter sequence |
|-----|-------------------|--------------|-----------|---|------|---|------------|-------------|--------------------|
| 26.62 | TAGATA-17-TCTAAT | ycgB | ycgB | < | > | dadA | dadAX | TTGACA-17-TATTCT |
| 26.79 | ycgR | ycgR | < | > | ymgE | ymgE | TTGACA-17-CATAAG |
| 26.92 | dhaKL | dhaK | < | > | dhaR* | dhaR | TTGACA-17-CATAAG |
| 27.10 | TGGTCA-17-TAAACT | pth | pth | < | > | ychH | ychH | TTGACA-17-CATAAC |
| 27.37 | TGGACA-16-TAAAAT | yabA | yabA | < | > | chaB | chaBC | TTGACA-18-TATTAT |
| 27.84 | ATGAAA-17-TATCAT | hns | hns | < | > | tdk | tdk | TTGACA-17-TATAT |
| 27.90 | TGGACA-16-TCTAAT | ynhE | ynhE | < | > | ynhE | ynhE | TTGACA-16-TCTAAT |
| 29.94 | ATGATA-19-TCTAAT | ychG | ychG | < | > | pprR* | pprR | TTGACA-18-TATTAT |
| 31.03 | TGGACA-17-TATAGT | ydhA | ydhA | < | > | ydhB | ydhB-ydhB | TTGACA-15-AAAAT |
| 31.90 | azOR | azOR | < | > | hpaA | hpaA | TTGACA-17-TATAT |
| 32.07 | TTAACA-19-TTTCAT | gapC | gapC | < | > | cybB | cybB | TTGACA-19-AAAAT |
| 32.16 | TTAATA-17-TCTAAT | ydhB | ydhB* | < | > | ydhC | ydhC | TTGACA-16-TATAT |
| 32.48 | TTAACA-18-TGTAAT | ynhC | ynhC | < | > | hicA | hicAB | TTAACA-19-TATAT |
| 33.50 | TGGACA-15-TATAAA | bdm-sra | bdm | < | > | osmC | osmC | TTGACA-17-TATAT |
| 34.73 | TTAATA-17-TATCCT | sad | sad | < | > | yneJ* | yneJ | TTGACA-17-AATAT |
| 34.84 | TGGACA-17-TGTAAT | ynhP | ynhP | < | > | marR* | marRAB | TTGACA-17-TATAT |
| 35.08 | TTAACA-17-TGTAAT | ynhC | ynhC | < | > | ynhC | ynhC | TTGACA-16-TATAA |
| 35.43 | relBE-hokD | relB | < | > | ydhV | ydhV | |
| 35.72 | TTGCCG-17-AATAAT | grxD | grxD | < | > | ydhO | ydhO | TTGACA-18-TAGAT |
| 36.41 | TTAATA-17-TCTAAT | ydhB | ydhB* | < | > | ydhC | ydhC | TTGACA-17-TATAT |
| 36.77 | TGGTCA-17-TATAAT | ynhC | ynhC | < | > | pykF | pykF | TTGACA-18-TATAT |
| 37.71 | TGGACA-17-TATCAT | thr5--infA | thr5 | < | > | arsB | arsB | ATGATA-16-TATAA |
| 39.13 | TGGACA-17-TGTAAT | ynhP | ynhP | < | > | ynhC | ynhC | TTGACA-17-TATAT |
| 40.39 | TTAACA-19-TGTAAT | ynhC | ynhC | < | > | ynhC | ynhC | TTGACA-18-TATAT |
| 40.48 | TTAATA-17-TCTAAT | znuA | znuA | < | > | znuC | znuC | ATGACA-17-TTAT |
| 40.90 | TGGTCA-15-TATAAA | yobI | yobI | < | > | yehB | yehB | CTGACA-17-TATAT |
| 41.96 | TTTCAA-17-TCTAAT | aspS | aspS | < | > | yehC | yehC | |
| 42.61 | TGGACA-17-TTAAAG | ydhA | ydhA | < | > | ydhK | ydhK | TTGACA-16-TATAA |
| 43.11 | CTAACA-19-AAAAAT | flcC | flcC | < | > | flcD | flcD | |
| 43.23 | TTAATA-18-TTCTAA | yedM | yedM | < | > | yedE | yedE | TTGACA-17-TGTAAT |
| 43.32 | TGGTCA-16-TAATT | yedM | yedM | < | > | intG | intG | TTGACA-16-TGAAT |
| 43.89 | TTAATA-16-TATAAC | yedWV | yedWV* | < | > | hluH | hluH | TGAAC-15-CAAAA |
| 44.99 | CTAACA-16-TAATT | yefM-yoeB | yefM* | < | > | hslL | hslL-FAF | TGAAC-19-TATAA |
| 45.28 | TGGTCA-17-TATAC | ygl | ygl | < | > | yegJ | yegJ | ATGATA-17-TATAT |
| 45.68 | TTAATA-18-TATATTT | yegR | yegR | < | > | yegS | yegS | TTGACA-17-TATAT |
| 46.18 | TCGTCA-18-TATCTT | ynhC | ynhC | < | > | ynhC | ynhC | TTGACA-17-TATAT |
| 46.80 | TTAATA-18-TATACAT | ducC | ducC | < | > | ynhK | ynhK | TTGACA-16-TGTAAT |
| 47.01 | ATGACA-15-AATAAT | ygaF | ygaF | < | > | yehC | yehC | TTGACA-17-TATAT |
| 47.42 | TTAATA-19-TATATTT | yehE | yehE* | < | > | yehH | yehH | TTGACA-15-TATAA |
| 49.19 | napF--ccmH | napF | < | > | yehO | yehO | TTGACA-16-TGTAAT |
| 50.02 | ATGACA-15-AATAAT | yefE-yoeB | yefE | < | > | yehC | yehC | TTGACA-17-TGTAAT |
| 50.41 | TGGTCA-16-TATACAT | ygaL | ygaL | < | > | ynhC | ynhC | TTGACA-17-TATAT |
| 51.81 | CTAACA-16-TATACAT | yNH | yNH | < | > | ynhC | ynhC | TTGACA-17-TATAT |
| 51.96 | TGAAC-15-TATAGT | yecB | yecB | < | > | yehC | yehC | TTGACA-15-TATAA |
| 52.11 | TGGTCA-17-TATAAT | yecB | yecB | < | > | yehC | yehC | TTGACA-17-TATAA |
| 53.46 | TGGTCA-17-TTTAAT | yecB | yecB | < | > | yehC | yehC | TTGACA-17-TTAT |

Constitutive Promoters of RpoD Sigma Factor
Table 1. Cont.

| Map   | Promoter sequence | Right Operon | Left Gene | D | RpoD | D | Right Operon | Left Operon | Promoter sequence |
|-------|-------------------|--------------|-----------|---|------|---|--------------|-------------|-------------------|
| 53.73 | TTGCCA-17-TACTAT  | yfdV         | yfdY      | < |  >   | lpxP| lpxP        |             | TCGACA-18-AAACAT |
| 54.02 | TTTACA-17-TAATTTT | glk          | glk       | < |  >   | yfeO | yfeO        |             | CTCACA-17-CATAAC |
| 54.09 | TTGATA-17-TAGCAT  | mntH         | mntH      | < |  >   | nupC | nupC        |             | ATGACT-17-TTAAAT |
| 54.40 | yfeR              | yfeR         |           | < |  >   | yfeH | yfeH        |             | TTTCA-17-TAGAAC  |
| 54.49 | TTAATA-16-GTGAAT  | zipA         | zipA      | < |  >   | sysZ | sysZK       |             | TTAAAT-19-TGAAAT |
| 54.96 | TTTGCT-17-CATAAT  | ypeA-ypeE    | ypeA      | < |  >   | amiA | amiA-heimF  |             | TTGTC-17-TAAAC  |
| 55.80 | TTCACC-17-TGAGAT  | yflM         | yflM      | < |  >   | yflB | yflB- yflM  |             | TTCA-16-TTAAAT  |
| 55.97 | dapa-bamC         | dapa         |           | < |  >   | gcvR | gcvR        |             | TTAABA-18-TCTGAT |
| 56.43 | TTTGCC-16-TATAAT  | upp-uraA     | upp       | < |  >   | purM | purMN       |             | ATGATA-17-TATTAT |
| 56.57 | ATGACA-18-TTAGAT  | yfgF         | yfgF      | < |  >   | yfgG | yfgG        |             | TTGACC-17-CITAGT |
| 56.70 | guaBA             | guaB         |           | < |  >   | xseA | xseA        |             | TTGCTG-16-TAGAGT |
| 57.34 | TCGGCA-18-XTAGAT  | trnJ         | trnJ      | < |  >   | suhB | suhB        |             | AAGACA-17-TTACT  |
| 57.68 | TTTAGT-16-TTAAAA  | glvB         | glvB      | < |  >   | hmp  | hmp         |             | TTTCA-16-TTAAAT |
| 58.35 | TAAACA-17-CATCAT  | rpoE-rseC    | rpoE      | < |  >   | nadB | nadB        |             | TAAAT-16-TTAAAT |
| 59.91 | TTGATA-16-T7ATGT  | aorF-yfaA    | aorF      | < |  >   | yflI | yflI        |             | TTTTCA-17-TTTTTA |
| 60.00 | ratAB             | ratA         |           | < |  >   | smpB | smpB        |             | TTTTCA-16-TTTTAT |
| 60.24 | TTAGATA-17-TTACAT | ileY         | ileY      | < |  >   | ygaQ | ygaQ-12     |             | TTCA-17-TGTAAT  |
| 60.27 | TTTACA-18-TTTAAAT | yqaE         | yqaE      | < |  >   | ygaV | ygaVF       |             | TTGACA-17-AACT  |
| 60.30 | TTTACA-17-TTTAAT  | stpA         | stpA      | < |  >   | alaE | alaE        |             | TTCA-17-TTACT  |
| 60.83 | TTGATAA-16-TTGGAT | mltB         | mltB      | < |  >   | srIa | srIa        |             | TTTAA-16-TTAAAT |
| 61.61 | TTGACA-18-T7ATGGT | ygbB*        | ygbB      | < |  >   | ygbJ | ygbJ        |             | TTTCA-16-TTAAAT |
| 62.45 | TTGAC-16-TGTGAT   | ygcW         | ygcW      | < |  >   | yqcE | yqcE-tycE   |             | TTCA-16-GATAAT  |
| 62.56 | TTTAAA-17-TGTGAT  | queE         | queE      | < |  >   | yqcG | yqcG        |             | TTCA-17-GATAAT  |
| 63.94 | TTGAGC-18-TTGTGAT | bppH-bpsP    | bppH      | < |  >   | mutH | mutH        |             | TCGGCA-18-TTAAAT |
| 65.80 | ygB- -visC        | ygB          |           | < |  >   | zapA | zapA        |             | TTGACT-17-TTACAT |
| 68.75 | TAGAGA-19-TTTTAGG | glgS         | glgS      | < |  >   | yqiJ | yqij        |             | TTCA-15-TTATT  |
| 69.27 | TTTACA-18- TTTAAT | ygiH         | ygiH      | < |  >   | yqiI | yqiI        |             | TTGACA-16-TTAAAT |
| 69.31 | TTGACT-19-TTACAT  | aer          | aer       | < |  >   | patA | patA        |             | TAAACA-19-TTAAAT |
| 69.68 | TTTTCA-16-TTACAT  | rlmG         | rlmG      | < |  >   | ygIP | ygIP        |             | TTTTC-17-TTACT  |
| 70.35 | TTGATA-16-TGTGAT  | trcA- -G     | trcA      | < |  >   | tdcR | tdcR        |             | TTAAT-16-TTAAAT |
| 71.48 | TTGACA-18-TTACAT  | metY- -pnp   | metY      | < |  >   | argG | argG        |             | ATGAA-17-AAAT  |
| 71.93 | TTTCA-18-TTAAAT   | mlaFEDCB     | mlaF      | < |  >   | yrbG | yrbG- -yrbA |             | TTTTCA-17-TTAAAT |
| 72.51 | TTACC-16-TTAAAT   | insH-10      | insH10    | < |  >   | yhCF | yhCF        |             | TTTAA-18-TTAAAT |
| 72.99 | ATTAA-18-TTATT    | aaeXAB       | aaeX      | < |  >   | aaeR | aaeR        |             | TTGATA-19-TGTAAT |
| 73.51 | TTTAAA-17-TTATTT  | envR         | envR      | < |  >   | acR  | acR         |             | TTGACT-19-AATAAT |
| 73.94 | CTGCTA-18-TTACAT  | smf          | smf       | < |  >   | def  | def-fnt    |             | TTGACT-18-GATAAT |
| 74.39 | TTTACA-17-TTAGAT  | gspA         | gspA      | < |  >   | gspC | gspD       |             | TTGTT-17-TTACT  |
| 76.72 | TTGACA-19-GTAGAT  | glpEGR       | glpE      | < |  >   | glpD | glpD        |             | TTGATAA-19-TTAAAT |
| 77.61 | TTGAC-18-TTGTGAT  | ftsYEX       | ftsY      | < |  >   | rsmD | rsmD- yhIhL |             | TTCA-18-TTACAT  |
| 78.27 | TTTACA-17-GATTAT  | yhiL         | yhiL      | < |  >   | yhiM | yhiM        |             | TTCGAA-16-TTAAAT |
| 78.33 | TTGACA-17-GTATA   | yhiN         | yhiN      | < |  >   | pitA | pitA        |             | TTCA-17- TATAAT |
| 78.58 | ATGAC-17-TTATAA   | dinQ         | dinQ      | < |  >   | arsR | arsRBC      |             | ATGAC-17-TTAAAT |
| 78.67 | TTGACA-18-TTATTT  | insH-11      | insH11    | < |  >   | slp  | slp-dctR    |             | TTGACT-17-TTAAT |
| 78.93 | TTGATA-16-TTAAAT  | gdoW         | gdoW      | < |  >   | gadY | gadY        |             | TTGAC-18-TTAAAT |
| 80.49 | TTCACA-17-TCAAAT  | bax          | bax       | < |  >   | malS | malS        |             | TTGATA-18-TCAAAT |
| 80.81 | TTGATA-19-TTACAT  | yiaT         | yiaT      | < |  >   | yiaU | yiaU        |             | TTCA-18-TTAAAT  |
### Table 1. Cont.

| Map | Promoter sequence | Right Operon | Left Gene | D | RpoD | D | Right Gene | Left Operon | Promoter sequence |
|-----|-------------------|--------------|-----------|---|------|---|------------|--------------|-------------------|
| 81.24 | TATAA-18-TATAAT | yibIH | yibI | < | > | mtaA | mtaADR | TATACA-19-TATAAT |
| 81.71 | TATAA-18-TATAAT | yibB | htrL | < | > | hldD | rfdA-waaCL | TATAA-17-TATAAA |
| 82.01 | TCTACA-17-TTTATT | waaQ-waaK | waaQ | < | > | waaA | waaA-coaD | CTGACA-16-TTTTAT |
| 82.28 | GTCACA-16-TAAAAG | ligB | ligB | < | > | gmk | gmk | GTCACA-16-AATAAG |
| 82.60 | TTGACA-19-TATAAT | yicJI | yicJ | < | > | selC | selC | TTATACA-18-TATAAA |
| 83.00 | TTGACA-17-GATACT | istR | istR | < | > | tisB | tisA | TTGACA-17-GATAAT |
| 84.59 | TTGACA-18-TATAAT | asnc-mnnmG | asnc* | < | > | asnA | asnA | TTGACA-16-TATAAT |
| 85.06 | TTGACA-19-TATAAT | rhlB | rhlB | < | > | trxA | trxA | TTGACA-16-AATAAA |
| 85.83 | TTGACA-18-TGATA | aslA | aslA | < | > | glmZ | glmZ | TTGACA-17-GATAAT |
| 85.95 | TTGACA-16-TTATT | yifB | yifB | < | > | ilvL | ilvL-XGM | TTGACA-16-TATAAT |

**Constitutive Promoters of RpoD Sigma Factor**

| Map | Promoter sequence | Right Operon | Left Gene | D | RpoD | D | Right Gene | Left Operon | Promoter sequence |
|-----|-------------------|--------------|-----------|---|------|---|------------|--------------|-------------------|

300 genes (a) | 178 (b) | K | HNS | 178 (a) | 291 genes (b) | 1.68 (a/b) | 26 TFs | 64 | 19 TFs | 1.63 (a/b)
and then a mixture of substrates and heparin was added to allow the single-round transcription. The final level of transcripts represents the amount of RNA synthesized in 15 min reaction after the addition of substrate mixture into open complexes formed during preincubation for various times up to 30 min (referred to parameter-I in this study). Parameter-I represents the binding affinity of RpoD holoenzyme to the test promoter. On the other hand, the slope of transcript increase represents the rate of open complex formation (referred to parameter-II).

Both parameter-I and -II were determined for each variant promoter for three times and the average values are shown in Fig. 4. In each panel, the promoter activity is compared between right-ward and left-ward transcription. The average numbers of genes under one promoter are 1.68 and 1.63 for left-ward and right-ward transcription, respectively, which correspond to 68 and 75%. Total number of genes under the control of 178 promoters were 300 for left-ward transcription, and 291 for right-ward transcription. The average numbers of genes under one promoter are 1.68 and 1.63 for left-ward and right-ward transcription, respectively. Among the total of 178 RpoD holoenzyme-binding sites, 64 (36%) overlap with the H-NS-binding sites.

Using a total of 669 sequences of the constitutive promoter with high-level of sequence conservation, the consensus sequence of RpoD promoter was reevaluated. The Logo plot, shown in Fig. 5A-2, indicates that: 1) the patterns of conserved bases of both -35 and -10 sequences are similar for both left and right genes with both type-A and type-B spacers (data not shown); and 2) a significant difference exists in the consensus sequence of both -35 and -10 boxes among the experimentally identified promoters [39].

The length of spacers between -35 and -10 signals ranges from minimum 14 to maximum 21 bp. Only 10–20% efficient promoters have been proposed to include either -35 or -10 box that resembles the consensus with five-out-of-six bases (5/6 agreement).

Unique features of the RpoD constitutive promoters

After sequence analysis of the entire genome of *E. coli* K-12, we realized that there is no ideal sequence of RpoD promoter with perfect matching to this consensus sequence. We then evaluated the RpoD promoters harbor unique sequences. By setting a rather severe screening condition of the sequence matching of more than 4 out of 6 bases for both -35 and -10 signals (total score, higher than 8; the highest score, 12) and with a spacer length of 17 plus/minus 2 (score 3 for 17 bp spacer, score 2 for 16 and 18 bp spacers, and score 1 for 15 and 19 bp spacers), a total of as many as 316 promoter sequences (89%) were identified among 354 predicted constitutive promoters within type-A spacer, and a total of 226 promoter sequences (82%) were identified among 276 predicted promoters excluding the internal promoters within type-B spacer. Overall the total amount of constitutive promoters with high-level (higher than 4/6) matching with the consensus RpoD promoter for both -36 and -10 signals is more than 85% (Fig. 5A-1). This is in sharp contrast with the collection of 582 experimentally defined promoters, in which the amount of promoters containing a high-level (4/6) agreement to both -35 or -10 signals are less than 40% (Fig. 5B-1). On average, only less than half of the 12 canonical bases of the -35 and -10 boxes are conserved among the experimentally identified promoters [39].

The length of spacers between -35 and -10 signals ranges from minimum 14 to maximum 21 bp. Only 10–20% efficient promoters have been proposed to include either -35 or -10 box that resembles the consensus with five-out-of-six bases (5/6 agreement).

Using a total of 669 sequences of the constitutive promoter with high-level of sequence conservation, the consensus sequence of RpoD promoter was reevaluated. The Logo plot, shown in Fig. 5A-2, indicates that: 1) the patterns of conserved bases of both -35 and -10 sequences are similar for both left and right genes with both type-A and type-B spacers (data not shown); and 2) a significant difference exists in the consensus sequence of both -35 and -10 boxes between a total of 543 newly identified constitutive promoters and a total of 582 experimentally identified promoters. The Logo pattern generated from the alignment of a total of 582 experimentally identified *E. coli* promoters, shown in Fig. 5B-2, indicates that the level of sequence conservation is significantly higher for -10 than -35, and the conserved bases are limited, TT (positions 1 and 2) of -35 signal and TA (positions 1 and 2) and T (position 6) of -10 signal. This Logo pattern agrees well with that analyzed by Mitchell et al. [40] using the promoter set of RegulonDB [41]. In the case of constitutive promoters, -35 and -10 sequences are equally well-conserved (Fig. 5A-2). Within -35 sequence, TT (position 1 and 2) and AA (positions 4 and 6) are well-conserved but G at position 3 and C at position 5 appear less important even for the constitutive promoters. On the other hand, all six bases of TATAAT within -10 signal are equally well-conserved, indicating their equal contribution to the promoter activity. The marked difference of the consensus sequence of experimentally identified RpoD promoters from that of constitutive promoters indicates that the promoter databases include a number of inducible promoters that were active under the support of positive TFs.
| Map | Promoter sequence | Left Operon | Left Gene | D | RpoD | D | Right Gene | Right Operon |
|-----|------------------|-------------|-----------|---|------|---|------------|-------------|
| 0.44 | CTGCCA-16-TATGAT | insAB-1     | insA-1    | < | <    | rpsT | rpsT       |             |
| 0.85 | caiABCDE         | caiA        | <         | < | caiT | caiABCDE |             |             |
| 3.96 | yaeH             | yaeH        | <         | < | yael | <       |             |             |
| 5.76 | TTGAAA-15-TATCAT | ykfA        | <         | < | ykfN | yagAB-ykgN |             |             |
| 5.99 | Internal Promoter | (insAB-afuBC) | afuB     | < | <    | ykgB | (yagAB-ykgN) |             |
| 6.02 | Internal Promoter | (yagAB-ykgN) | ykgN     | < | <    | ykgC | ykgC       |             |
| 6.28 | TTGACA-16-TATAAT | yagK        | HNS       | < | yagL | yagL      |             |             |
| 6.34 | TTTAAAA-17-TATCAT | yagN      | HNS       | < | intF | intF     |             |             |
| 6.72 | TTGAAA-17-TACTT | ykgMO      | HNS       | < | ykgR | ykgR     |             |             |
| 6.84 | ykgB             | ykgB       | HNS       | < | ykgC | ykgC     |             |             |
| 9.47 | Internal Promoter | (lacZYA) | lacY       | < | lacY | (lacZYA) |             |             |
| 9.63 | Internal Promoter | (frmRAB) | frmA       | < | frmR | frmRAB  |             |             |
| 9.77 | Internal Promoter | (xseB-ispA-dxs) | ispA | < | <    | xseB     | (xseB-ispA-dxs) |             |
| 10.27 | TCCACA-17-TACACT | ylaB       | ylaC      | < | <    | maa      | maa         |             |
| 10.30 | hha              | hha        | <         | < | tomB | tomB-hha |             |             |
| 10.33 | ATGAAA-17-TATGAT | tomB-hha   | HNS       | < | acrB | (acrAB)  |             |             |
| 14.19 | ATGAAA-17-TATGAT | lipA       | lipA      | < | <    | ybeF     | ybeF        |             |
| 14.22 | TTGACA-15-TATATT | ybeF       | HNS       | < | lipB  | lipB     |             |             |
| 15.30 | TTGTTAA-18-TACAAT | uof-fur   | uof       | < | <    | fda      | fda         |             |
| 15.31 | TGGGCA-18-AATAAG | fda        | fda       | < | ybfE  | ybfE     |             |             |
| 15.32 | TGGGCC-18-TATAT | ybfE       | ybfE      | < | <    | ybfF     | ybfF        |             |
| 16.20 | TAAACA-16-TAAAAA | ybdG       | HNS       | < | gltA  | gltA     |             |             |
| 16.94 | TTGTTAA-18-TACAAT | uof-fur   | uof       | < | <    | fda      | fda         |             |
| 17.03 | TGGTCC-17-TAAAAA | gallK      | gallE     | < | <    | modF     | modF        |             |
| 17.38 | TTGATG-19-TGAAAT | ybfE       | ybfE      | < | <    | ybfF     | ybfF        |             |
| 18.04 | TTGACA-17-AATAAT | focA-pflB  | focA      | < | <    | ycaO     | ycaO        |             |
| 19.38 | TTGACA-17-AATAAT | hspQ       | hspQ      | < | <    | rml     | rml         |             |
| 19.45 | TTGACA-17-AATAAT | hspQ       | hspQ      | < | <    | rml     | rml         |             |
| 20.54 | TTTAAT-19-AATAAT | ybfE       | ybfE      | < | <    | ybfF     | ybfF        |             |
| 22.15 | TTGACA-17-AATAAT | ybdG       | HNS       | < | gltA  | gltA     |             |             |
| 25.16 | TTGCTA-16-GATAAT | mfd        | mfd       | < | <    | ycfT     | ycfT        |             |
| 25.49 | Internal Promoter | (potABCD) | potB      | < | <    | potA     | (potABCD)   |             |
| 26.11 | TTCCATA-17-CATAAT | iraM       | HNS       | < | ycgX  | ycgX     |             |             |
| 26.13 | TTAAAA-16-AATAAT | bluR       | HNS       | < | bluF  | bluF     |             |             |
| 27.07 | Internal Promoter | (ychF)     | ychF      | < | <    | pth      | pth         |             |
| 28.17 | TTTACA-17-TAGAT | kch        | kch       | < | <    | yciI     | yciI        |             |
| 28.91 | TTGACA-17-TATTAT | osmB       | HNS       | < | yciT  | yciT     |             |             |
| 29.02 | TTGAGG-16-TATTTT | deoT       | deoT      | < | <    | yciZ     | yciZ        |             |
| 29.06 | TTGAAA-17-GAGAAT | fabl       | fabl      | < | <    | yciD     | yciD        |             |
| 29.08 | TTGAAA-16-CAGAAT | yciD       | yciD      | < | <    | sapF     | sapDF       |             |
| 29.17 | ATGACA-15-TTAAT | sapABCD    | sapA      | < | <    | ymiA     | ymiA        |             |
| 30.85 | TGTACA-16-AATAAT | pinR       | pinR      | < | <    | ynaE     | ynaE        |             |
| Map             | Promoter sequence  | Left Operon | Left Gene | D    | RpoD  | D    | Right Gene | Right Operon |
|-----------------|--------------------|-------------|-----------|------|-------|------|------------|--------------|
| 30.86           | CTGACA-17-TACCAT   | ynaE        | ynaE*     | <    | HNS   | <    | uspF       | uspF         |
| 31.54           | CTGACA-17-AATAAC   | ynbG        | ynbG      | <    |       | <    | insC-2     | insCD-2      |
| 33.21           | TTCAC-16-TCTAA     | narU        | narU      | <    | yddU  | <    | yddJ       | yddL         |
| 33.25           | Internal Promoter  | (yddLKJ)    | yddK      | <    | yddL  | <    | yddLKJ     |              |
| 33.42           | Internal Promoter  | yddM        | yddM      | <    |       | <    | adhP       | adhP         |
| 33.58           | Internal Promoter  | (dmpXABCFD) | dmpB      | <    |       | <    | dmpA       | (dmpXABCFD)  |
| 34.03           | TTGTCA-16-TATATA   | ydeN        | ydeN      | <    |       | <    | ydeo       | ydeo         |
| 34.10           | TTGAAG-16-TATATT   | ydeP        | ydeP      | <    | HNS   | <    | ydeo       | ydeo         |
| 34.27           | TTGACT-16-TAAAAC   | hipB        | hipB      | <    |       | <    | yneO       | yneO         |
| 34.28           | CTGACA-17-TTTAT    | yneO        | yneO      | <    | HNS   | <    | lsrK       | lsrK         |
| 34.68           | TTGCCG-19-TATCTT   | yneF        | yneF      | <    |       | <    | yneG       | yneG         |
| 35.31           | TTATATA-16-AATAAT  | essQ        | essQ      | <    | HNS   | <    | cspB       | cspB         |
| 35.58           | Internal Promoter  | (rpsAB)     | rpsB      | <    |       | <    | rpsA       | rpsAB        |
| 35.60           | TTGTC-17-TATAG     | rpsAB       | rpsA      | <    |       | <    | ynfA       | ynfA         |
| 36.28           | TTACG-17-AAAAT     | fumC        | fumC      | <    |       | <    | fumA       | fumA         |
| 36.52           | TTACG-17-TATAG     | uidR        | uidR*     | <    |       | <    | lsaH       | lsaH         |
| 37.12           | TTCAAA-15-TACATT   | sodC        | sodC      | <    |       | <    | ydhF       | ydhF         |
| 37.98           | ATCACA-16-GATAAT   | sufABCDSE   | sufA      | <    |       | <    | yndB       | yndB         |
| 38.00           | TTGTCA-16-CATATT   | ydiH        | ydiH      | <    |       | <    | ydiI       | ydiI         |
| 38.76           | TTATATA-16-AATAAT  | infC        | infC      | <    |       | <    | thrS       | thrS-ripT-pheST|
| 39.02           | ATGACT-16-AATAAT   | ydjO        | ydjO      | <    | HNS   | <    | cedA       | cedA         |
| 39.28           | ATGACT-16-AATAAT   | ydjO        | ydjO      | <    | HNS   | <    | cedA       | cedA         |
| 39.92           | TTGACA-17-GATAAT   | ydiF        | ydiF*     | <    |       | <    | ydjG       | ydjG         |
| 40.46           | TTGCC-19-TTTTAT    | yeaQ        | yeaQ      | <    |       | <    | yoaG       | yoaG         |
| 40.65           | TTATAA-15-TACATT   | fadD        | fadD      | <    |       | <    | lsaH       | lsaH         |
| 41.07           | TTGCC-15-GATAAC    | yobF-cspC   | yobF      | <    | HNS   | <    | yebO       | yebO         |
| 41.18           | Internal Promoter  | (fliAZY)    | fliA      | <    |       | <    | proQ       | proQ         |
| 41.53           | TTGTC-15-CACAAT    | yebE        | yebE      | <    |       | <    | yebF       | yebF         |
| 41.62           | TTAACG-17-TACATT   | edd-eda     | edd       | <    |       | <    | zwf        | zwf          |
| 41.89           | TGTAT-16-TATCAT    | ruvAB       | ruvA      | <    | HNS   | <    | yobl       | yobl         |
| 42.14           | GTGAC-18-TAAAAG    | torYZ       | torY      | <    | HNS   | <    | cutC       | cutC         |
| 42.90           | TTGAC-17-AATAAT    | sdlA        | sdlA*     | <    | HNS   | <    | yecC       | yecC         |
| 42.98           | TTGAA-17-AATAAT    | sdiA        | sdiA*     | <    | HNS   | <    | yecC       | yecC         |
| 43.07           | Internal Promoter  | (fliAZY)    | fliA      | <    |       | <    | proQ       | proQ         |
| 44.47           | TTGAC-17-TATATT    | cobUST      | cobU      | <    |       | <    | insH-6     | insH-6       |
| 44.49           | TTGACTA-17-TATATT  | yoeG        | yoeG      | <    |       | <    | yoeH       | yoeH         |
| 44.55           | TTGATAC-18-TATTAT  | yoeH        | yoeH      | <    | HNS   | <    | insD-3     | insD-3       |
| 44.77           | TTGACTA-15-TATATT  | yeeX        | yeeX      | <    |       | <    | yeaA       | yeaA         |
| 44.95           | yeeY               | yeeY        | <    |       | <    | yeeZ       | yeeZ         |
| 45.26           | ATTACCA-19-TATATT  | insH-7      | insH-7    | <    | HNS   | <    | wbbK       | (wbbBUK)     |
| 45.37           | glf-wbbH           | glf         | <    |       | <    | rfbX       | (rfbBDACX)   |
| 45.48           | TTCACA-18-TGGAAT   | rbDBDACX    | rbB       | <    |       | <    | galF       | galF         |
| 46.67           | ogrK               | ogrK        | <    |       | <    | yegZ       | yegZ         |
| 46.92           | Internal Promoter  | (gatZABCD)  | gatB      | <    |       | <    | gatA       | (gatZABCD)   |
| 47.00           | yeG                | yeG         | <    |       | <    | thlD       | thlD         |
| 47.63           | yehS               | yehS        | <    |       | <    | yehT       | yehT         |
| Map | Promoter sequence | Left Operon | Left Gene | D | RpoD | D | Right Gene | Right Operon |
|-----|------------------|-------------|-----------|---|------|---|------------|--------------|
| 47.89 | TTGACG-18-TATGAT | pbpG | pbpG | < | < | yohC | yohC |
| 49.05 | TTGATG-17-TATCAT | yejG | yejG | < | < | bcr | bcr |
| 49.64 | TTGATG-16-TGCAAT | mpo | mpo | < | HNS | yojl | yojl |
| 51.11 | GTGAAA-16-GATAAT | pmrD | pmrD | < | < | menE | menE |
| 51.61 | Internal Promoter | (nuoABC- -HIJKL) | nuoH | < | < | nuoG | (nuoAB- -GHJKL) |
| 51.73 | Internal Promoter | (nuoABC- -HIJKL) | nuoC | < | < | nuoB | (nuoAB- -GHJKL) |
| 52.25 | TTGAAA-16-TTTAAT | hisJQMP | hisJ | < | < | argT | argT |
| 52.87 | TTGAAA-17-TATGAT | yfcV | yfcV | < | HNS | sixA | sixA |
| 53.56 | CTGACA-19-CATTAT | yfdV | yfdV | < | < | oct | oct |
| 53.67 | TTTATA-18-TTCAAT | mazEFG | mazR | < | < | relA | relA |
| 54.26 | TTGACG-17-TATCAT | kduL | kduL | < | HNS | yqeF | yqeF |
| 54.40 | TTGACG-15-TTCAAAT | ygfF | ygfF | < | < | gcvP | gcvP |
| 56.18 | TCCCA-16-TGCTAT | epd-gpk-fbaA | epd | < | HNS | yggC | (yggDC) |
| 56.70 | TTGACG-15-TTCAAT | yghF | yghF | < | < | yghG | yghG |
| 57.77 | TTGCGT-17-GACAAT | yghW | yghW | < | < | yghX | yghX |
| 60.24 | TTGAAAA-17-TAATAAT | mqsR | mqsR | < | < | ygiV | ygiV |
| 60.40 | Internal Promoter | (nuoABC- -HIJKL) | nuoH | < | < | nuoG | (nuoAB- -GHJKL) |
| 61.81 | TTGACG-17-TTAAT | sibD | sibD | < | < | sibE | sibE |
| 74.76 | rimP-nusA-infB | rimP | rimP | < | < | metY | metY |
| 74.76 | folP-glmM | folP | folP | < | < | ftsH | ftsH |
| 77.24 | TTGCGC-18-TGATAA | arcB | arcB | < | < | yhcC | yhcC |
| 77.77 | TGGGCA-18-TAATAAT | ssPAB | ssP | < | < | rpsI | rpsI |
| 78.66 | TGGAAT-17-TATTAT | bfd-bfr | bfr | < | < | chiA | chiA |
| 78.47 | TGGGAA-16-TATTAT | yheO-tusDCB | yheO* | < | < | fkpA | fkpA |
| 79.16 | TTGCA-17-TATGAT | argD | argD | < | < | pabA | pabA |
| 79.67 | glpE | glpE | < | < | glpG | glpG |
| 79.70 | TTAGCA-17-TTATAT | gntR | gntR* | < | < | yhwW | yhwW |
| 79.75 | TGGGCA-19-TATGAT | rpoH | rpoH | < | < | ftsX | ftsX |
| 79.79 | TGGGAA-16-TATGAT | yhi-rbbA-yhhJ | yhi | < | < | yhiL | yhiL |
| 80.87 | CAGGCA-17-TATGAT | yiaWV | yiaW | < | < | aldB | aldB |
| 80.18 | TTAGCA-19-TATGAT | yilH | yilH | < | HNS | yilL | yilL |
| 80.76 | TGGAAT-17-TAAAG | (hdeABD) | hdeB | < | HNS | hdeA | hdeABD |
| 80.87 | TTAGCA-17-TTATTAT | gadX | gadX* | < | HNS | gadA | gadA |
| 79.17 | Internal Promoter | (dppABCD) | dppC | < | < | dppB | (dppABCD) |
| 80.87 | CAGGCA-17-TTATTAT | yiaWV | yiaW | < | < | aldB | aldB |
| 80.90 | TTAGCA-19-TTATTAT | ivbL- -uhpABC | ivbL | < | < | istR | istR |
Constitutive promoters carrying the binding site of silencer H-NS

The constitutive promoters direct transcription by the RpoD holoenzyme alone in the absence of positive transcription factors. Expression in vivo of the constitutive promoters must be repressed under conditions where the target genes are not needed. One possible transcription factor with the silencing activity is H-NS, one of the major core nucleoid proteins with functional dichotomy [42]. H-NS is known as a universal silencer for repression of a number of xenogeneic genes [43,44]. In parallel with the mapping of RpoD promoters, we identified a total of 987 H-NS-binding sites, 39 (24%) of which is shown in bold and italic while the promoter sequence with 5-out-of-6 match is shown in bold. The spacers including H-NS binding sites are marked as HNS in the spacer column. Total number of genes under the control of these 181 promoters were 290 (1.80 gene per promoter). Among the total of 181 RpoD holoenzyme-binding sites, 39 (24%) overlap with the H-NS-binding sites.

doi:10.1371/journal.pone.0090447.t002

Table 2. Cont.

| Map     | Promoter sequence | Left Operon | Left Gene | D | RpoD | D | Right Gene | Right Operon |
|---------|------------------|-------------|-----------|---|------|---|------------|--------------|
| 83.53   | gyrB             | gyrB        | HNS       | < | recF | recF |            |              |
| 84.18   | pstB-phoU        | pstB        | HNS       | < | pstA | (pstCA) |            |              |
| 85.30   | CTGACA-18-GATCAT | ppIC        | HNS       | < | yifG | yifG |            |              |
| 86.22   | yigF             | yigF        | HNS       | < | yigF | yigF |            |              |
| 87.36   | glnLG            | glnL        | HNS       | < | glnA | glnA |            |              |
| 88.73   | TTGCCA-18-TATACT | rraA        | HNS       | < | menA | menA-raA |            |              |
| 92.24   | TTCCCA-15-TAAACT | yjcF        | HNS       | < | actP | actP |            |              |
| 92.56   | yjCO             | yjCO        | HNS       | < | fdhF | fdhF |            |              |
| 93.18   | TTATCA-16-TATAAA | yjdN        | HNS       | < | yjdM | yjdM |            |              |
| 93.46   | TTATCA-17-GATACT | adIA        | HNS       | < | melR | melR |            |              |
| 93.90   | TTAGGT-19-TATAAT | cadBA       | HNS       | < | cadC | cadC |            |              |
| 94.05   | ATACA-17-TAAAAGA | dcaA        | HNS       | < | aspA | aspA-dcuA |            |              |
| 94.22   | TTAACC-17-TAGAGT | yjeJ        | HNS       | < | epmB | epmB |            |              |
| 95.44   | TTGTTCA-16-AAAAT | nrdD        | HNS       | < | treC | treC |            |              |
| 97.36   | TTAGGA-15-TTTAAC | yihu        | HNS       | < | yihu | yihu |            |              |
| 97.49   | TTCAAA-17-TTTAAC | yihuHG      | HNS       | < | yihu | yihu |            |              |
| 97.61   | TTACC-17-TATCAGC | sgxBCAER    | HNS       | < | yhQP | yhQP |            |              |
| 98.20   | iadA             | iadA        | HNS       | < | yijG | (yijGR-aDA) |            |              |
| 98.30   | CTGACC-19-TACAA | yijk        | HNS       | < | yijL | yijL |            |              |
| 98.63   | TTAACC-16-AAAAT | mcrBC       | HNS       | < | symE | symE |            |              |
| 99.60   | lplA             | lplA        | HNS       | < | ytB  | ytB  |            |              |
|         | 290 (a)           | 181 (b)     | HNS       |   |      |      |            |              |
|         | 1.80 (a/b)        | 15 TFs      | 39        |   |      |      |            |              |
|         | 69 Y-genes        | 24%         |           |   |      |      |            |              |
|         | 14 essential      |             |           |   |      |      |            |              |

Among the total of 1,075 RpoD holoenzyme-binding sites, 181 are located within type-B spacers upstream of left-side genes and downstream of right-side genes, indicating that these promoters direct leftward transcription. The genes and operons under the control of these 181 promoters were estimated, of which 16 represent putative internal promoters. Descriptions and symbols are as in Table 1. A total of 15 genes encoding transcription factors are indicated by star symbols (*). The essential genes listed in the PEC database are underlined within the operons. The promoter sequence with complete match with the canonical promoter (see Fig. 4) is shown in bold and italic while the promoter sequence with 5-out-of-6 match is shown in bold. The spacers including H-NS binding sites are marked as HNS in the spacer column. Total number of genes under the control of these 181 promoters were 290 (1.80 gene per promoter). Among the total of 181 RpoD holoenzyme-binding sites, 39 (24%) overlap with the H-NS-binding sites.

doi:10.1371/journal.pone.0090447.t002

Constitutive promoters carrying the binding site of silencer H-NS

The constitutive promoters direct transcription by the RpoD holoenzyme alone in the absence of positive transcription factors. Expression in vivo of the constitutive promoters must be repressed under conditions where the target genes are not needed. One possible transcription factor with the silencing activity is H-NS, one of the major core nucleoid proteins with functional dichotomy [42]. H-NS is known as a universal silencer for repression of a number of xenogeneic genes [43,44]. In parallel with the mapping of RpoD promoters, we identified a total of 987 H-NS-binding sites, 39 (24%) of which is shown in bold and italic while the promoter sequence with 5-out-of-6 match is shown in bold. The spacers including H-NS binding sites are marked as HNS in the spacer column. Total number of genes under the control of these 181 promoters were 290 (1.80 gene per promoter). Among the total of 181 RpoD holoenzyme-binding sites, 39 (24%) overlap with the H-NS-binding sites.

doi:10.1371/journal.pone.0090447.t002

Constitutive promoters carrying the binding site of silencer H-NS

The constitutive promoters direct transcription by the RpoD holoenzyme alone in the absence of positive transcription factors. Expression in vivo of the constitutive promoters must be repressed under conditions where the target genes are not needed. One possible transcription factor with the silencing activity is H-NS, one of the major core nucleoid proteins with functional dichotomy [42]. H-NS is known as a universal silencer for repression of a number of xenogeneic genes [43,44]. In parallel with the mapping of RpoD promoters, we identified a total of 987 H-NS-binding sites, 39 (24%) of which is shown in bold and italic while the promoter sequence with 5-out-of-6 match is shown in bold. The spacers including H-NS binding sites are marked as HNS in the spacer column. Total number of genes under the control of these 181 promoters were 290 (1.80 gene per promoter). Among the total of 181 RpoD holoenzyme-binding sites, 39 (24%) overlap with the H-NS-binding sites.

doi:10.1371/journal.pone.0090447.t002

Constitutive promoters carrying the binding site of silencer H-NS

The constitutive promoters direct transcription by the RpoD holoenzyme alone in the absence of positive transcription factors. Expression in vivo of the constitutive promoters must be repressed under conditions where the target genes are not needed. One possible transcription factor with the silencing activity is H-NS, one of the major core nucleoid proteins with functional dichotomy [42]. H-NS is known as a universal silencer for repression of a number of xenogeneic genes [43,44]. In parallel with the mapping of RpoD promoters, we identified a total of 987 H-NS-binding sites, 39 (24%) of which is shown in bold and italic while the promoter sequence with 5-out-of-6 match is shown in bold. The spacers including H-NS binding sites are marked as HNS in the spacer column. Total number of genes under the control of these 181 promoters were 290 (1.80 gene per promoter). Among the total of 181 RpoD holoenzyme-binding sites, 39 (24%) overlap with the H-NS-binding sites.
during the systematic search for transcription factors. For instance, the csgD (the master regulator of biofilm formation), the ndk gene (NADH:ubiquinone oxidoreductase), and the gadX gene (the stress-response regulator for acid resistance) are all under the control of more than 10 transcription factors including H-NS [4,22]. By using the newly developed PS-TF (promoter-specific transcription factor) screening system, we have identified the involvement of H-NS for regulation of a set of genes that are all under the control of multiple transcription factors [49]. These findings altogether suggest that the involvement of H-NS in transcription silencing of a number of constitutive promoters will be established once PS-TF screening is performed.

Discussion

Definition of the constitutive promoters

The pattern of genome transcription in E. coli is determined by controlling the utilization of a limited number of RNA polymerase [4,22]. High-throughput systems have been developed for identification of the RNA polymerase distribution within the genome. For instance, large-scale mapping of the functioning promoters in vivo on the E. coli genome has been performed by the chromatin immuno-precipitation and microarray (ChiP-chip) analysis with specific antibodies against RNA polymerase subunits [14,15]. To prevent moving RNA polymerase along DNA, E. coli cells were treated with rifampicin, which binds to the RpoB subunit of RNA polymerase and blocks transcription initiation, thereby fixing the initiated RNA polymerase on promoters [50]. For identification of the distribution pattern of functioning RNA polymerase, RNA-Seq analysis of high throughput sequencing of transcriptomes is becoming the method of choice [16-18]. Concomitant with the increase in the number of promoters detected under various stressful culture conditions, the variation of promoter sequence is expanding. The fluctuation of functioning promoters is attributable to the modulation of promoter selectivity of RNA polymerase by sigma factors and transcription factors with positive and negative regulatory functions [3,4].

The constitutive promoter has been considered to represent a set of promoters associated with the genes constitutively expressed in vivo in all circumstances. Based on this definition, however, it is practically impossible to identify the whole set of constitutive promoters of E. coli under various environmental conditions. At present, the whole sets of promoters determined by high-throughput put procedures (see above) are accumulating, but the experiments have been performed using different E. coli strains, under different culture conditions and using different experimental systems. In addition, even in laboratory culture conditions, it now turned clear that the steady state of cell growth does not exist and moreover, the genome expression pattern varies between individual cells within the same culture. In contrast, the whole set of constitutive promoters can be identified in vitro because in the case of E. coli, the faithful transcription can be established in vitro using purified RNA polymerase and pure DNA template under defined conditions. To avoid the complexity arisen from in vivo determination of the functioning promoters, an attempt was then made in this study to identify the whole set of the constitutive promoters using an in vitro system. Based on the results, we propose to revise the definition of the constitutive promoter as the promoter that is recognized and transcribed by the RNA polymerase holoenzyme alone in the absence of supporting TFs.

SELEX-chip search for the constitutive promoters

For this purpose, we used the improved method of Genetic SELEX screening [21], which has been successfully employed for identification of binding sites of a number of TFs on the E. coli genome [22]. As a result, the number of regulation targets markedly increased even for the TFs with known regulatory functions. For instance, the number of regulation targets increased more than 2 fold from 150 to 350 even for the best characterized transcription factor, CRP (cAMP receptor protein) or CAP (catabolite activator protein) [51]. This experimental system is particularly useful for short-cut estimation of the regulation targets of uncharacterized TFs including YhiK (renamed to RedA), YcdC (renamed to RutR), YgjZ (renamed to PgrR), YdhM (renamed to NemR) and YgjP (renamed to Dan) [22]. The Genomic SELEX screening system has also been successfully employed for detection of the alteration of promoter-recognition properties of transcription factors after phosphorylation (in the case of two-component systems) or interaction with effector ligands. For instance, the selection of regulation targets of SdaA, a regulator of genes for cell division and differentiation, was found to alter differently in the presence of each of three homoserine lactone analogs, the QS signals [52].

As an extension of screening of the whole set of RpoD promoters, we have successfully performed the screening of promoters recognized by RNA holoenzymes containing minor sigma factors, RpoS, RpoH, RpoF, RpoE and FecI. In the case of RpoN sigma factor, it requires enhancer-binding proteins such as NtrC and for formation of stable RNA polymerase-promoter complexes. Results will be described elsewhere.

Physiological roles of the constitutive promoters

The genes under the control of constitutive promoters are supposed to be expressed constitutively although unnecessary genes are subject to repression by silencers (see the H-NS chapter). One short-cut interpretation is that the essential genes are under the control of constitutive promoters. After systematic deletion of the genome, the total number of essential genes in the E. coli genome has been minimized to generate the minimal genome consisting of 302 essential genes ([53]; also see PEC database). Among 302 essential genes, 70 (23%) were identified to be under the control of constitutive promoters. Noteworthy is that the expression of some essential genes should be under the control of constitutively expressed positive transcription factors. In fact, we have identified the presence of approximately 100 species of E. coli transcription factor throughout cell growth.

Marked variation of the promoter sequences listed in databases must have been arisen from several different factors: 1) the list of RpoD-dependent promoters include those recognized by RpoS (and possibly other minor sigma factors), of which the promoter recognition properties overlaps with RpoD [54,55]; 2) the list also includes promoters that are functional in vivo only under the support of positive TFs [56-58]; and 3) a variety of promoter-like sequences have been identified in silico to be promoters. In fact approximately 60% of the RpoD holoenzyme-binding sites are located inside open reading frames (see Fig. 1A). Binding of RNA polymerase on some open reading frames has been recognized [59]. The unexpected high number of RpoD holoenzyme-binding sites inside open reading frames raises a possibility of an as yet unidentified functional role(s) for the RpoD holoenzyme. These promoter-like sequences may contribute transcription initiation from internal promoters and/or blocking the migration of elongation complexes. Clustering of promoter-like sequences within the E. coli genome were predicted by Collado-Vides and colleagues [60] while Ozoline and colleagues proposed the presence of as many as 78 ‘Promoter Islands’ [59,61]. These promoter-like sequences could form transcriptionally inactive
| Map   | Left Operon | Left Gene | D | RpoD | D | Right Gene | Right Operon | Promoter sequence |
|-------|-------------|-----------|---|------|---|------------|---------------|-------------------|
| 0.37  | sokC        | sokC      | > | HNS  | > | nhaA       | nhaAR         | TTAACC-17-TCTAAT  |
| 0.61  | dapB        | dapB      | > | carA | > | carAB      |               | TTTGACT-17-CAGAAT |
| 0.99  | yaaU        | yaaU      | > | kefF | > | kefC       |               | TTGAGA-17-TGGAAC |
| 2.27  | (mraZ- -lpXc) | ftsZ    | > | lpXc | > | lpXc       |               | TTTAAA-17-TAAAAT  |
| 2.76  | (pdrR- -lpid) | aceF    | > | lpd  | > | lpd        |               | TTGACG-17-TATTAT  |
| 3.55  | hrpB        | hrpB      | > | mrcB | > | mrcB       |               | TTGACG-15-TCTCAT  |
| 3.61  | mrcB        | mrcB      | > | fluA | > | fluACDB    |               | TTGACG-18-TATTAT  |
| 3.81  | clkA        | clkA      | > | erpA | > | erpA       |               | TAGATA-19-TAGAAT  |
| 4.33  | (bamA- -lpxA) | lpxD    | > | fabZ | > | fabZ       | (bamA- -lpxA)  | TGGCAA-15-TCTCAT  |
| 4.36  | (bamA- -lpxA) | fabZ    | > | lpXa | > | lpXa       | (bamA- -lpxA)  | Internal Promoter |
| 5.41  | dinB-yafNOP | dinB     | > | yafN | > | yafN       | yafNOP         | TTGACG-15-TTTAAC  |
| 5.62  | (proBC)     | proA      | > | thrW | > | thrW       |               | TGGACT-16-TATAAT  |
| 7.15  | betT        | betT      | > | yahA | > | yahA       |               | TGGACT-16-TATAAT  |
| 8.44  | yaiU        | yaiU      | > | yaiV*| > | yaiV       |               | TGCAC-18-TTTAAT  |
| 8.73  | yaiI        | yaiI      | > | yaiV | > | yaiV       |               | TTAATA-17-TGTAAT  |
| 9.33  | (nrdR- -pgpA) | ribD    | > | ribE | > | ribE       | (nrdR- -pgpA)  | TGGCAA-17-TGATG  |
| 9.79  | bloA        | bloA      | > | tig  | > | tig        |               | TGGACG-15-TATTAT  |
| 9.83  | tig         | tig       | > | clpP | > | clpP       |               | TGGAA-17-CATAAC  |
| 10.08 | cof         | cof       | > | ybaO*| > | ybaO       |               | TGGTGC-17-TAAAT  |
| 10.70 | adk         | adk       | > | hemH | > | hemH       |               | TTAATA-15-GATAAT  |
| 11.26 | (ybbAP)     | ybbP      | > | rhsD | > | rhsD       | ybbP           | TTAAAT-17-TGTAAT  |
| 11.36 | (rhsD- -yibH) | yibH    | > | ybbD | > | ybbD       | yibH           | TTAAAT-17-TGTAAT  |
| 11.38 | ybbD        | ybbD      | > | yblI | > | yblI       |               | TCTGCA-19-TAAAT  |
| 12.23 | (renD-emrE) | emrE      | > | ybcK | > | ybcK       |               | GTGACC-17-TAAAAA |
| 12.24 | ybcK        | ybcK      | > | ybcL | > | ybcL       |               | GTGGCA-17-TACAAT |
| 12.29 | ybcLM       | ybcLM     | > | ybcM | > | ybcM       |               | Internal Promoter |
| 12.55 | refereX     | refereX   | > | appY*| > | appY       |               | TTATCA-17-TTTAAT |
| 13.06 | SokE        | SokE      | > | hokE | > | hokE       |               | TGGACA-17-TGACT  |
| 13.56 | (entcEBAH)  | entH      | > | csaA | > | csaA       |               | TGGACA-15-AAAAT  |
| 14.15 | pagP        | pagP      | > | cspE | > | cspE       |               | TGGACA-17-TGACT  |
| 14.57 | ybeR-djiB   | ybeR      | > | djiB | > | djiB       | (ybeR-djiB)    | Internal Promoter |
| 15.84 | (ybiKO)     | ybiC      | > | ybiQ | > | ybiQ       |               | TTTTCA-17-AATACT |
| 16.61 | (mngAB)     | mngB      | > | cydA | > | cydA       |               | TCTACA-17-TATAAT |
| 17.49 | (biobFCD)   | bioD      | > | uvrB | > | uvrB       |               | TTGCA-17-TAAAT  |
| 17.64 | (mokaABCDE) | mokaE     | > | ybhL | > | ybhL       | mokaE          | TTGCAC-17-TACTC  |
| 17.65 | ybhL        | ybhL      | > | ybhM | > | ybhM       |               | ybhM            | ACGACA-16-TAAATA  |
| 19.21 | (ybjc- -ybjiN) | nsfA    | > | rimK | > | rimK       | ybjc- ybjiN    | Internal Promoter |
| 20.23 | (lola-rara) | rara      | > | serS | > | serS       | lola-rara      | TGGAC-17-GATAAG  |
| 20.68 | (sarc-aroA) | aroA      | > | ycaL | > | ycaL       |               | TGGTGA-17-ATTAT  |
| 23.37 | (efeOB)     | efeB      | > | phoH | > | phoH       |               | TTTATA-16-TATAAT |
| 24.80 | (rpmF- -fabHDG) | fabG    | > | acpP | > | acpP       | (rpmF- -fabHDG)| TGGACA-16-TACAAT |
| 24.84 | (acpF- -fabF) | fabF    | > | pabC | > | pabC       | (acpF- -fabF)  | TGGACA-15-GATAAG |
| 25.12 | (hinT- -ycfP) | ycfP    | > | ndh  | > | ndh        | (hinT- -ycfP)  | CTCACA-17-ACAAAT |
| 25.92 | (ymfTL- -ymfS) | ymfL    | > | ymfM | > | ymfM       | (ymfTL- -ymfS) | Internal Promoter |
| 26.07 | pinE        | pinE      | > | mcrA | > | mcrA       | pinE           | TTGTCG-17-ATTAT  |
| 26.19 | ycgZ- -ymgC | ycgZ      | > | ymgA | > | ymgA       | ycgZ- ymgC     | Internal Promoter |
| Map   | Left Operon | Left Gene | D | Right Gene | Right Operon | Promoter sequence |
|-------|-------------|-----------|---|------------|--------------|-------------------|
| 26.20 | (ycgZ- -ymgC) | arIR      | > | >         | ymgC         | Internal Promoter |
| 26.22 | (ycgZ- -ymgC) | ymgC      | > | >         | ycgG         | TGGACG-19- TATTTT |
| 26.27 | ymgF        | ymgF      | > | >         | ycgH         | TGGACA-19- TATAAG |
| 28.00 | ychE        | ychE      | > | >         | oppA         | TTAACA-17- AAGAAT |
| 28.51 | (yciVOQ)    | yciO      | > | >         | yciQ         | Internal Promoter |
| 28.71 | topA        | topA      | > | >         | ydbC         | TCAACA-15- TATAAA |
| 28.84 | pgpB        | pgpB      | > | >         | ysiS         | TTGATT-18- AAATC |
| 31.32 | (paaAB- -GHIJ) | paaB     | > | >         | paaC         | Internal Promoter |
| 31.44 | (paaAB- -GHIJ) | paaJ     | > | >         | paaK         | Internal Promoter |
| 31.63 | ins-2       | ins-2     | > | >         | ydbC         | TCAACA-17- TGAAT |
| 32.89 | yncH        | yncH      | > | >         | rhsE         | TTGACT-17- TATTAC |
| 34.53 | (lsrACD- -tam) | lsrD   | > | >         | lsrB         | Internal Promoter |
| 35.16 | ydfK        | ydfK      | > | >         | pinQ         | TGTACA-16- AATAT |
| 38.01 | dicF        | dicF      | > | >         | dicB         | CTAACA-17- TGAAT |
| 38.12 | rpra        | rpra      | > | >         | ydiL         | CTGATA-15- TATTGT |
| 38.73 | yniC        | yniC      | > | >         | ydiE         | TTGATA-16- TATCAA |
| 38.99 | ydiM        | ydiM      | > | >         | ydjM         | CTGAA-17- AATAT |
| 39.36 | nadE        | nadE      | > | >         | cho          | TGGTA-15- TTAAT |
| 41.46 | holE        | holE      | > | >         | yobB         | TATACA-16- CATAAC |
| 43.96 | (iedyVZ)    | yedZ      | > | >         | zinT         | TTTACA-18- AATAT |
| 44.44 | amn         | amn       | > | >         | yeeN         | TAGAC-18- TATAT |
| 44.60 | yeeP        | yeeP      | > | >         | isrC         | TGGTAC-17- TAGAT |
| 44.60 | isrC        | isrC      | > | >         | flu          | TGGTC-17- TACAT |
| 46.62 | (mdtA- -baeSR) | baeR    | > | >         | yegP         | CTGGC-17- CATACT |
| 48.45 | yehH        | yehH      | > | >         | nfo          | TATACA-16- TACAT |
| 48.47 | nfo         | nfo       | > | >         | yeil         | TATACA-16- TACAT |
| 49.23 | (yjgLM)     | yjgM      | > | >         | proL         | TGGAA-16- TAGAT |
| 49.61 | yjgO        | yjgO      | > | >         | eco          | GCGACA-15- TATAA |
| 51.29 | rbn         | rbn       | > | >         | elaD         | TTTAAA-18- TTTAT |
| 54.54 | cysZ        | cysZ      | > | >         | cysK         | ATGTC-16- TATAA |
| 55.21 | (yffOP)     | yffP      | > | >         | yffQ         | ATGTC-16- TATAA |
| 55.22 | (yffQR)     | yffR      | > | >         | yffS         | ATGTC-16- TATAA |
| 56.02 | bcp         | bcp       | > | >         | hycA         | TTTACA-17- CAGAT |
| 56.02 | hyfAB- -GHIJ | hyfA     | > | >         | hyfB         | Internal Promoter |
| 57.13 | sseA        | sseA      | > | >         | rfa          | TGGCA-16- TATTTT |
| 58.93 | bmd         | bmd       | > | >         | raiA         | CTGCA-18- TTTAT |
| 59.27 | nadK        | nadK      | > | >         | recN         | TTTACA-17- TATAA |
| 59.35 | smbB        | smbB      | > | >         | ssrA         | TTTACA-18- TATAA |
| 59.59 | (mLAB)      | mlB       | > | >         | yfpP         | TTTACA-15- TATAA |
| 59.79 | yfjW        | yfjW      | > | >         | yfjX         | TTTACA-15- TATAA |
| 59.82 | (yfjYZF)    | yfjF      | > | >         | psaA         | CTGCA-15- AACAAT |
| Map   | Left Operon | Left Gene | D | RpoD | D | Right Gene | Right Operon | Promoter sequence |
|-------|-------------|-----------|---|------|---|------------|--------------|------------------|
| 60.31 | ygaM        | ygaM      | > |      |   | nrdH       | nrdHIEF      | TCAACA-18- TATCAT |
| 60.46 | (proVWX)    | proW      | > |      |   | proX       | (proVWX)     | TATCA-16- AATAAC |
| 60.51 | (ygaXY)     | ygaY      | > |      |   | ygaZ       | ygaZHi       | TTAAGA-17- TATAAT |
| 60.92 | (srIAEBD)   | srID      | > |      |   | gutM*      | gutM-xrIRQ   |                  |
| 61.59 | mutS        | mutS      | > |      | HNS | pphB       | pphB         | TTAAGC-17- AAAAA |
| 64.43 | ygeF        | ygeF      | > |      | HNS | ygeG       | ygeG         | TTTAAA-17- TATCA |
| 64.49 | ygeI        | ygeI      | > |      | pbi | pbi         |              | TGACC-16- GATAC |
| 65.32 | yqfG        | yqfG      | > |      | idi | idi         |              | TTGTCG-18- AATCA |
| 65.82 | zapA        | zapA      | > |      | ssrS | ssrS-fau   |              |                  |
| 65.83 | (ssrS-fau)  | fau       | > |      | sibC | sibC        | TTTGACA-15- CCAAT |
| 66.49 | metK        | metK      | > |      | HNS | galP       | galP         | ATAAAA-18- TATAAC |
| 67.93 | metC        | metC      | > |      | yghB | yghB        | TGGACA-15- TATTGT |
| 68.61 | yqiC        | yqiC      | > |      | HNS | ygiL       | ygiL         | TCGATA-17- TATAAA |
| 69.10 | (ttdABT)    | ttdH      | > |      | ttdT | (ttdABT)   | Internal Promoter |
| 69.42 | ebgR        | ebgR      | > |      | ebgA | ebgAC      | TTGCCG-15- TATTTT |
| 69.55 | ygiJK       | ygiJ      | > |      | ygiJ | (ygiJK)    | Internal Promoter |
| 69.79 | aix         | aix       | > |      | sstT | sstT        | CTGACC-17- TGTCAT |
| 70.93 | exuR        | exuR      | > |      | yqiA | yqiA-msrA  | TTTTCA-16- TAAAAT |
| 70.39 | tdcR        | tdcR      | > |      | yhaB | yhaBC      | TTGATA-19- GAAAT |
| 70.79 | (agaS--agaBCD) | agal   | > |      | HNS | yraH       | yraHi        | TTGATA-17- TCAAT |
| 71.84 | ispB        | ispB      | > |      | sfsB* | sfsB       | TTGATA-18- TATAGT |
| 72.93 | argR        | argR      | > |      | HNS | yhcN       | yhcN         | TTGAAA-18- AATAAC |
| 73.46 | (panF-prmA) | prmA      | > |      | dusB | dusB-fis   | GTGCCA-18- AAAAT |
| 73.50 | yheST       | yheS      | > |      | yheT | (yheST)    | Internal Promoter |
| 73.35 | (nirBDC-cysG) | cysG       | > |      | HNS | yhlL       | yhlL         | TTAAGC-19- TATAAT |
| 73.39 | yhhL        | yhhL      | > |      | friA | friABCDR   | CTGACA-18- TTAAT |
| 76.26 | yhfG        | yhfG      | > |      | HNS | floA       | floABC       | TTGA1-15- TTAAT |
| 77.16 | yhhZ        | yhhZ      | > |      | insA-6 | insAB-6 | TTGAAA-17- TTAAT |
| 77.20 | (insAB-6)   | insB6     | > |      | HNS | yrhD       | yrhD         | TTGATA-18- TATATT |
| 78.71 | slp-dctR    | slp       | > |      | HNS | dctR       | dctR         | TTATAA-17- TATTAT |
| 80.65 | yiaK        | yiaK      | > |      | yial | (yiaK--sgbHUE) | Internal Promoter |
| 81.13 | yibA        | yibA      | > |      | yibJ | yibJ        |              |                  |
| 81.15 | yibJ        | yibJ      | > |      | yibG | yibG        |              |                  |
| 81.20 | yibV        | yibV      | > |      | yibU | yibU        |              |                  |
| 81.79 | (rfaD-waaFCL) | waaC       | > |      | waaL | (rfaD-waaFCL) | Internal Promoter |
| 82.24 | yicC        | yicC      | > |      | dinD | dinD        | GTGAGA-15- TATAAA |
| 82.34 | dinD        | dinD      | > |      | yicG | yicG        | TTATCA-16- AAAAT |
| 82.48 | xanP        | xanP      | > |      | yicH | yicH        |              |                  |
| 85.10 | ilvLXGMEDA  | ilvL      | > |      | ilvX | (ilvLXGMEDA) | Internal Promoter |
| 85.43 | trxA        | trxA      | > |      | rhoL | rhoL-rho   | TTGACT-17-TATAAA |
| 85.74 | (rfe- -tffT- -rffM) | rffM | > |      | yfK | yfK        | ATTACA-15- TTTAT |
| 87.25 | polA        | polA      | > |      | spf  | spf         | CTGTCG-17- TAGAAA |
| 87.47 | typA        | typA      | > |      | yihL | yihLM       |              |                  |
| 88.49 | fieF        | fieF      | > |      | pfkA | pfkA        |              |                  |
| 90.08 | (rplKJL-rpoBC) | rplL    | > |      | rpo8 | (rplKJL-rpoBC) | TATGCA-15- TGAAG |
| 91.98 | aphaA       | aphaA     | > |      | yibQ | yibQR       |              |                  |
The consensus sequence of constitutive promoters

Local concentrations of RNA polymerase on the genome. Complexes with RNA polymerase but might contribute to increase long spacers including the constitutive promoters often correspond with long UTR sequences but lacking protein-coding sequences of identified using 10 signal, each being separated by 17 bp linker was originally this -10 signal function, supporting the hypothesis that the finding is that all six bases of TATAAT are equally important for agreement with the previous proposals [63,64], but the novel

The most significant feature of constitutive promoters is the high-level of conservation of canonical TTGACA(-35)-17bp-TATAAT(-10) sequence. We also identified the roles and -35 and -10 signals depending inducible promoters. The promoter -10 TATAAT plays a key role in binding the RNA polymerase (see Fig. 4) but G at position 3 and C at position 5 are relatively less important for the constitutive promoters, suggesting that both play roles in TF-dependent inducible promoters. The promoter -10 TATAAT signal plays a major role in promoter opening (see Fig. 4) in agreement with the previous proposals [63,64], but the novel finding is that all six bases of TATAAT are equally important for this -10 signal function, supporting the hypothesis that the cooperativity of energy threshold, but not interaction of individual bases with RpoD, are important factors guiding the dynamics and selectivity of promoter open complex formation [63].

Regulation of the conserved promoter

The constitutive promoters must be repressed in cases when the genes under their control are not necessary. A group of silencing proteins play roles in preventing the potentially harmful effects of uncontrolled expression of the constitutive promoters. In E. coli, the H-NS family proteins are the major players of anti-silencing [44,45,66]. The high-level of overlapping was observed in the distribution between the constitutive promoters (see Tables 1, 2 and 3) and the binding sites of silencer H-NS [45]. A total of 203 (30%) of the constitutive promoters were predicted to be under the control of H-NS silencer, but silencing proteins for other 70% promoters remain unidentified. The spectrum of silencing targets by H-NS should be modulated by interaction with the members of Hha/YdgT family of small-sized co-regulators [67,68]. Possible involvement of other growth condition-specific nucleoid proteins in silencing such as Dps (DNA-binding protein in starved cells) [69] and Dan (DNA-binding protein under anaerobic conditions) [32] awaits further studies.

It should be noted that the constitutive promoters are also subject to activation by positive TFs for enhanced expression of the target genes. For instance, the csgD gene encoding the master regulator of biofilm formation carries one of the Type-A constitutive promoters, which is controlled by as many as 20 positive and negative TFs [22,33]. The level of constitutive csgD promoter is higher than that of csgB promoters directing transcription toward opposite orientation (see Fig. 2A[b]), but

Table 3. Cont.

| Map | Left Operon | Left Gene | D | RpoD | D | Right Gene | Right Operon | Promoter sequence |
|-----|-------------|-----------|---|------|---|------------|--------------|------------------|
| 92.52 | (nrfABCDEFG) | nrfG | > | HNS | > | gltP | gltP | ATGCCA-18-TATTAT |
| 94.29 | ecnB | ecnB | > | sugE | sugE | TTAGAA-16-CAAAAT |
| 94.89 | yjeT | yjeT | > | purA | purA | CGTAGA-19-TTTAAG |
| 95.01 | (nfrR-mnr- yjfl) | rnfB | > | yjfl | (nfrR-mnr- yjfl) | TATAAT-17-TGGAAAT |
| 95.35 | (rpsF-prbR- rplI) | rpsR | > | rplI | (rpsF-prbR- rplI) | Internal Promoter |
| 97.06 | (yjhc) | yjhc | > | ythA | ythA | TTGA-17-CATAAT |
| 97.85 | fimB | fimB | > | HNS | > | fimE | fimE | TTGCA-17-TATAGT |
| 98.12 | (uxuAB) | uxuB | > | uxuR* | uxuR | GTGCA-17-TATAA |
| 98.51 | yjiS | yjiS | > | HNS | > | yjiT | yjiT | TTGAGA-18-TATAAAA |
| 99.35 | osmY | osmY | > | yja | yjiA | Internal Promoter |
| 99.36 | yja | yja | > | yjiU | yjiU | |
| 99.52 | (deoCABD) | deoA | > | deoB | (deoCABD) | |
| 99.57 | (deoCABD) | deoD | > | HNS | > | yjiw* | yjiw | TTGGCA-18-TCTAT |
| 99.57 | yjiJ | yjiJ | > | HNS | > | yjID | yjID | TTGGCA-17-AATTAT |
| HNS | 154 (b) | 271 (a) |
| 38 | 9 TFs | 1.76 (a/b) |
| 25% | 17 essential | 62 Y-genes |

Among the total of 1,075 RpoD holoenzyme-binding sites, 154 are located within type-B spacers upstream of right-side genes and downstream of left-side genes, indicating that these promoters direct rightward transcription. The genes and operons under the control of these 154 promoters were estimated, of which 18 represent putative internal promoters. A total of 9 genes encoding transcription factors are indicated by star symbols (*). The essential genes listed in the PEC database are underlined within the operons. The promoter sequence with complete match with the canonical promoter (see Fig. 4) is shown in bold and italic while the promoter sequence with 5-out-of-6 match is shown in bold. The spacers including H-NS binding sites are marked as HNS in the spacer column. Total number of genes under the control of 154 promoters were 271 (1.76 gene per promoter). Among the total of 154 RpoD holoenzyme-binding sites, 38 (25%) overlap with the H-NS-binding sites.

doi:10.1371/journal.pone.0090447.t003
once the regulator CsgD is produced, it antagonizes H-NS silencer and the expression of the csgBAC operon is markedly enhanced, leading to production of curli fimbriae for biofilm formation.

In conclusion, we classified E. coli promoters into the constitutive promoters recognized by RNA polymerase holoenzyme alone and the transcription factor-assisted inducible promoters. Using the information of RpoD holoenzyme-binding sites identified by Genomic SELEX screening system, we predicted a total of 669 constitutive promoters with high-level conservation of the promoter consensus sequences. This finding indicates that the majority of hitherto identified promoters represent the TF-dependent inducible promoters.

Materials and Methods

Bacterial strains and plasmids

E. coli K12 W3350 type-A [70] was used for purification of RNA polymerase and the template DNA for Genomic SELEX screening of RpoD promoters. E. coli BL21(DE3) was used for the expression and purification of sigma and core enzyme subunit proteins. Expression plasmids for the core enzyme subunits (pRpoD, pRpoA, pRpoB and pRpoC) and all seven sigma subunits (pRpoD, pRpoS, pRpoN, pRpoH, pRpoF, pRpoE and pFec) were constructed by ligating the respective coding sequences, which were prepared by PCR amplification of the E. coli K12 W3350 type-A genome DNA as template, into pET21 expression vector essentially according to the standard procedure used for expression of all seven sigma subunits and all 300 transcription factors in this laboratory [71,72].

Purification of core RNA polymerase

RNA polymerase was purified from log-phase cells of E. coli K-12 W3350 by the standard procedure [26]. Separation of the core and holoenzymes was performed by passing the purified RNA polymerase through P11 phosphocellulose column in the presence of 50% glycerol [26]. To remove trace amounts of the core enzyme-associated sigma factors, the purified RNA polymerase in the storage buffer containing 50% glycerol was dialyzed against the same buffer but containing 5% glycerol and fractionated by phosphocellulose column chromatography in the presence of 5% glycerol [26]. The level of remaining sigma factors was less than 0.1%, if any, as checked by immuno-staining with antibodies against each of seven sigma factors.

Purification of core and sigma subunits

The core enzyme subunits (RpoA, RpoB, RpoC and RpoZ) were expressed using the respective expression plasmids and purified by two cycles of column chromatography through DEAE (DE52) and P11 phosphocellulose [26]. Sigma subunits were expressed and purified by ion-exchange column chromatography through DE52 and P11 followed by Sephacryl S-300 gel filtration column [26]. The purified sigma and core subunit proteins were
more than 99% pure as judged by protein staining of SDS-PAGE gels.

Preparation of antibodies

Antibodies against RpoD sigma and core enzyme subunits were produced in rabbits by injecting purified sigma proteins. The protocol for antibody production was raised following the Ethical Guidelines proposed by the Science Council of Japan and the Japanese Government, and approved by the Committee on the Ethics of Animal Experiments in the Animal Laboratory of Mitsubishi Chemical Medience Co. (Uto, Kumamoto, Japan). Antibodies against each RNA polymerase proteins were produced in two rabbits, and after examination of antibody activity using immune-blot analysis, the batch of higher activity was used in this study. Anti-RpoD, anti-RpoC used in this study did not cross-react with each other.

Genomic SELEX screening of RpoD holoenzyme-binding sequences

The genomic SELEX method was carried out as previously described [21]. A mixture of DNA fragments of the E. coli K-12 W3110 genome was prepared after sonication of purified genome DNA, and cloned into a multi-copy plasmid pBR322. In each SELEX screening, the DNA mixture was regenerated by PCR. For SELEX screening, 5 pmol of the mixture of DNA fragments and 10 pmol RNA polymerase RpoD holoenzyme were mixed in a binding buffer (10 mM Tris- HCl, pH 7.8 at 4°C, 3 mM magnesium acetate, 150 mM NaCl, and 1.25 mg/ml bovine serum albumin) and incubated for 30 min at 37°C. The DNA-RNA polymerase mixture was treated with anti-RpoC antibody and DNA fragments recovered from the complexes were PCR-amplified and subjected to next cycle of SELEX for enrichment of RNA polymerase-bound DNA fragments.

For SELEX-chip analysis, DNA samples were isolated from the DNA-protein complexes at the final state of SELEX, PCR-amplified and labeled with Cy5 while the original DNA library was labeled with Cy3. The fluorescent labeled DNA mixtures were hybridized to a DNA microarray consisting of 43,450 species of 60 b-long DNA probe, which are designed to cover the entire E. coli genome at 105 bp interval (Oxford Gene Technology, Oxford, UK) [27–29]. The fluorescent intensity of test sample at each probe was normalized with that of the corresponding peak of original library. After normalization of each pattern, the Cy5/Cy3 ratio was measured and plotted along the E. coli genome.

Mixed transcription assay in vitro

The promoter sequence of 195 bp-long lacUV5 template and its 5’ truncated 175 bp-long DNA was modified to the consensus sequence, TTGACA(-35)-17 bp-TATAAT(-10), each producing run-off transcripts of 42 and 22 nucleotides in length, respectively. Starting from these DNA fragments with the ideal promoter, the complete set of variant consensus promoters, each containing one base substitution at all positions of the hexanucleotide sequences of promoter -35 and -10, was prepared as described previously [23]. The test promoters of variant consensus promoters directed the synthesis of 42 b-long RNA while the reference template containing the original consensus promoter directed the synthesis of 22 b-long RNA.
of 22 b-long RNA. The in vitro mixed transcription was performed under the standard single-round reaction conditions in the presence of [32P]UTP as a labeled substrate [23–25]. In brief, a mixture of these two templates was preincubated with RpoD holoenzyme for various times (0.5, 1.0, 2.5, 5.0, 10, 15 and 30 min) at 37°C in the standard transcription assay buffer, and then a mixture of substrates and heparin was added to allow a single-round transcription for 15 min. Heparin inactivates RNA polymerase that was not involved in open complex formation. RNA products were separated by electrophoresis on 10% PAGE containing 8.3 M urea. The amount of RNA was determined by measuring the intensity of 32P radioactivity and corrected for the U content of each transcript.

Promoter sequence analysis

The complete genome sequence of E. coli K-12 MG1655 (U00096.2) was used for the promoter sequence analysis. The perl scripts for finding the candidate sequences of promoters include the following functions: 1) extraction of every 6 bases in the genome sequences with sliding 1 bp; 2) comparison of the 6 base sequences with TTGACA (-35) and calculation of the score by setting one match scoring 1 point, one mis-match scoring 0 point, and without gap; 3) comparison of the 6 bases with TATAAT (-10) under the same conditions; and 4) extraction of a pair of -35 and -10 hexanucleotide sequences by setting the best-match score of 12 (6 at -35 and at -10 signal); 6) extraction of a pair of -35 and -10 signals with a spacer length of 17 plus/minus 2, giving the score of +3 for 17 bp, +2 for 16 and 18 bp; and +1 for 15 and 19 bp, respectively. Thus, the maximum score for the best-match promoter in both -35 and -10 sequences separated by 17 bp spacer is 15.

The Logo pattern analysis of promoter sequences was performed using the established sequence logo generator [73].

Supporting Information

Table S1 RNA polymerase RpoD holoenzyme-binding sites on the E. coli genome. A total of 1,075 RpoD holoenzyme-binding sites were identified within spacers on the entire E. coli K-12 W3110 genome. The binding sites identified within intergenic spacers were classified into Type-A, Type-B and Type-C (see Fig. 2 and Table 1). The constitutive promoters were predicted based on the location of the RpoD holoenzyme-binding sites. [S1A] Promoters within type-A spacers. The genes under the constitutive promoters on left- and right-sides of type-A intergenic spacers and their functions are described. The genes shown under filled background encodes DNA-binding transcription factors (TFs), each regulating a set of target genes. The spacers under grey background contain two distinct peaks of holoenzyme binding. The positions of these genes on the genome are described on the left-end column while the levels of SELEX-peaks relative to the highest peak are described on the right-end column. [S1B] Promoters within type-B spacers. The genes under the constitutive promoters on either left- or right-sides of type-B intergenic spacers and their functions are described. The positions of these genes on
the genome are described on the left-end column while the levels of SELEX-peaks relative to the highest peak are described on the right-end column. The genes shown under black background encode DNA-binding TFs. The spacers under grey background contain two distinct peaks of holoenzyme binding.

**Table S2** Operons under the control of the constitutive promoters. Operons under the control of constitutive promoters are listed on the operon columns. Promoters shown by shaded background represent those listed in RegulonDB and EcoCyc databases. Spacers containing H-NS binding sites are indicated by HNS. [S2A] Operons under the control of constitutive promoters within type-A spacers. The constitutive promoters identified within type-A spacers direct bidirectional transcription of the operons indicated on left- or right-operon columns. The map positions of the promoter-proximal genes are indicated on the left- and right-end columns. [S2B] Operons under the control of constitutive promoters within type-B spacers. The constitutive promoters identified within type-B spacers direct transcription toward one direction.

**Author Contributions**
Conceived and designed the experiments: AI TS. Performed the experiments: TS AI. Analyzed the data: YY TS AI. Contributed reagents/materials/analysis tools: AI. Wrote the paper: AI. Supervised the project: AI KT.

**References**
1. Gruber TM, Gross CA (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. Annu Rev Microbiol 57: 441–466.
2. Paget MSB, Helmann JD (2003) The sigma-70 family of sigma factors. Genome Biol 4: 203.
3. Ishihama A (2000) Functional modulation of Escherichia coli RNA polymerase. Annu Rev Microbiol 54: 499–518.
4. Ishihama A (2010) Prokaryotic genome regulation: Multi-factor promoters, multi-target regulators and hierarchic networks. FEMS Microbial Rev 34: 628–645.
5. Gourse RL, Ross W, Rutherford ST (2006) General pathway for turning on promoters transcribed by RNA polymerase containing alternative sigma factors. J Bacteriol 188: 4309–4391.
6. Louette M, Grishkov M, Gross CA (1992) The sigma-70 family: sequence conservation and evolutionary relationships. J Bacteriol 174: 3843–3849.
7. Saecker RM, Record MT Jr, Dehaseth PL (2011) Mechanism of bacterial transcription initiation: RNA polymerase-promoter binding, isomerization to initiation competent open complex. J Mol Biol 412: 754–771.
8. Burgess RR, Anthony L (2001) How sigma docks to RNA polymerase and what sigma does. Curr Opin Microbiol 4: 126–131.
9. Young BA, Gruber TM, Gross CA (2004) Minimal machinery of RNA polymerase holoenzyme sufficient for promoter melting. Science 303: 1382–1384.
10. Pribnow D (1975) Bacteriophage T7 early promoters: nucleotide sequences of two RNA polymerase binding sites. J Mol Biol 15: 419–443.
11. Sierwin U, Simpson RB, Gilbert W (1980) E. coli RNA polymerase interacts homologously with two different promoters. Cell 20: 269–281.
12. Busby S, Ehrig RH (1994) Promoter structure, promoter recognition, and transcription activation in prokaryotes. Cell 79: 743–746.
13. Gralla JD (1996) Activation and repression of $E$. coli promoters. Curr Opin Genet Dev 6: 526–530.
14. Herring CD, Raffaelli M, Allen TE, Kanin EI, Landick R, et al. (2005) Immobilization of Escherichia coli RNA polymerase and location of binding sites by use of chromatin immunoprecipitation and microarrays. J Bacteriol 187: 6106–6147.

**Figure 5. The composition of constitutive promoters.** [A] The known consensus sequences of RpoD-dependent promoter, TTGAAC (-35) and TATAAT (-10) separated by 17 plus/minus 2 bp, were searched for all type-A (177) and type-B (315) spacers (see Experimental Procedure for the analysis method). Most of the constitutive promoters carry high-levels of the consensus sequence as listed in Table 1. The composition of promoter -35 and -10 sequence was classified into 8 groups based on the conservation level of consensus sequences. About 89% of type-A promoters and 82% of type-B promoters (or 86% of A- plus B-type promoters) contain the sequence higher than 4/6 agreement with the consensus sequence at both -35 or -10 positions (A1), while only 39% of a total of 582 known promoters carries this level of consensus sequences (B1). [B] The whole set of constitutive promoter sequences were subjected to Logo analysis [73]. The Logo patterns of -35 and -10 sequences are essentially the same among the constitutive promoters within type-A and type-B spacers. The Logo pattern of the whole set of constitutive promoters was compared with the Logo pattern generated using the total of 582 experimentally identified promoters [19,20]. The contribution of each base of the consensus -35 and -10 sequences is significantly different between the constitutive promoters and the set of known promoters.

doi:10.1371/journal.pone.0090447.g005
15. Mooney RA, Davis SE, Peters JM, Rowland JL, Ansari AZ, et al. (2009) Regulator trafficking on bacterial transcription units in vivo. Mol Cell 33: 97–101.
16. Cho BK, Zengler K, Qiu Y, Park YS, Knight EM, et al. (2009) The transcriptome unit architecture of the Escherichia coli genome. Nat Biotechnol 27: 1034–1048.
17. Raghavan R, Sage A, Ochman H (2011) Genome-wide identification of transcription start sites yields a novel thermostrasing RNA and new cyclic AMP receptor protein-regulated genes in Escherichia coli. J Bacteriol 193: 2071–2074.
18. Gomes dos Santos G, Cötz PD, Huang K, Haas BJ, Jard D, et al. (2012) Efficient and robust RNA-seq process for cultured bacteria and complex communities transcribed. Genome Biol 13: R23.
19. Salgado H, Perez-Gil M, Gama-Castro S, Santos-Zavaleta A, Muniz-Rascado L, et al. (2013) RegulatorDB v.10: omics data sets, evolutionary conservation and regulatory phrases, cross-validated gold standards and more. Nucleic Acids Res 41: D203–213.
20. Keshler DM, Collado-Vides J, Santos-Zavaleta A, Perez-Gil M, Gama-Castro S, et al. (2013) EcoCyc: a comprehensive database of Escherichia coli biology. Nucleic Acids Res 39: D583–D590.
21. Shimada T, Fujita N, Maeda M, Ishihama A (2005) Systematic search for the E. coli sigma 70 promoter using genomic SELEX. Genes Cells 10: 907–918.
22. Ishihama A (2012) Prokaryotic genome regulation: Multi-factor promoters and multi-target regulators. Proc Jpn Acad, Ser B, Phys Biol Sci 88: 455–508.
23. Koyahara M, Nagata K, Ishihama A (1999) Promoter selectivity of Escherichia coli RNA polymerase: Effect of base substitution in the promoter -35 region on promoter strength. Nucleic Acids Res 27: 7546–7572.
24. Kajitani M, Ishihama A (1983) Determination of the promoter strength in the mixed transcription system, II. Promoters of ribosomal RNA, ribosomal protein S1 and σ32 protein operons from Escherichia coli. Nucleic Acids Res 11: 3073–3082.
25. Kajitani M, Ishihama A (1983) Determination of the promoter strength in the mixed transcription system, I. Promoters of lactose, tryptophan and ribosomal protein L10 operons from Escherichia coli. Nucleic Acids Res 11: 3071–3073.
26. Fujita N, Ishihama A (1996) Reconstitution of RNA polymerase. Meth Enzymol 273: RNA Polymerase and Associated Factors, Part A, Adhya Sed., pp. 121–130, Academic Press, New York, 1996.
27. Ellington AD, Szostak JW (1990) In vitro selection of DNA molecules that bind specific ligands. Nature 346: 818–822.
28. Singer BS, Shatatland T, Brown D, Gold L (1997) Libraries for genomic SELEX. Nucleic Acids Res 25: 781–786.
29. Tjok, Gold L (1999) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249: 505–510.
30. Shimada T, Ishihama A, Bushby SJW, Grainger DC (2008) The Escherichia coli RpoK transcription factor binds at targets within genes as well as intragenic regions. Nucleic Acids Res 36: 3956–3954.
31. Shimada T, Hiro K, Kori A, Yamamoto K, Ishihama A (2007) RurK is the uracil/thymine-sensing master regulator of a set of genes for synthesis and degradation of pyrimidines. Mol Microbiol 66: 744–779.
32. Teramoto J, Yoshimura S, Takeyasu K, Ishihama A (2010) A novel nucleoid protein of Escherichia coli induced under anaerobic growth conditions. Nucleic Acids Res 38: 3605–3618.
33. Gottesman S, Storz G (2011) Bacterial small RNA regulators: versatile roles and complex transcription regulation in Escherichia coli. Microbiology 157: 379–395.
34. Ali SS, Whitney JC, Stevenson J, Robinson H, Howell PL (2013) Structural logo generator. Genome Res 14: 1188–1190.
35. Ramseier TM, Chien SY, Saier MH (1996) Cooperative interaction between Cra and Fnr in the regulation of the protein L10 operons from Escherichia coli. J Bacteriol 179: 2730–2742.
36. Barnt CB, Reynolds RP (1987) Analysis of E. coli promoter sequence. Nucleic Acids Res 15: 2397–2391.
37. Mooney RA, Davis SE, Peters JM, Rowland JL, Ansari AZ, et al. (2009) Regulator trafficking on bacterial transcription units in vivo. Mol Cell 33: 97–101.
38. Harley CB, Reynolds RP (1987) Analysis of Escherichia coli histone-like protein H-NS preferentially binds to horiontally acquired DNA of Salmonella. Science 231: 236–238.
39. Shimada T, Bridier A, Brierud R, Ishihama A (2011) Novel roles of LexO in transcription regulation in E. coli. Antagonistic interplay with the universal sigma H-NS. Mol Microbiol 82: 376–397.
40. Dame RT, Wyman C, Goosen N (2001) Structural basis for preferential binding of H-NS to curved DNA. Biochim Biophys Acta 153: 231–234.
41. Gold BR, Li Y, Cote A, Weirnsch MT, Ding P, et al. (2011) Structural basis for recognition of AT-rich DNA by unrelated xenogeneic silencing proteins. Proc Natl Acad Sci USA 26: 10690–10695.
42. Lim CJ, Lee SY, Kenney IJ, Yan J (2012) Nucleoprotein filament formation is the structural basis for bacterial H-NS gene sensing. Sci Rep 2: 509.
43. Shimada T, Ogawa-Hara H, Ishihama A, Kohara S, Shimura M, Kori A, et al. (2013) Screening of promoter-specific transcription factors: multiple regulators for the σ80 gene involved in cell division control and quorum sensing. Microbiology 159: 2501–2512.
44. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogos: a sequence logo generator. Genome Res 14: 1188–1190.