

# Whole set of constitutive promoters for RpoN sigma factor and the regulatory role of its enhancer protein NtrC in *Escherichia coli* K-12

Tomohiro Shimada<sup>1,\*</sup>, Shun Furuhashi<sup>1</sup> and Akira Ishihama<sup>2</sup>

## Abstract

The promoter selectivity of *Escherichia coli* RNA polymerase (RNAP) is determined by its promoter-recognition sigma subunit. The model prokaryote *E. coli* K-12 contains seven species of the sigma subunit, each recognizing a specific set of promoters. Using genomic SELEX (gSELEX) screening *in vitro*, we identified the whole set of 'constitutive' promoters recognized by the reconstituted RNAP holoenzyme alone, containing RpoD ( $\sigma^{70}$ ), RpoS ( $\sigma^{38}$ ), RpoH ( $\sigma^{32}$ ), RpoF ( $\sigma^{28}$ ) or RpoE ( $\sigma^{24}$ ), in the absence of other supporting regulatory factors. In contrast, RpoN sigma ( $\sigma^{54}$ ), involved in expression of nitrogen-related genes and also other cellular functions, requires an enhancer (or activator) protein, such as NtrC, for transcription initiation. In this study, a series of gSELEX screenings were performed to search for promoters recognized by the RpoN RNAP holoenzyme in the presence and absence of the major nitrogen response enhancer NtrC, the best-characterized enhancer. Based on the RpoN holoenzyme-binding sites, a total of 44 to 61 putative promoters were identified, which were recognized by the RpoN holoenzyme alone. In the presence of the enhancer NtrC, the recognition target increased to 61–81 promoters. Consensus sequences of promoters recognized by RpoN holoenzyme in the absence and presence of NtrC were determined. The promoter activity of a set of NtrC-dependent and -independent RpoN promoters was verified *in vivo* under nitrogen starvation, in the presence and absence of RpoN and/or NtrC. The promoter activity of some RpoN-recognized promoters increased in the absence of RpoN or NtrC, supporting the concept that the promoter-bound NtrC-enhanced RpoN holoenzyme functions as a repressor against RpoD holoenzyme. Based on our findings, we propose a model in which the RpoN holoenzyme fulfils the dual role of repressor and transcriptase for the same set of genes. We also propose that the promoter recognized by RpoN holoenzyme in the absence of enhancers is the 'repressive' promoter. The presence of high-level RpoN sigma in growing *E. coli* K-12 in rich medium may be related to the repression role of a set of genes needed for the utilization of ammonia as a nitrogen source in poor media. The list of newly identified regulatory targets of RpoN provides insight into *E. coli* survival under nitrogen-depleted conditions in nature.

## DATA SUMMARY

The genomic SELEX data for RpoN RNA polymerase (RNAP) holoenzyme (<https://shigen.nig.ac.jp/ecoli/tec/download/export/1/RpoN>), NtrC (<https://shigen.nig.ac.jp/ecoli/tec/download/export/2/NtrC>) and RpoN RNAP holoenzyme +NtrC (<https://shigen.nig.ac.jp/ecoli/tec/download/export/1/RpoNNtrC>) have been deposited in the Transcription Factor Profiling of *Escherichia coli* (TEC) database at the National Institute of Genetics, Japan (<https://shigen.nig.ac.jp/ecoli/tec/>).

## INTRODUCTION

The specificity of the promoter recognition of RNA polymerase (RNAP), responsible for environmental changes in bacteria, is modulated by replacement of the  $\sigma$  subunit, which controls differential gene expression [1, 2]. The sigma subunit  $\sigma^{54}$ , encoded by the *rpoN* gene, was first discovered during an analysis of glutamine synthetase and nitrogen assimilation in enteric bacteria [3]. Subsequent studies have confirmed its role in nitrogen assimilation but have also shown that it is involved in a variety of nitrogen-related, and seemingly unrelated, functions [4–6].

Received 13 April 2021; Accepted 15 July 2021; Published 17 November 2021

**Author affiliations:** <sup>1</sup>School of Agriculture, Meiji University, Kawasaki, Kanagawa, Japan; <sup>2</sup>Micro-Nanotechnology Research Center, Hosei University, Koganei, Tokyo, Japan.

**\*Correspondence:** Tomohiro Shimada, [tomoshimada@meiji.ac.jp](mailto:tomoshimada@meiji.ac.jp)

**Keywords:** *Escherichia coli*; gSELEX; nitrogen metabolism; NtrC; RNA polymerase; RpoN sigma factor.

**Abbreviations:** gSELEX, genomic SELEX; NTA, nitritotriacetic acid; RNAP, RNA polymerase; RT-qPCR, reverse-transcription quantitative real-time PCR; sRNA, small RNA; TCS, two-component system; TF, transcription factor.

**Data statement:** All supporting data, code and protocols have been provided within the article or through supplementary data files. Four supplementary tables and one supplementary figure are available with the online version of this article.

000653 © 2021 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

Ammonia is considered the preferred nitrogen source for *Escherichia coli* grown in minimal medium [7]. Under nitrogen depletion, one of the enhancers of NtrC encoded by *glmG* is activated by the phosphorylation of the kinase NtrB encoded by *glmL* [8]. These two proteins form the NtrBC two-component system (TCS). Phosphorylated NtrC binds to the nitrogen-regulated gene promoters, and in conjunction with the RpoN holoenzyme, transforms the promoter closed complex to an open complex [8–11]. NtrC directly or indirectly controls the majority of nitrogen-regulated genes [6, 12]. Based on this well-characterized NtrC-dependent transcription initiation system as a model, the RpoN RNAP holoenzyme is believed to be unique, with respect to the requirement of the enhancer (or activator), as follows: RNAP RpoN holoenzyme (RpoN-core enzyme complex) binds to the promoter to form an inactive closed complex, which is converted into the active open promoter complex after addition of hydrolysing ATP, with the help of a distinct class of transcriptional activators called enhancer-binding proteins [13, 14]. Moreover, the requirement of enhancers is unique to RpoN and not the other six sigma factors [15]. At present, 12 transcription factors (TFs) are known to be *E. coli* enhancers, belonging to two families: AtoC, NorR, NtrC, PrpR, PspF, QseF, RtcR and ZraR, which belong to the NtrC family, and DhaR, FhlA, HyfR and TyrR, which belong to the TyrR family [16].

To understand the regulatory role of RpoN *in vivo*, attempts have been made to identify RpoN-regulated promoters, such as with ChIP-seq analysis for detecting the RpoN-binding sites on the *E. coli* genome [17] and RNA-seq analysis to determine the mRNA levels after deletion of RpoN [18]. These genome-wide approaches have indicated the presence of novel RpoN targets. However, it is difficult to distinguish between the direct and indirect effects of the RpoN holoenzyme and/or the enhancer. For instance, the intracellular concentrations of RNAP sigma factor and TFs change depending on the growth phase and growth conditions [19, 20]. Therefore, it is difficult to identify the whole set of direct regulatory targets of NtrC *in vivo*, even though the genome-wide transcriptome [21] and genome-wide distribution [22] have been analysed. To avoid the problems associated with *in vivo* experiments, we performed genomic SELEX (gSELEX) screening of genomic DNA sequences recognized by the RNAP holoenzyme containing RpoN sigma (without other sigma factors) in the presence or absence of a single species NtrC enhancer. The original SELEX screening uses synthetic oligonucleotides with all possible sequences, and is able to identify the target DNA sequence. However, after a computer-based homology search for the consensus sequences, it is difficult to identify the whole set of target genes from the entire genome, because of difficulty in distinguishing positive target and false positive; whereas the gSELEX screening system uses genome fragments with all possible target sequences. The gSELEX screening system was developed to directly identify DNA sequences recognized *in vitro* by DNA-binding TFs [23, 24] and successfully applied to identify regulatory targets of more than 200 TFs from a single species: *E. coli* K-12 W3350 [16, 25]. Using this gSELEX method, we also identified the whole set of promoters recognized by a single species of sigma factor, including the major sigma subunit RpoD [26] and four species of the minor sigma subunits, RpoS, RpoH, RpoF and RpoE [27]. Promoter

### Impact Statement

The promoter selectivity of *Escherichia coli* RNA polymerase (RNAP) is determined by the promoter-recognition sigma subunit. RpoN sigma ( $\sigma^{54}$ ), involved in the expression of nitrogen-related genes and also other cellular functions, requires an enhancer, such as NtrC, for transcription initiation. Using the genomic-SELEX-chip method, we identified the number of target genes or operons as 44–61, 30–41 and 61–81 for RpoN RNAP holoenzyme, NtrC and RpoN RNAP holoenzyme +NtrC, respectively. Newly identified NtrC-dependent RpoN target genes include not only nitrogen-related genes but also nitrogen-related metabolism genes, such as those involved in carbon source metabolism. Based on our findings, we propose a dual function for the RNAP RpoN holoenzyme, as a repressor (in the absence of NtrC) and as a NtrC-activated transcriptase. This type of promoter, recognized by the RpoN holoenzyme alone, was termed as a repressive promoter.

search using the gSELEX system enabled the detection of the whole set of constitutive promoters recognized by each RNAP holoenzyme alone in the absence of other supporting factors, as well as in the absence of interfering proteins, including other sigma factors. Thus, the numbers of constitutive promoters for each sigma factor identified were as follows: 1320 for RpoD, 235 for RpoS, 331 for RpoH, 260 for RpoF and 493 for RpoE [26–28]. Based on the list of constitutive promoters, we could also predict the ‘inducible’ promoters recognized and activated in the presence of additional supporting factors. Under *in vivo* conditions, it is impossible to obtain the whole set of binding sites for both RNAP and TFs. In addition, the transcription-related data listed in the databases include different levels of accuracy. For instance, a number of TF-binding sites are estimated *in silico*, relying on consensus sequences that often include inaccurate predictions. Another significant problem originates from the use of various *E. coli* strains with different genetic backgrounds and the use of different culture conditions in each experiment.

In this study, we identified the whole set of RpoN-dependent promoters and the whole set of NtrC-binding sites using gSELEX screening. Furthermore, gSELEX analysis of the RpoN holoenzyme was performed in the presence of NtrC to identify the RpoN promoters regulated by the NtrC enhancer. The promoter activity of some RpoN promoters was examined using a gel-shift assay *in vitro* and reverse-transcription quantitative real-time PCR (RT-qPCR) assay *in vivo*. The promoter activity of some of the promoters recognized by the RpoN holoenzyme alone increased in the absence of RpoN. Furthermore, the binding of the RpoD holoenzyme to the test promoter was interfered with by the binding of the NtrC-enhanced RpoN holoenzyme to the promoter, suggesting a repressor function of the RpoN holoenzyme with competition against other RNAP holoenzymes. We designated this promoter as a repressive promoter, alongside the constitutive

promoter for RpoN-family sigma factors. The whole set of repressive promoters described herein provides fundamental catalogues for the promoters recognized by RpoN sigma factors and a useful resource for further analysis combined with other enhancers.

## METHODS

### Bacterial strains and plasmids

*E. coli* K-12 W3350 type-A, containing the full set of seven sigma factors [29], was used for the purification of RNAP and as a template DNA for the gSELEX screening of RpoN promoters and NtrC target genes. *E. coli* DH5 $\alpha$  was used for plasmid amplification. *E. coli* BL21(DE3) was used for the expression and purification of sigma N and sigma D, core enzyme subunit proteins, and NtrC. Expression plasmids for the core enzyme subunits and sigma N subunits (pRpoA, pRpoB, pRpoC, pRpoD and pRpoN) and NtrC (pNtrC) were constructed by ligating the corresponding coding sequences, which were prepared via PCR amplification of the *E. coli* K-12 W3350 type-A genomic DNA as a template, into the pET21 expression vector, according to a standard procedure used for the expression of sigma and TFs [30, 31]. *E. coli* BW25113 [32] and its single-gene knockout mutants, JW3169 for *rpoN* and JW3839 for *ntrC* [33], were obtained from the *E. coli* Stock Centre (National Bio-Resource Centre, Japan).

Cells were grown in LB medium or 3 mM NH<sub>4</sub>Cl (for nitrogen-starvation experiments) Gutnick minimal medium [34] (33.8 mM KH<sub>2</sub>PO<sub>4</sub>, 77.5 mM K<sub>2</sub>HPO<sub>4</sub>, 5.74 mM K<sub>2</sub>SO<sub>4</sub> and 0.41 mM MgSO<sub>4</sub> supplemented with Ho-LE trace elements and 0.2%, w/v, glucose), using NH<sub>4</sub>Cl as the sole nitrogen source, at 37 °C with constant shaking at 150 r.p.m. When necessary, 20  $\mu$ g kanamycin ml<sup>-1</sup> was added to the medium. Cell growth was monitored by measuring the turbidity at 600 nm.

### Purification of core RNAP

RNAP was purified from log-phase cells of *E. coli* K-12 W3350 using a standard procedure [35]. The native core was separated from the holoenzymes by passing the purified RNAP through a P11-phosphocellulose column in the presence of 50% (v/v) glycerol. To remove trace amounts of the core-enzyme-associated sigma factors, the purified RNAP in the storage buffer containing 50% (v/v) glycerol was dialysed against the same buffer containing 5% (v/v) glycerol and fractionated by P11-phosphocellulose column chromatography in the presence of 5% (v/v) glycerol. The level of remaining sigma factors was less than 0.1%, if any, as verified using SDS-PAGE gels by both protein staining with a silver reagent and immunostaining with antibodies against each of the seven sigma factors.

### Purification of core and sigma subunits

The core enzyme subunits (RpoA, RpoB, RpoC and RpoZ) were expressed using corresponding expression plasmids and purified by two cycles of column chromatography using DEAE (DE52) and P11-phosphocellulose [35]. The sigma subunits were expressed and purified via ion-exchange

column chromatography using DE52 and P11, followed by a Sephacryl S-300 gel filtration column. The purified sigma and core subunit proteins were over 99% pure, as determined by both protein staining and immunostaining of SDS-PAGE gels.

### Purification of antibodies

Antibodies against core enzyme subunits were produced in rabbits by injecting purified proteins [36, 37]. Antibodies against each RNAP protein were produced in two rabbits, and after the examination of antibody activity using immunoblot analysis, a batch with higher activity was used in this study. The anti-RpoD, anti-RpoS, anti-RpoN, anti-RpoH, anti-RpoF, anti-RpoE, anti-FecI and anti-RpoC used in this study did not cross-react with each other. These antibodies were produced by the Nippon Institute for Biological Science and the Animal Laboratory of Mitsubishi Chemical Medience.

### gSELEX screening of the binding sequences of RpoN RNAP holoenzyme and NtrC

gSELEX screening was performed using a standard procedure [23, 24]. A mixture of DNA fragments of the *E. coli* K-12 W3350 genome was prepared by sonicating purified genomic DNA and cloning it into multi-copy plasmid pBR322 at the *EcoRV* site. In each gSELEX screening, the DNA mixture was regenerated by PCR using a pair of primers with the flanking sequences of pBR322 *EcoRV*. For gSELEX screening, 5 pmol of the mixture of the DNA fragments and 10 pmol RpoN RNAP holoenzyme or NtrC 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 BSA ml<sup>-1</sup>) and incubated for 30 min at 37 °C. For reconstitution of the RpoN holoenzyme, the sigma-free core enzyme and fourfold molar excess of RpoN sigma subunit were mixed and incubated. For NtrC, acetylphosphate (0.1 mM) was added for NtrC auto-phosphorylation. The DNA-RpoN RNAP mixture was treated with anti-RpoC antibody and A/G beads, whereas the DNA-NtrC mixture was treated with a Ni-nitrilotriacetic acid (NTA) agarose column for purified DNA-test protein complexes. DNA fragments recovered from the complexes were PCR amplified and subjected to the next cycle of gSELEX to enrich the test protein-bound DNA fragments.

For gSELEX-chip analysis, DNA samples were isolated from the DNA-protein complexes at the final state of gSELEX, PCR-amplified and labelled with Cy5, while the original DNA library was labelled with Cy3. The fluorescently labelled DNA mixtures were hybridized to a DNA microarray consisting of 43 450 species of 60 bp long DNA probes, which were designed to cover the entire *E. coli* K-12 MG1655 genome at 105 bp intervals (Agilent). The fluorescence intensity of the test sample at each probe was normalized to that of the corresponding peak of the original library. After the normalization of each pattern, the Cy5/Cy3 ratio was measured and plotted along the *E. coli* K-12 MG1655 genome. The gene organization was almost identical between the two well-characterized *E. coli* K-12 strains, except for a long-range inversion between the *rrnD* and *rrnE* operons.



## Consensus-sequence analysis

To analyse the RpoN promoter motif and the NtrC binding sequence, each set of 500 bp binding sequences centred on the binding peak identified by gSELEX-chip was analysed using the program MEME SUITE [38]. The sequences were aligned, and a consensus-sequence logo was created using WEBLOGO (<http://weblogo.berkeley.edu/logo.cgi>).

## Gel-shift assay

The gel-shift assay was performed according to standard procedures [39]. Probes of the NtrC-dependent RpoN holoenzyme-binding target sequences were generated by PCR amplification using a pair of primers (Table S1a, available with the online version of this article) and *Ex Taq* DNA polymerase (TaKaRa). A mixture of each probe and RpoN holoenzyme, NtrC and RpoD holoenzyme was incubated at 37 °C for 30 min in the gel-shift buffer with 25 mM acetylphosphate and ATP. After the addition of the DNA loading solution, the mixture was directly subjected to 3.5% PAGE. DNA in gels was stained with GelRed (Biotium) and detected using LuminoGraph I (Atto).

## RT-qPCR analysis

RT-qPCR analysis was performed according to a standard procedure [40]. The nitrogen replete and deplete conditions were made according to Brown *et al.* with some modifications [22]. *E. coli* cells were inoculated in Gutnick minimal medium supplemented with 0.2% glucose and 3 mM NH<sub>4</sub>Cl at 37 °C under aeration with constant shaking at 150 r.p.m. until an OD<sub>600</sub> of 0.4 (nitrogen replete condition in exponential phase) or 0.9 (nitrogen deplete condition, 20 min after growth stopped) was reached. Then, the total RNA was extracted. Total RNA was transcribed to cDNA with random primers using the THUNDERBIRD SYBR qPCR RT kit (Toyobo). qPCR (quantitative real-time PCR) was conducted using THUNDERBIRD SYBR qPCR mix (Toyobo) and was performed using the LightCycler 96 system (Roche). The primer pairs used are listed in Table S1(b). The cDNA templates were serially diluted fourfold and used for qPCR assays. The qPCR mixtures, each containing 10 µl THUNDERBIRD SYBR qPCR mix, 1 µl each primer (50 µM stock), 7 µl water and 1 µl cDNA, were amplified under the following thermal cycling conditions: 95 °C treatment for 2 min; 45 cycles of 10 s at 95 °C and 20 s at 55 °C; and incubation for 20 s at 72 °C. The expression levels of 16S rRNA were used for the normalization of the levels of the test samples, and the relative expression levels were quantified using Relative Quantification software (Roche). Results are presented as the mean values of three independent experiments.

## Western blot analysis

Western blot analysis for PAGE gels was carried out by a standard method as described previously, with some modification [37]. After the gel-shift assay, gels were blotted onto PVDF membrane using a semi-dry transfer apparatus. Proteins on the membranes were immuno-detected with

anti-RpoN or anti-RpoD antibodies, and then detected with ImmunoStar Zeta (Fujifilm). Images were analysed with LuminoGraph I (Atto).

## RESULTS AND DISCUSSION

### gSELEX screening *in vitro*

Transcription by the RNAP RpoN holoenzyme is believed to depend on an enhancer (or activator), which promotes the transition from a closed promoter complex to an open complex for transcription initiation. This scenario was established using a single species of the enhancer NtrC [8–11]. To identify the whole set of RpoN-dependent promoters in the entire genome of *E. coli* K-12 W3110, and to identify the role of each enhancer, we performed a mass-screening *in vitro* of the whole set of sequences that are recognized by the reconstituted RNAP RpoN holoenzyme and the well-characterized NtrC as a model system. The sigma-free core enzyme was prepared by passing the purified RNAP (stored in a storage buffer containing 50% (v/v) glycerol) through a phosphocellulose chromatography column, three times in the presence of 5% (v/v) glycerol [35]. The level of remaining sigma subunits, if any, was less than 0.1%, as detected by both protein staining and immunostaining against each of the seven sigma factor species (RpoD, RpoN, RpoS, RpoH, RpoF, RpoE and FecI). The stoichiometry between core enzyme subunits was also checked by immunostaining with antibodies against each core subunit, RpoA, RpoB, RpoC and RpoZ. The holoenzymes fully saturated with each sigma subunit were reconstituted by mixing this sigma-free core enzyme and fourfold molar excess of purified RpoN sigma factor. As these sigma subunits alone are unable to bind to promoter DNA, the presence of excess sigma does not interfere with the function of RNAP holoenzymes. For the identification of DNA sequences that are recognized by each holoenzyme, we employed the gSELEX screening system [24], in which a library of *E. coli* genomic DNA fragments of 200–300 bp was used instead of synthetic oligonucleotides with all possible sequences used in the original SELEX method [41–43].

A multi-copy plasmid library of 200–300 bp random DNA fragments was constructed from the *E. coli* K-12 W3350 genome [23, 24]. The library used in this study contained 5.5-fold molar excess of the entire genome and, thus, a single and the same sequence may be included in five different overlapping segments on average, thereby increasing the resolution of SELEX fragment mapping. In each experiment of gSELEX screening, the mixture of genomic DNA fragments, which was regenerated by PCR from the genomic DNA library, was mixed with a twofold molar excess of the reconstituted RpoN holoenzyme and subjected to gSELEX screening. The DNA-holoenzyme complexes formed were recovered using the anti-RpoC antibody, which gave the highest level of RNAP recovery among all the anti-core subunit antibodies. RNAP-associated DNA was isolated from the antibody precipitates, amplified by PCR and subjected to cycles of gSELEX. After repeated gSELEX screening, the final products of RpoN holoenzyme-bound DNA fragments were mapped onto the

genome using a DNA tiling microarray (Oxford Gene Technology) [44]. The binding intensity was measured as the ratio of RpoN holoenzyme-bound DNA labelled with Cy3 against the original library DNA labelled with Cy5 on an array, and plotted along the *E. coli* genome for each holoenzyme. In the case of NtrC, His-tagged NtrC was purified and the binding reaction with the genomic DNA library was performed in the presence of acetylphosphate for NtrC activation by phosphorylation [45]. The NtrC-associated DNA was isolated using a Ni-NTA agarose affinity column and subjected to DNA tiling array analysis. During the DNA tiling array, 60 bp long probes were aligned along the *E. coli* genome at 105 bp intervals; therefore, approximately 300 bp long gSELEX fragments were bound to two or more consecutive probes. This criterion was employed to avoid background noise of the non-specific binding of holoenzyme-bound DNA fragments to the tiling array; it is worth noting that peaks showing hybridization to only a single probe were judged as false-positive noise.

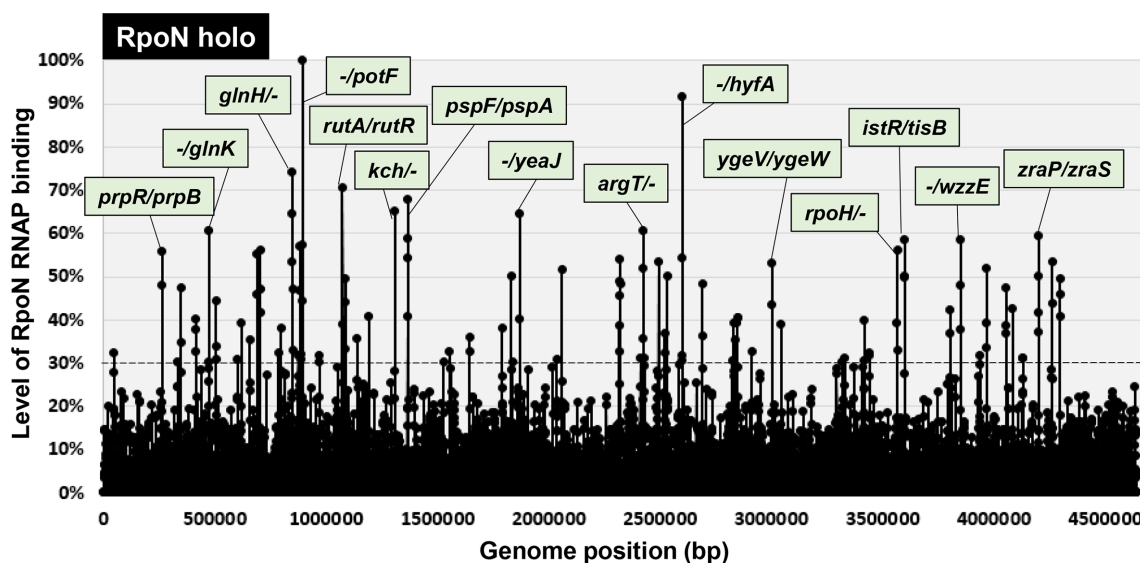
Binding sites were classified into two groups: one ‘within spacers’ and the other ‘inside genes’. Binding sites of the within spacers group were further classified into three types: a type-A spacer, located between bidirectional transcription units; a type-B spacer, located upstream of one transcription unit but downstream of another transcription unit; a type-C spacer, located downstream of both transcription units. Based on the transcription direction of flanking genes, the total number of constitutive promoters was predicted to range between the minimum (number of type-A spacers + number of type-B spacers) and maximum (number of type-A spacers  $\times$  2 + number of type-B spacers). The height of the

binding intensity identified by the gSELEX-chip system is generally in good agreement with the number of clones identified by the gSELEX-clos (cloning-sequencing) system, indicating that these two parameters correlate with the binding affinity of the test regulatory protein to DNA [24].

### Identification of the whole set of constitutive promoters recognized by the RpoN holoenzyme alone

To identify the whole set of RpoN recognition promoters, we performed gSELEX screening for the RpoN RNAP holoenzyme. After seven cycles of gSELEX screening, the sequences with a binding affinity to the RpoN holoenzyme formed a number of peaks along the entire *E. coli* genome (Fig. 1). By setting the cut-off level to 30% relative to the highest peak located upstream of *potF* (putrescine transporter), a total of 71 RpoN holoenzyme-binding peaks were identified, of which 44 (62%) were located within intergenic spacers (Fig. 1; detailed in Table S2), in addition to 17 peaks inside type-A spacers and 27 peaks inside type-B spacers (Table S2). From the RpoN holoenzyme-binding sites inside type-A and type-B spacers, a total of 44 (17 type-A + 27 type-B) to 61 (17  $\times$  2 type A + 27 type B) promoters were tentatively identified as constitutive promoters recognized by the RpoN holoenzyme (Table 1). A total of 27 peaks (38%) were located inside the ORFs (Fig. 1; detailed in Table S2).

Of the 71 binding targets of the RpoN holoenzyme, 23 sites (32%) were listed as RpoN targets in the RegulonDB database (RegulonDB column in Table S2). In contrast, 30 sites (42%)



**Fig. 1.** gSELEX-chip search for the binding sequences of the RpoN RNAP holoenzyme on the *E. coli* K-12 genome. gSELEX was performed to search for the binding sites of the RNAP RpoN holoenzyme. The y-axis represents the ratio against the highest peak at the *potF* promoter region and shows the level of RpoN holoenzyme-bound DNA fragments, whereas the x-axis represents the position on the *E. coli* K-12 genome in bp. The adjacent gene on the *E. coli* K-12 genome of the peak position was indicated for high intensity peaks (>60%). A list of binding sites of the RpoN holoenzyme is provided in Table 1 (detailed in Table S2).

**Table 1.** RpoN holoenzyme-binding sites on the *E. coli* K-12 genome

gSELEX was performed to search for binding sites of the RpoN RNAP holoenzyme. By setting the cut-off level to 30%, a total of 71 binding sites were identified (see Fig. 1 for gSELEX pattern), which have been aligned along the map of the *E. coli* K-12 genome. Binding intensity of the RNAP RpoN holoenzyme is shown in the RpoN holo column (see Table S2; the dark orange shading shows the intensity 61–100%, medium orange shading shows 41–60% and pale orange shading shows 30–40%). A total of 44 sites are located within intergenic spacers; 17 within type-A spacers and 27 within type-B spacers (see Table 2). Columns D indicate the direction of transcription. The potential target genes or operons of RpoN were predicted based on the adjacent genes and the gene orientation (shown with green shading). The grey shading shows genes that are not potential targets.

No.	gSELEX peak type	Map position (bp)	RpoN holo	Left gene function	Operon	Left gene	D	RpoN holo	D	Right gene	Operon	Right gene function
1	A	347864	47%	DNA-binding transcriptional activator	<i>prpR</i>	<i>prpR</i>	<		>	<i>prpB</i>	<i>prpBCDE</i>	2-Methylisocitrate lyase
2	B	471846	61%			<i>mdlB</i>	>		>	<i>glnK</i>	<i>glnK-amtB</i>	Nitrogen assimilation regulatory protein for GlnL, GlnE and AmtB
3	B	619432	39%	Iron-enterobactin transporter subunit	<i>fepC</i>	<i>fepC</i>	<		<	<i>fepG</i>		
4	A	655760	35%	Anaerobic C4-dicarboxylate transport	<i>dcuC</i>	<i>dcuC</i>	<		>	<i>pagP</i>	<i>pagP</i>	Palmitoyl transferase for lipid A
5	B	688560	55%	IS5 transposase and trans-activator	<i>insH</i>	<i>insH</i>	<		<	<i>lnt</i>		
6	A	784656	32%	Conserved protein	<i>ybgS</i>	<i>ybgS</i>	<		>	<i>aroG</i>	<i>aroG</i>	3-Deoxy-D-arabino-heptulosonate-7-phosphate synthase
7	B	847362	74%	Glutamine transporter subunit	<i>glnHPQ</i>	<i>glnH</i>	<		<	<i>dps</i>		
8	B	874568	32%			<i>yliE</i>	>		>	<i>yliF</i>	<i>yliF</i>	Predicted diguanylate cyclase
9	A	882830	57%	Undecaprenyl pyrophosphate phosphatase	<i>ybjG</i>	<i>ybjG</i>	<		>	<i>cmr</i>	<i>cmr</i>	Multidrug efflux system protein
10	B	891170	44%			<i>rfsA</i>	>		>	<i>rimK</i>	<i>rimK-ybjN</i>	Ribosomal protein S6 modification protein
11	B	892632	100%			<i>ybjN</i>	>		>	<i>potF</i>	<i>potFGHI</i>	Putrescine transporter subunit: periplasmic-binding component of ABC superfamily
12	A	1073268	71%	Predicted monooxygenase	<i>rutABCDEFG</i>	<i>rutA</i>	<		>	<i>rutR</i>	<i>rutR</i>	Predicted DNA-binding transcriptional regulator
13	B	1191232	41%	Adenylosuccinate lyase	<i>purB</i>	<i>purB</i>	<		<	<i>hflD</i>		
14	B	1308556	65%	Voltage-gated potassium channel	<i>kch</i>	<i>kch</i>	<		<	<i>yciI</i>		
15	A	1366070	68%	DNA-binding transcriptional activator	<i>pspF</i>	<i>pspF</i>	<		>	<i>pspA</i>	<i>pspABCDE</i>	Regulatory protein for phase-shock-protein operon
16	B	1527534	30%			<i>yncH</i>	>		>	<i>ycdD</i>	<i>ycdD</i>	Predicted protein

Continued

Table 1. Continued

No.	gSELEX peak type	Map position (bp)	RpoN holo	Left gene function	Operon	Left gene	D	RpoN holo	D	Right gene	Operon	Right gene function
17	A	1830436	50%	Succinylornithine transaminase, PLP-dependent	<i>astCADBE</i>	<i>astC</i>	<		<	<i>xthA</i>	<i>xthA</i>	Exonuclease III
18	A	2036832	31%	Predicted DNA-binding response regulator in TCS with YedV	<i>yedWV</i>	<i>yedW</i>	<		<	<i>hiuH</i>	<i>hiuH</i>	Hydroxysourate hydrolase/transthyretin-related protein
19	A	2060070	52%	DNA-binding transcriptional dual regulator of nitrogen assimilation	<i>nac</i>	<i>nac</i>	<		<	<i>asnV</i>	<i>asnV</i>	Asn tRNA
20	B	2321470	48%			<i>atoC</i>	>		>	<i>atoD</i>	<i>atoDAEB</i>	Acetyl-CoA:acetoacetyl-CoA transferase, alpha subunit
21	A	2411432	31%	Conserved inner membrane protein	<i>yfbV</i>	<i>yfbV</i>	<		<	<i>ackA</i>	<i>ackA-pta</i>	Acetate kinase A and propionate kinase 2
22	B	2425832	61%	Lysine/arginine/ornithine transporter subunit	<i>argT-hisIQMP</i>	<i>argT</i>	<		<	<i>ubiX</i>		
23	B	2429072	31%	Membrane protein required for colicin V production	<i>cvpA-purF-ubiX</i>	<i>cvpA</i>	<		<	<i>dedD</i>		
24	A	2493362	54%	Predicted inner membrane protein	<i>yfdY</i>	<i>yfdY</i>	<		<	<i>lpxP</i>	<i>lpxP</i>	Palmitoleoyl-acyl carrier protein-dependent acyltransferase
25	B	2520564	32%	DNA-binding transcriptional activator	<i>xapR</i>	<i>xapR</i>	<		<	<i>xapB</i>		
26	B	2531464	50%			<i>cysK</i>	>		>	<i>ptsH</i>	<i>ptsHI-ctr</i>	Phosphohistidinoprotein-hexose phosphotransferase component of PTS system
27	B	2599140	92%			<i>bcp</i>	>		>	<i>hyfA</i>	<i>hyfABCDEFGHIJR-jocB</i>	Hydrogenase 4, 4Fe-4S subunit
28	B	2689364	48%	ncRNA	<i>glmY</i>	<i>glmY</i>	<		<	<i>purL</i>		
29	B	2825748	31%			<i>srIB</i>	>		>	<i>srID</i>	<i>srID-gutM-srIR-gutQ</i>	Sorbitol-6-phosphate dehydrogenase
30	B	2836270	35%	Formate dehydrogenase-H, [4Fe-4S] ferredoxin subunit	<i>hydN-hypF</i>	<i>hydN</i>	<		<	<i>ascG</i>		
31	A	2848650	40%	Regulator of the transcriptional regulator FhIA	<i>hycABCDEFGHI</i>	<i>hycA</i>	<		<	<i>hypA</i>	<i>hypABCDE-fhIA</i>	Protein involved in nickel insertion into hydrogenases 3
32	A	3004270	53%	Predicted DNA-binding transcriptional regulator	<i>ygeV</i>	<i>ygeV</i>	<		<	<i>ygeW</i>	<i>ygeW</i>	Conserved protein
33	B	3043930	39%	Predicted NAD(P)-binding oxidoreductase with NAD(P)-binding Rossmann-fold domain	<i>ygfF</i>	<i>ygfF</i>	<		<	<i>gcvP</i>		

Continued

Table 1. Continued

No.	gSELEX peak type	Map position (bp)	RpoN holo	Left gene function	Operon	Left gene	D	RpoN holo	D	Right gene	Operon	Right gene function
34	B	3417032	40%			<i>yhdV</i>	>		>	<i>yhdX</i>	<i>yhdXYZ</i>	Predicted amino-acid transporter subunit
35	B	3440634	32%	30S ribosomal subunit protein S13	<i>rpsMKD-rpoA-rplQ</i>	<i>rpsM</i>	<		<	<i>rpmJ</i>		
36	B	3598870	59%	RNAP sigma 32 (sigma H) factor	<i>rpoH</i>	<i>rpoH</i>	<		<	<i>fisX</i>		
37	A	3851352	59%	ncRNA	<i>istR</i>	<i>istR</i>	<		<	<i>fisB</i>	<i>fisB</i>	LexA-regulated toxic peptide
38	B	3967058	52%			<i>rfe</i>	>		>	<i>wzzE</i>	<i>wzzE-wecBC-rffGHCA-wzxE-rffT-wzyE-rffM</i>	Enterobacterial common antigen polysaccharide chain length modulation protein
39	A	4056244	47%	Glutamine synthetase	<i>glnALG</i>	<i>glnA</i>	<		<	<i>typA</i>	<i>typA</i>	GTP-binding protein
40	A	4083972	43%	Formate dehydrogenase-O <sub>2</sub> large subunit	<i>fdoGHI-fdhE</i>	<i>fdoG</i>	<		<	<i>fdhD</i>	<i>fdhD</i>	Formate dehydrogenase formation protein
41	B	4131538	31%			<i>metF</i>	>		>	<i>katG</i>	<i>katG</i>	Catalase/hydroperoxidase HPI(1)
42	A	4199860	60%	Zn-binding periplasmic protein	<i>zraP</i>	<i>zraP</i>	<		<	<i>zraS</i>	<i>zraSR</i>	Sensory histidine kinase in two-component regulatory system with ZraR
43	B	4260864	54%			<i>dusA</i>	>		>	<i>pspG</i>	<i>pspG</i>	Phage shock protein G
44	B	4297530	50%	Formate dehydrogenase-H <sub>2</sub> selenopolypeptide subunit	<i>fdhF</i>	<i>fdhF</i>	<		<	<i>mdtP</i>		
A=17, B=27		Cut-off	>30%									
		Spacer	44									

Constitutive promoters=44-61



were detected by ChIP-chip analysis (ChIP-chip column in Table S2) [17]. A total of 39 (55%) were newly identified as RpoN targets in the *E. coli* genome in this study (see below).

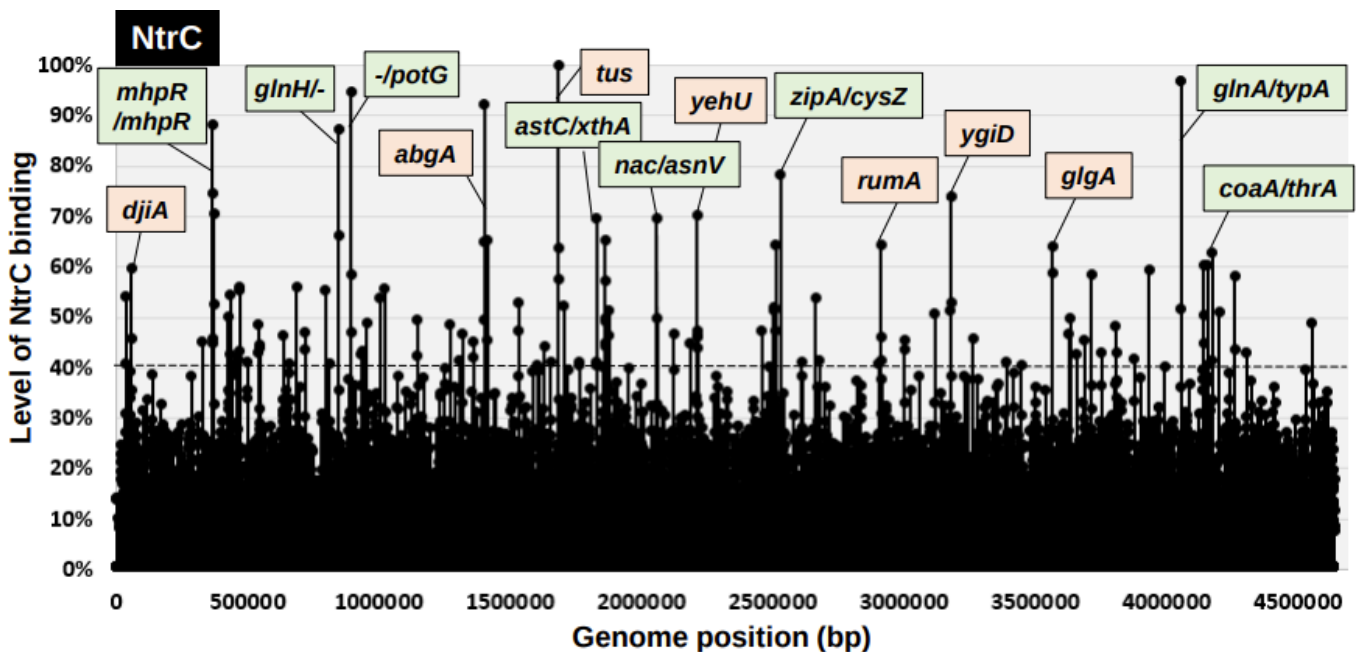
### Identification of the whole set of NtrC target genes

NtrC was isolated as the nitrogen assimilation regulator encoded by *glnG* [46], and is known as an enhancer for nitrogen assimilation under nitrogen-limited conditions [47]. To understand genome regulation by NtrC TCS in *E. coli*, we attempted to identify the whole set of regulatory target promoters, genes and operons under the control of phosphorylated NtrC. For this purpose, we independently employed gSELEX screening using purified His-tagged NtrC in the presence of 10 mM acetylphosphate for NtrC phosphorylation *in vitro*. From a mixture of *E. coli* K-12 W3110 genome fragments, NtrC-bound DNA fragments were affinity-purified using Ni-NTA agarose and then subjected to tiling array analysis to identify NtrC recognition sequences. This gSELEX screening was repeated for up to five cycles. The original mixture of genomic DNA fragments formed smeared bands during PAGE. However, after repeated gSELEX screening, the NtrC-bound DNA formed sharper bands during PAGE, indicating the enrichment of specific DNA fragments with specific binding activity to NtrC. Here, we identified a total of 93 high-intensity peaks by setting the cut-off level above an intensity of 40%, relative to the highest peak located on the *tus* ORF (Fig. 2; for details, see Table S3). Of these 93 high-level binding peaks, seven binding sites are

listed as NtrC target genes or operons in the RegulonDB database (RegulonDB column in Table S3). A total of 31 (33 %) NtrC-binding sites were found within the spacers, while 62 (67 %) were found inside the ORFs (Table S3). Of these 31 NtrC-binding sites within spacers, 11 were located within spacers of bidirectional transcription units (Table 2), 19 were located inside spacers upstream of one ORF but downstream of another ORF (Table 2), and two were located inside the type-C spacer (Table 2). Using these results, we predicted that the total number of regulatory targets of NtrC was between 30 (11 type A+19 type B) and 41 (11×2 type A+19×1 type B). We performed gSELEX screening for approximately 200 species of *E. coli* K-12 TFs. Although the binding of TFs inside ORFs is variable between TF species, the level of 67 % binding of NtrC inside ORFs was rather high, and its unidentified regulatory roles should be analysed in detail.

### Identification of the whole set of promoters recognized by NtrC-controlled RpoN holoenzyme

For transcription initiation by the RpoN holoenzyme, one of the NtrC- or TyrR-family TFs, such as NtrC, is believed to be necessary [8–11]. To understand the intrinsic role of the RpoN sigma factor, we performed gSELEX screening of the regulatory target promoters of the RpoN holoenzyme in the presence of an excess amount of NtrC under the same conditions used for the screening of constitutive promoters by the RpoN holoenzyme alone. After seven cycles of gSELEX screening, RpoN holoenzyme-bound DNA segments were



**Fig. 2.** gSELEX-chip search for the binding sequences of NtrC on the *E. coli* K-12 genome. gSELEX was performed to search for the binding sites of NtrC, in the presence of acetylphosphate, with respect to NtrC phosphorylation. The y-axis represents the ratio against the highest peak at the *tus* ORF and shows the level of NtrC-bound DNA fragments, whereas the x-axis represents the position on the *E. coli* K-12 genome in bp. The adjacent gene on the *E. coli* K-12 genome of the peak position was indicated for high intensity peaks (>70%). Peaks located within the spacer regions are shown with green labels, while peaks located within ORFs are shown with orange labels. A list of the binding sites of NtrC is provided in Table 2 (detailed in Table S3).

**Table 2.** NtrC-binding sites on the *E. coli* K-12 genome

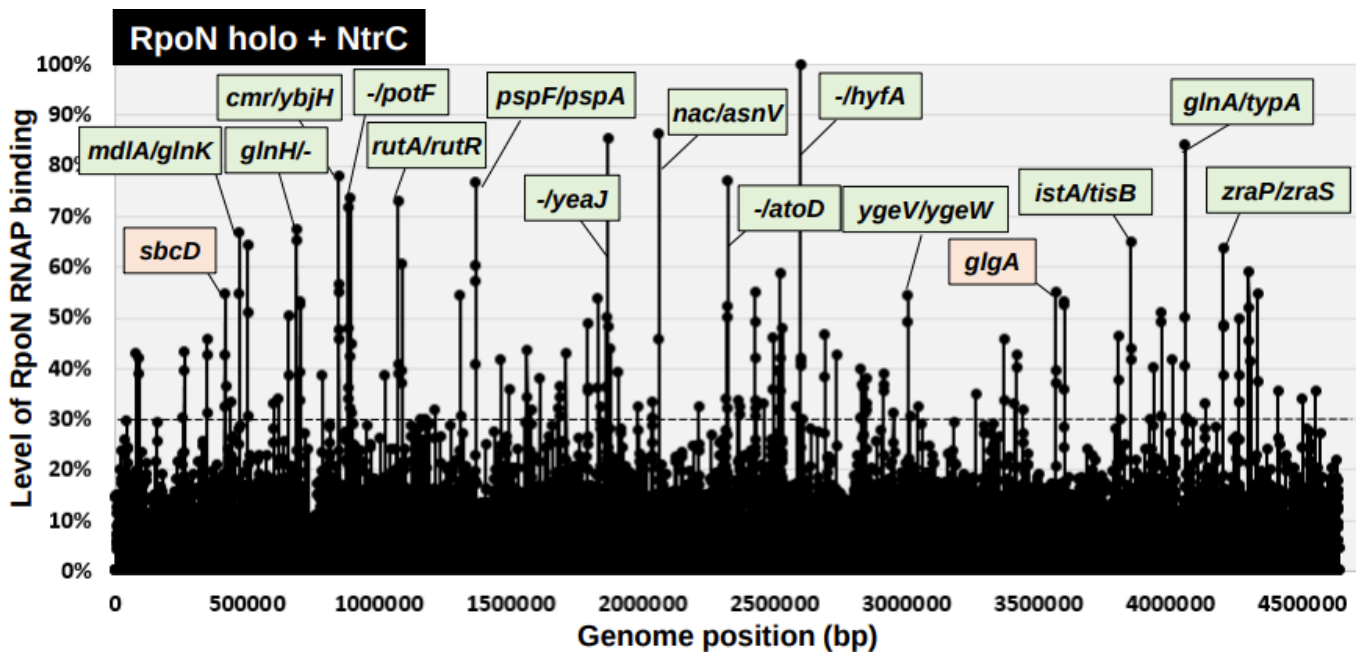
gSELEX was performed to search for the binding sites of NtrC. By setting the cut-off level to 40%, a total of 93 binding sites were identified (see Fig. 2 for gSELEX pattern), which are aligned along the map of the *E. coli* K-12 genome. Binding intensity of NtrC is shown in the NtrC column (the dark orange shading shows the intensity 61–100%, medium orange shading shows 41–60% and pale orange shading shows 30–40%). A total of 32 sites are located within intergenic spacers: 11 within type-A spacers; 19 within type-B spacers; 2 within type-C spacers (see Table S3). Columns D indicate the direction of transcription. Potential target genes or operons of NtrC were predicted based on the adjacent genes and gene orientation (shown with green shading). The grey shading shows genes that are not potential targets.

No.	gSELEX peak type	Map position (bp)	NtrC	Left gene function	Operon	Left gene	D	NtrC	D	Right gene	Operon	Right gene function
1	A	367 650	88%	DNA-binding transcriptional activator; 3HPP-binding	<i>mhpR-lacI</i>	<i>mhpR</i>	<		>	<i>mhpA</i>	<i>mhpABCDFE</i>	3-(3-Hydroxyphenyl)propionate hydroxylase
2	B	371 336	71%			<i>mhpC</i>	>			<i>mhpD</i>	<i>mhpDFE</i>	2-Keto-4-pentenoate hydratase
3	B	433 872	43%			<i>ribD</i>	>			<i>ribE</i>	<i>ribE-nusB-thiL-pgpA</i>	Riboflavin synthase beta chain
4	A	443 846	41%	2-Dehydropanoate reductase, NADPH-specific	<i>panE-yajL</i>	<i>panE</i>	<			<i>yajQ</i>	<i>yajQ</i>	Predicted nucleotide binding protein
5	B	471 846	56%			<i>mdtB</i>	>			<i>glnK</i>	<i>glnK-amtB</i>	Nitrogen assimilation regulatory protein for GlnL, GlnE and AmtB
6	B	547 672	44%			<i>ftrA</i>	>			<i>ybfF</i>	<i>ybfE-ybfF</i>	Conserved protein
7	A	655 760	41%	Anaerobic C4-dicarboxylate transport	<i>dcuC</i>	<i>dcuC</i>	<			<i>pegP</i>	<i>pegP</i>	Palmitoyl transferase for lipid A
8	B	688 560	56%	IS5 transposase and trans-activator	<i>insH</i>	<i>insH</i>	<			<i>Int</i>		
9	B	847 362	87%	Glutamine transporter subunit	<i>glnHPQ</i>	<i>glnH</i>	<			<i>dps</i>		
10	B	894 130	95%			<i>potF</i>	>			<i>potG</i>	<i>potGHI</i>	Putrescine transporter subunit; ATP-binding component of ABC superfamily
11	A	1 250 156	40%	Dihydroxyacetone kinase, N-terminal domain	<i>dhaKLM</i>	<i>dhaK</i>	<			<i>dhaR</i>	<i>dhaR</i>	Predicted DNA-binding transcriptional regulator; dihydroxyacetone
12	B	1 308 556	42%	Voltage-gated potassium channel	<i>kch</i>	<i>kch</i>	<			<i>yciI</i>		
13	B	1 613 766	40%			<i>yneJ</i>	>			<i>yneK</i>	<i>yneK</i>	Predicted protein
14	A	1 630 062	44%	Predicted mannose dehydrogenase	<i>ydfJ</i>	<i>ydfJ</i>	<			<i>ydfK</i>	<i>ydfK</i>	Qin prophage; predicted DNA-binding transcriptional regulator
15	B	1 653 158	41%	Predicted dehydratase	<i>rspAB</i>	<i>rspA</i>	<			<i>ynfA</i>		

Continued

Table 2. Continued

No.	gSELEX peak type	Map position (bp)	NtrC	Left gene function	Operon	Left gene	D	NtrC	D	Right gene	Operon	Right gene function
16	A	1 830 436	70%	Succinylornithine transaminase, PLP-dependent	<i>astCADBE</i>	<i>astC</i>	<		>	<i>xthA</i>	<i>xthA</i>	Exonuclease III
17	B	1 863 654	50%	Predicted oxidoreductase	<i>yeaE</i>	<i>yeaE</i>	<		<	<i>mipA</i>		
18	A	2 060 070	50%	DNA-binding transcriptional dual regulator of nitrogen assimilation	<i>nac</i>	<i>nac</i>	<		>	<i>asnV</i>	<i>asnV</i>	Asn tRNA
19	B	2 184 766	45%			<i>rcnA</i>	>		>	<i>rcnB</i>	<i>rcnB</i>	Periplasmic protein involved in nickel/cobalt export
20	A	2 458 968	47%	Conserved protein	<i>yfcZ</i>	<i>yfcZ</i>	<		>	<i>fadL</i>	<i>fadL</i>	Long-chain fatty acid outer membrane transporter
21	A	2 529 354	78%	Cell division protein involved in Z ring assembly	<i>zipA</i>	<i>zipA</i>	<		>	<i>cysZ</i>	<i>cysZ</i>	Predicted inner membrane protein
22	B	3 001 538	44%			<i>xdhB</i>	>		>	<i>xdhC</i>	<i>xdhC</i>	Xanthine dehydrogenase, Fe-S binding subunit
23	B	3 446 170	40%	50S ribosomal subunit protein L14	<i>rpINXE-rpsNH-rpIFR-rpsE-rpmD-rplO-secY-rpmJ</i>	<i>rpIN</i>	<		<	<i>rpsQ</i>		
24	B	3 933 336	59%			<i>rbsA</i>	>		>	<i>rbsC</i>	<i>rbsCBKR</i>	D-Ribose transporter subunit
25	B	3 994 336	40%			<i>yigA</i>	>		>	<i>xerC</i>	<i>xerC-yigB</i>	Site-specific tyrosine recombinase
26	A	4 056 244	97%	Glutamine synthetase	<i>ghnALG</i>	<i>ghnA</i>	<		>	<i>typA</i>	<i>typA</i>	GTP-binding protein
27	A	4 173 336	63%	Pantothenate kinase	<i>coaA</i>	<i>coaA</i>	<		>	<i>thrU</i>	<i>thrU-tyrU-glyT-thrT-tufB</i>	Thr tRNA
28	B	4 260 864	58%			<i>dusA</i>	>		>	<i>pspG</i>	<i>pspG</i>	Phage shock protein G
29	B	4 304 842	43%	Predicted alkyl sulfatase	<i>yjcS</i>	<i>yjcS</i>	<		<	<i>alsK</i>		
30	B	4 530 150	40%	KpLE2 phage-like element; predicted endoglucanase with Zn-dependent exopeptidase domain	<i>sgcXBCQAFR</i>	<i>sgcX</i>	<		<	<i>yjhp</i>		
			>40%	NtrC targets=30-41								
A=11, B=19		Cut-off	>40%									
		Spacer	30									



**Fig. 3.** gSELEX-chip search for the binding sequences of the RpoN RNAP holoenzyme in the presence of NtrC on the *E. coli* K-12 genome. gSELEX was performed to search for binding sites of the RpoN holoenzyme in the presence of NtrC. The y-axis represents the ratio against the highest peak at the *hyfA* promoter region and shows the level of RpoN holoenzyme-bound DNA fragments in the presence of NtrC, whereas the x-axis represents the position on the *E. coli* K-12 genome in bp. The adjacent gene on the *E. coli* K-12 genome of the peak position was indicated for high intensity peaks (>60 %). The peaks located within the spacer regions are shown with green labels, while the peaks located within the ORFs are shown with orange labels. A list of the binding sites of the RpoN holoenzyme in the presence of NtrC is described in Table 3 (detailed in Table S4).

isolated using anti-RpoC antibody and then subjected to tiling array analysis. By setting the cut-off level to 30% relative to the highest peak located upstream of *hyfA* (subunit of hydrogenase 4), a total of 108 NtrC-controlled RpoN holoenzyme-binding peaks were identified (Fig. 3), of which 61 peaks (56%) were located within intergenic spacers and 47 peaks (44%) were located inside the ORFs (for details see Table S4). Of the 61 RpoN holoenzyme-binding peaks within spacers, 19 peaks were located within type-A spacers of bidirectional transcription units (Table 3, type-A), and 42 were located inside type-B spacers upstream of one transcription unit but downstream of another transcription unit (Table 3, type-B). The promoters recognized by a combination of the RpoN holoenzyme and NtrC were predicted to be located in the type-A and type-B spacers (Table 3). Of the 108 NtrC-controlled RpoN holoenzyme-binding sites, 57 sites were detected in the absence of NtrC (Table 4, Fig. 4a), suggesting that these promoters (44 to 61) could be recognized by the RpoN holoenzyme alone without the support of NtrC (Fig. 4b). This finding indicates the presence of constitutive promoters for RpoN sigma, as in the case of the other six sigma factors [26, 27]. In contrast, a total of 21 to 27 promoters required NtrC for binding to the target promoters (Fig. 4b), of which 4 to 5 promoters were recognized by NtrC alone, while the other 17 to 22 promoters were recognized by a combination of RpoN and NtrC (Fig. 4b).

### Sequences recognized by the RpoN holoenzyme and NtrC

Using the RpoN-binding sequence from a small number of RpoN targets, a 17 bp long sequence consisting of conserved GG at the -24 site and GC at the -12 site was proposed as the RpoN promoter motif [6, 17], which is different from the well-known TTGACA (-35) and TATAAT (-10) promoter sequences of RpoD group sigma factors. The RpoN promoter motif was then re-evaluated using the entire set of 71 RpoN holoenzyme-binding targets (see Table 4, RpoN-holo row), which includes 32 known targets (Table S2). To identify the RpoN promoters within the binding sites of the RpoN holoenzyme, a collection of 500 bp sequences from 71 targets was analysed *in silico* search using the MEME program [38]. Subsequently, we identified a 15 bp long sequence, (-24 side) TGGCACnnTnTTGC (-12 side) (Table S2), which included the proposed RpoN promoter motif TGC at the -12 bp site and TGGCA at the -24 bp site (Fig. 4a). Previous studies have performed promoter sequence prediction using the experimental data obtained *in vivo* for enhancer-dependent promoters. Therefore, this study is to our knowledge the first to analyse the promoter sequence recognized by the RpoN holoenzyme alone in the absence of supporting TFs.

Using the DNA-binding sequences of several NtrC targets, a 17 bp long palindromic sequence consisting of a 17 bp



**Table 3.** RpoN holoenzyme-binding sites in the presence of NtrC on the *E. coli* K-12 genome

gSELEX was performed to search for binding sites of the RNAP RpoN holoenzyme in the presence of NtrC. By setting the cut-off level to 30%, a total of 108 binding sites were identified (see Fig. 3 for gSELEX pattern), which are aligned along the map of the *E. coli* K-12 genome. Binding intensity of NtrC is shown in the RpoN holo+NtrC column (the dark orange shading shows the intensity 61–100%, medium orange shading shows 41–60% and pale orange shading shows 30–40%). A total of 61 sites are located within intergenic spacers; 19 within type-A spacers; 42 within type-B spacers (see Table S4). Columns D indicate the direction of transcription. Potential target genes or operons of RpoN in the presence of NtrC were predicted based on the adjacent genes and gene orientation (shown with green shading). The grey shading shows genes that are not potential targets.

No.	gSELEX peak type	Map position (bp)	RpoN holo+NtrC	Left gene function	Operon	Left gene	D	RpoN+NtrC	D	Right gene	Operon	Right gene function
1	A	42 372	30%	Predicted transporter	<i>cai</i> TABCADE	<i>cai</i> T	<		>	<i>fixA</i>	<i>fixABCX</i>	Predicted electron transfer flavoprotein subunit, ETFP adenine nucleotide-binding domain
2	A	77 346	43%	DNA-binding transcriptional regulator	<i>sgrR-sroA-thpA-thiPQ</i>	<i>sgrR</i>	<		>	<i>sgrS</i>	<i>sgrST-setA</i>	ncRNA
3	B	257 850	30%			<i>fixA</i>	>		>	<i>ctrl</i>	<i>ctrl</i>	DNA-binding transcriptional regulator
4	A	347 864	43%	DNA-binding transcriptional activator	<i>prpR</i>	<i>prpR</i>	<		>	<i>prpB</i>	<i>prpBCDE</i>	2-Methylisocitrate lyase
5	B	418 832	37%			<i>phoR</i>	>		>	<i>brnQ</i>	<i>brnQ-proY</i>	Predicted branched chain amino acid transporter (LIV-II)
6	B	471 846	67%			<i>mdlB</i>	>		>	<i>glnK</i>	<i>glnK-amtB</i>	Nitrogen assimilation regulatory protein for GlnL, GlnE and AmtB
7	B	619 432	34%	Iron-enterobactin transporter subunit	<i>fepC</i>	<i>fepC</i>	<		<	<i>fepG</i>		
8	A	655 760	50%	Anaerobic C4-dicarboxylate transport	<i>dcuC</i>	<i>dcuC</i>	<		>	<i>pagP</i>	<i>pagP</i>	Palmitoyl transferase for lipid A
9	B	688 560	67%	IS5 transposase and trans-activator	<i>insH</i>	<i>insH</i>	<		<	<i>Int</i>		
10	A	784 656	39 %	Conserved protein	<i>ybgS</i>	<i>ybgS</i>	<		>	<i>aroG</i>	<i>aroG</i>	3-Deoxy-D-arabino-heptulosonate-7-phosphate synthase, phenylalanine repressible
11	B	847 362	78%	Glutamine transporter subunit	<i>glnHPQ</i>	<i>glnH</i>	<		<	<i>dps</i>		
12	A	882 830	72%	Undecaprenyl pyrophosphate phosphatase	<i>ybjG</i>	<i>ybjG</i>	<		>	<i>cmr</i>	<i>cmr</i>	Multidrug efflux system protein
13	B	891 170	32%			<i>rfsA</i>	>		>	<i>rimK</i>	<i>rimK-ybjN</i>	Ribosomal protein S6 modification protein

Continued

**Table 3.** Continued

No.	gSELEX peak type	Map position (bp)	RpoN holo+NtrC	Left gene function	Operon	Left gene	D	RpoN+NtrC	D	Right gene	Operon	Right gene function
14	B	892 632	74 %			<i>ybjN</i>	>		>	<i>potF</i>	<i>potFGHI</i>	Putrescine transporter subunit; periplasmic-binding component of ABC superfamily
15	B	894 130	45%			<i>potF</i>	>		>	<i>potG</i>	<i>potGHI</i>	Putrescine transporter subunit; ATP-binding component of ABC superfamily
16	A	1 073 268	73%	Predicted monoxygenase	<i>rutABCDEFGHI</i>	<i>rutA</i>	<		>	<i>rutR</i>	<i>rutR</i>	Predicted DNA-binding transcriptional regulator
17	B	1 177 842	30%			<i>lolE</i>	>		>	<i>nagK</i>	<i>nagK-cobB</i>	N-Acetyl-D-glucosamine kinase
18	B	1 308 556	54%	Voltage-gated potassium channel	<i>kch</i>	<i>kch</i>	<		<	<i>ycil</i>		
19	A	1 366 070	77%	DNA-binding transcriptional activator	<i>pspF</i>	<i>pspF</i>	<		>	<i>pspA</i>	<i>pspABCDE</i>	Regulatory protein for phage-shock-protein operon
20	B	1 561 132	44%	D-Ala-D-Ala dipeptidase, Zn-dependent	<i>ddpXABCDF</i>	<i>ddpX</i>	<		<	<i>dos</i>		
21	B	1 608 732	38%	Altronate oxidoreductase, NAD-dependent	<i>uxaB</i>	<i>uxaB</i>	<		<	<i>yneF</i>		
22	B	1 678 972	32%			<i>ydgI</i>	>		>	<i>folM</i>	<i>folM</i>	Dihydrofolate reductase isozyme
23	B	1 709 534	43%			<i>rsxE</i>	>		>	<i>nth</i>	<i>nth</i>	DNA glycosylase and apyrimidinic lyase (endonuclease III)
24	A	1 830 436	54%	Succinylornithine transaminase, PLP-dependent	<i>astCADBE</i>	<i>astC</i>	<		<	<i>xthA</i>	<i>xthA</i>	Exonuclease III
25	B	1 863 654	50%	Methylglyoxal reductase	<i>yeaE</i>	<i>yeaE</i>	<		<	<i>mipA</i>		
26	B	1 905 652	39%	Predicted protein	<i>yobF-cspC</i>	<i>yobF</i>	<		<	<i>yebO</i>		
27	A	2 036 832	33%	Predicted DNA-binding response regulator in TCS with YedV	<i>yedWV</i>	<i>yedW</i>	<		>	<i>hiuH</i>	<i>hiuH</i>	Hydroxyisourate hydrolase/transthyretin-related protein
28	A	2 060 070	86%	DNA-binding transcriptional dual regulator of nitrogen assimilation	<i>nac</i>	<i>nac</i>	<		>	<i>asnV</i>	<i>asnV</i>	Asn tRNA
29	B	2 321 470	77%			<i>atoC</i>	>		>	<i>atoD</i>	<i>atoDAEB</i>	Acetyl-CoA:acetoacetyl-CoA transferase, alpha subunit

Continued

Table 3. Continued

No.	gSELEX peak type	Map position (bp)	RpoN holo+NtrC	Left gene function	Operon	Left gene	D	RpoN+NtrC	D	Right gene	Operon	Right gene function
30	B	2 360 468	34%	Predicted DNA-binding transcriptional regulator	<i>yfaX-rhmD-yfaVU</i>	<i>yfaX</i>	<		<	<i>yfaY</i>		
31	B	2 425 832	55%	Lysine/arginine/ornithine transporter subunit	<i>argT-hisIQMP</i>	<i>argT</i>	<		<	<i>ubiX</i>		
32	B	2 429 072	32%	Membrane protein required for colicin V production	<i>cvpA-purE-ubiX</i>	<i>cvpA</i>	<		<	<i>dedD</i>		
33	A	2 458 968	33%	Conserved protein	<i>yfcZ</i>	<i>yfcZ</i>	<		<	<i>fadL</i>	<i>fadL</i>	Long-chain fatty acid outer membrane transporter
34	A	2 493 362	46%	Predicted inner membrane protein	<i>yfdY</i>	<i>yfdY</i>	<		>	<i>lpxP</i>	<i>lpxP</i>	Palmitoleyl-acyl carrier protein (ACP)-dependent acyltransferase
35	B	2 520 564	59%	DNA-binding transcriptional activator	<i>xapR</i>	<i>xapR</i>	<		<	<i>xapB</i>		
36	B	2 522 072	36%	Xanthosine transporter	<i>xapB</i>	<i>xapB</i>	<		<	<i>xapA</i>		
37	B	2 531 464	48%			<i>cysK</i>	>		>	<i>ptsH</i>	<i>ptsH-crr</i>	Phosphohistidinoprotein-hexose phosphotransferase component of PTS system
38	B	2 599 140	100%			<i>bcp</i>	>		>	<i>hyfA</i>	<i>hyfABCDEFGHIJ-foxB</i>	Hydrogenase 4, 4Fe-4S subunit
39	B	2 689 364	47%	ncRNA	<i>glmY</i>	<i>glmY</i>	<		<	<i>purL</i>		
40	B	2 825 748	40%			<i>srIB</i>	>		>	<i>srID</i>	<i>srID-gutM-srIR-gutQ</i>	Sorbitol-6-phosphate dehydrogenase
41	A	2 830 336	30%	DNA-binding transcriptional activator	<i>norR</i>	<i>norR</i>	<		<	<i>norV</i>	<i>norVW</i>	Flavorubredoxin oxidoreductase
42	B	2 836 270	36%	Formate dehydrogenase-H, [4Fe-4S] ferredoxin subunit	<i>hydN-hypF</i>	<i>hydN</i>	<		<	<i>ascG</i>		
43	A	2 848 650	38%	Regulator of the transcriptional regulator FhIA	<i>hycABCDEFGHI</i>	<i>hycA</i>	<		>	<i>hypA</i>	<i>hypABCDE-fhlA</i>	Protein involved in nickel insertion into hydrogenases 3
44	A	3 004 270	55%	Predicted DNA-binding transcriptional regulator	<i>ygeV</i>	<i>ygeV</i>	<		>	<i>ygeW</i>	<i>ygeW</i>	Conserved protein
45	B	3 043 930	33%	Predicted NAD(P)-binding oxidoreductase with NAD(P)-binding Rossmann-fold domain	<i>ygfF</i>	<i>ygfF</i>	<		<	<i>gcvP</i>		
46	B	3 370 654	46%	Sialic acid transporter	<i>nanTEK-yhcH</i>	<i>nanT</i>	<		<	<i>nanA</i>		

Continued

Table 3. Continued

No.	gSELEX peak type	Map position (bp)	RpoN holo+NtrC	Left gene function	Operon	Left gene	D	RpoN+NtrC	D	Right gene	Operon	Right gene function
47	B	3 408 032	33%			<i>prnA</i>	>			<i>dusB</i>	<i>dusB-fis</i>	tRNA-dihydrouridine synthase B
48	B	3 417 032	43 %			<i>yhdV</i>	>			<i>yhdX</i>	<i>yhdXYZ</i>	Predicted amino-acid transporter subunit
49	B	3 440 634	32%	30S ribosomal subunit protein S13	<i>rpsMKD-rpoA-rplQ</i>	<i>rpsM</i>	<			<i>rpmJ</i>		
50	B	3 598 870	53%	RNAP sigma 32 (sigma H) factor	<i>rpoH</i>	<i>rpoH</i>	<			<i>fisX</i>		
51	B	3 809 172	30%	Formamidopyrimidine/5-formyluracil/5-hydroxymethyluracil DNA glycosylase	<i>mutM</i>	<i>mutM</i>	<			<i>rpmG</i>		
52	A	3 851 352	65%	ncRNA	<i>istR</i>	<i>istR</i>	<			<i>tisB</i>	<i>tisB</i>	LexA-regulated toxic peptide
53	B	3 967 058	51%			<i>rje</i>	>			<i>wzzE</i>	<i>wzzE-wecBC-rffGHCA-wzxE-rffT-wzyE-rffM</i>	Enterobacterial common antigen polysaccharide chain length modulation protein
54	B	4 008 248	42%			<i>pldB</i>	>			<i>yigL</i>	<i>yigL</i>	Predicted hydrolase
55	A	4 056 244	84%	Glutamine synthetase	<i>ghnALG</i>	<i>ghnA</i>	<			<i>typA</i>	<i>typA</i>	GTP-binding protein
56	B	4 131 538	33%			<i>metF</i>	>			<i>katG</i>	<i>katG</i>	Catalase/hydroperoxidase HPI(I)
57	A	4 199 860	64 %	Zn-binding periplasmic protein	<i>zraP</i>	<i>zraP</i>	<			<i>zraS</i>	<i>zraSR</i>	Sensory histidine kinase in two-component regulatory system with ZraR
58	B	4 260 864	50%			<i>dusA</i>	>			<i>pspG</i>	<i>pspG</i>	Phage shock protein G
59	B	4 297 530	59%	Formate dehydrogenase-H, selenopolypeptide subunit	<i>fdhF</i>	<i>fdhF</i>	<			<i>mdtP</i>		
60	B	4 304 842	41%	Predicted alkyl sulfatase	<i>yjcS</i>	<i>yjcS</i>	<			<i>alsK</i>		
61	B	4 331 330	55%	Sensory histidine kinase in two-component regulatory system with BasR	<i>basS</i>	<i>basS</i>	<			<i>basR</i>		
<b>Regulatory targets=61-80</b>												
		<b>Cut-off</b>	<b>&gt;30%</b>									
		<b>Spacer</b>	<b>108</b>									



**Table 4.** Summary of binding sites of the RpoN holoenzyme and NtrC

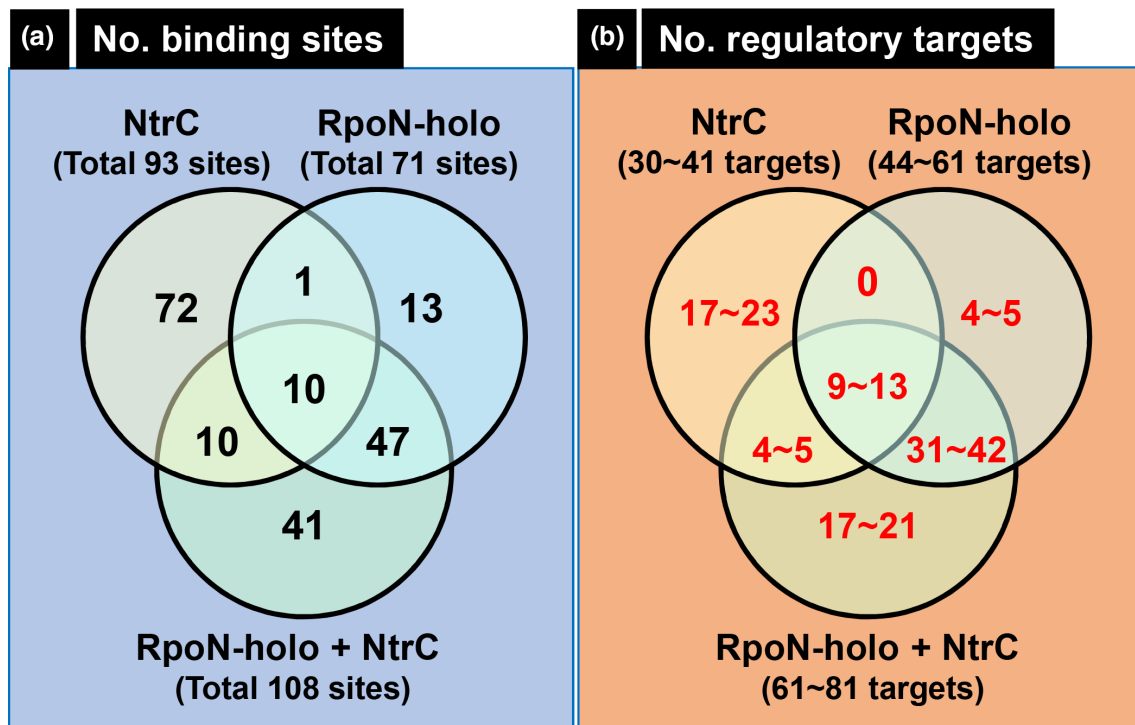
The binding site of each RpoN holoenzyme and NtrC on the *E. coli* K-12 W3110 genome was determined *in vitro* using the gSELEX screening system. Details of the experimental procedures are described in a previous study [23]. The number of the target transcription units was estimated based on the location of the binding sites

Regulator	Total no. of binding sites	Inside spacer			Inside ORF	No. of regulatory targets		
		Type-A	Type-B	Type-C		Type-A spacer	Type-B spacer	Total
RpoN-holo	71	17	27	0	27 (38%)	17–34	27	44–61
			Total 44 (62%)					
NtrC	93	11	19	2	61 (66%)	11–22	19	30–41
			Total 32 (34%)					
RpoN-holo+NtrC	108	20	41	0	47 (44%)	20–40	41	61–81
			Total 61 (56%)					

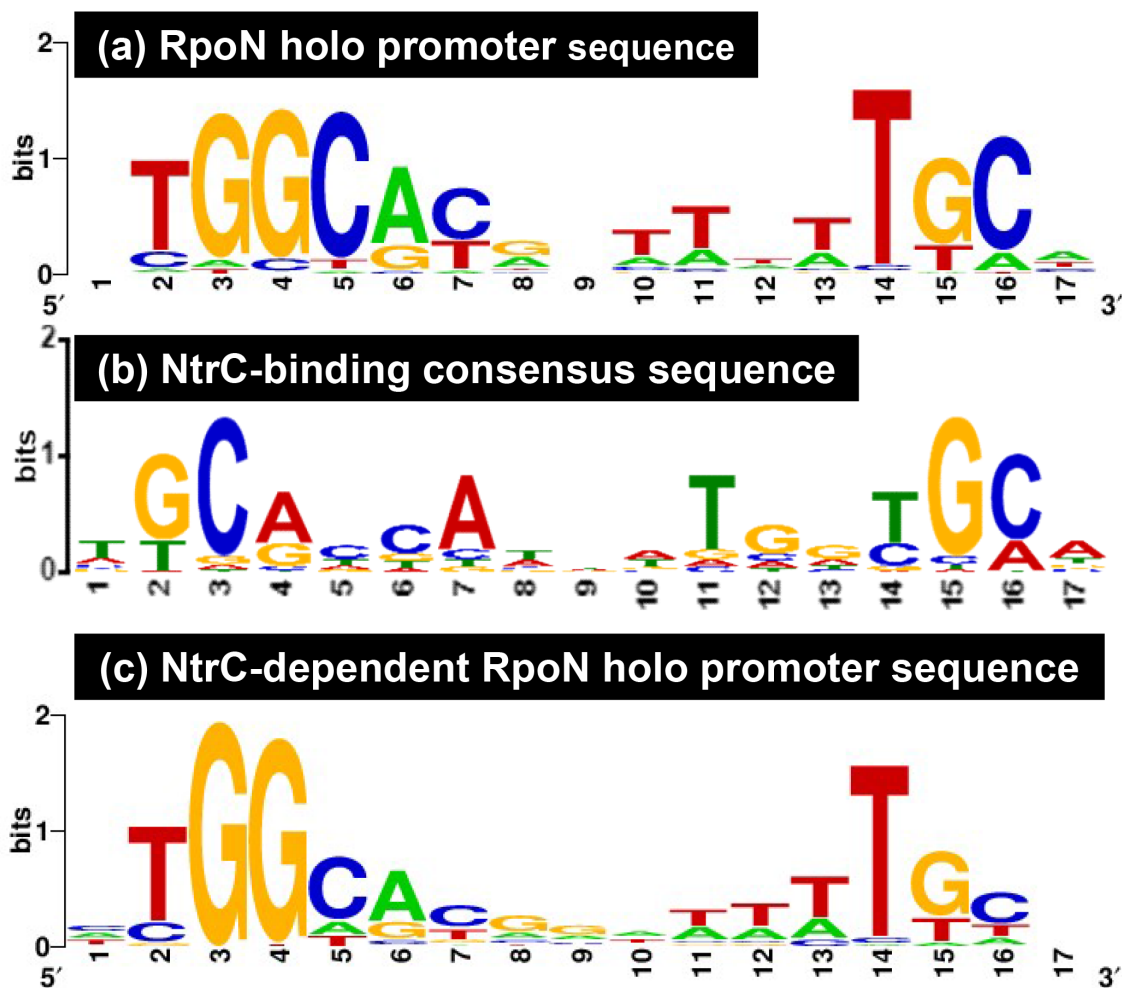
long sequence of TGCACCAnnnTGGTGCA was proposed as the consensus recognition sequence of NtrC [8]. As we obtained a large number of NtrC-binding sites by gSELEX, the consensus sequence of NtrC binding was re-evaluated using the whole set of 93 targets, including 7 known targets (Table 2). A collection of 500 bp sequences from these targets was analysed using the MEME program. Subsequently, we identified a 17 bp long sequence (Fig. 5b), which contained highly conserved GCAnnA and TnnTGC. This sequence is in good agreement with a previous report using *in vitro*

experimental evidence [8]. Thus, we concluded that this highly conserved (T)GCA(CC)AnnnT(GG)TGC(A) 17 bp long NtrC-box sequence is required for the tight binding of NtrC.

Finally, we analysed the promoter sequences recognized by the RpoN holoenzyme in the presence of excess NtrC. Some of the conserved sequences of promoters recognized by the RpoN holoenzyme alone were lost in the presence of NtrC, suggesting a certain level of alteration of the promoter



**Fig. 4.** Correlation diagrams of the targets between RpoN and NtrC. Venn diagram summarizing the correlation of target sites of RpoN holoenzyme and NtrC. The number of binding sites is shown in (a), while the number of regulatory targets is shown in (b). All the 14 sites detected in the RpoN holoenzyme but not in RpoN holoenzyme +NtrC showed over 24% intensity in RpoN holoenzyme +NtrC (for counting the number of targets, the cut-off level was set as 30%) (Table 5).



**Fig. 5.** Consensus sequences of the RpoN holoenzyme promoter and NtrC binding. The promoter motif of RpoN holoenzyme, in the presence or absence of NtrC and binding sequences of NtrC, was analysed using the MEME program. The sequences are listed in Tables S2–S4, and were subjected to Logo analysis for the determination of the consensus sequences for the following samples: (a) the whole set of RpoN holoenzyme targets (total 71 sequences in Table S2); (b) the whole set of NtrC targets (total 93 sequences in Table S3); (c) the whole set of RpoN holoenzyme targets in the presence of NtrC, not included in the RpoN holoenzyme targets in the absence of NtrC (total 51 sequences in Table S4).

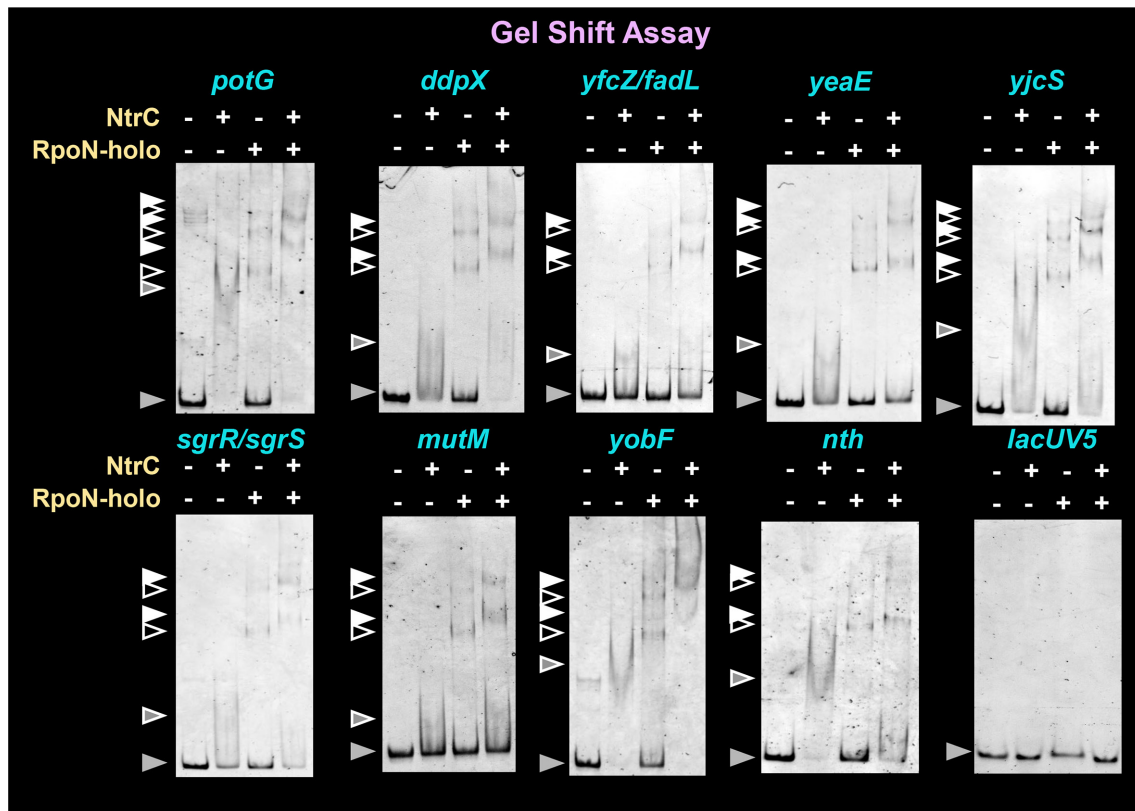
recognition property in the presence of NtrC. Compared with the promoter sequence of the RpoN holoenzyme alone (Fig. 5a), the NtrC-dependent RpoN promoter sequence showed high-level conservation at the 3rd G and 4th G, but low-level conservation at the 5th, 15th and 16th C (Fig. 5c). These results suggest that NtrC modulates the promoter recognition property of the RpoN holoenzyme to recognize sequences of low-level conservation at the position of the -12 GC element.

The RpoN holoenzyme binds to promoters with conserved sequence elements at -24 GG and -12 GC. One unique feature of the NtrC-dependent RpoN promoter identified in this study is the conservation of these elements, which was low for -12 GC and high for -24 GG (Fig. 5). This -12 GC element is involved in the stability of the RpoN holoenzyme–target promoter complex *in vitro* [48], while the -24 GG element

is the dominant element for promoter binding by the RpoN holoenzyme [49]. In conjunction with the results of the gel-shift assay (Fig. 6), NtrC appears to support the stability of the formation of RpoN holoenzyme–promoter complex, which has a low-level conservation of the -12 GC element.

#### Confirmation of the interaction of newly identified regulatory targets with RpoN holoenzyme and NtrC Gel-shift assay *in vitro*

Based on gSELEX-chip analysis, we identified 44–61 promoters for the RpoN holoenzyme alone and 61–81 promoters for the NtrC-supported RpoN holoenzyme (Table 4, Fig. 4), including approximately 40 hitherto identified RpoN promoters. Similarly, we identified a total of 30–41 regulatory targets for NtrC via gSELEX screening (Table 4, Fig. 4). To experimentally confirm the regulation of newly



**Fig. 6.** NtrC-dependent RpoN holoenzyme–DNA complex formation. The target promoter fragments were mixed with the RpoN holoenzyme (0.3  $\mu$ M, lane 2), NtrC protein (15  $\mu$ M, lane 3), or both in combination with the addition of 25 mM acetylphosphate (lane 4). After incubation at 37  $^{\circ}$ C for 30 min, the reaction mixture was subjected to 3.5% PAGE. Grey triangles indicate the free probe; grey triangles with white frame indicate the NtrC–probe complex; black triangles with white frame indicate the RpoN holoenzyme–probe complex; white triangles indicate the RpoN holoenzyme–NtrC–probe complex.

identified target promoters with RpoN and/or NtrC, the interaction of the RpoN holoenzyme or NtrC with some representative targets was analysed by gel-shift assay *in vitro*, and RT-qPCR assay *in vivo*, of target mRNA.

To confirm the binding activity of both the RpoN holoenzyme and NtrC to the target promoters *in vitro*, we performed a gel-shift assay to detect the test protein–target DNA complexes. From the newly identified target genes, seven independent spacer probes containing nine representative targets were selected: *yfcZ* (uncharacterized conserved protein)/*fadL* (long-chain fatty acid outer membrane transporter), *yeaE* (aldo-keto reductase), *yjcS* (metallo-lactamase family protein), *sgrR* (TF)/*sgrS* (small RNA antisense regulator), *mutM* (DNA glycosylase), *yobF* (hypothetical protein) and *nth* (endonuclease III) (targets are indicated in red in Table 5). In addition, we used two hitherto known RpoN promoters, namely, *potG* encoding putrescine transporter and *ddpX* encoding D-Ala-D-Ala dipeptidase [12] (these two genes were also detected in gSELEX screening), and *lacUV5* as a reference control. Each of these test probes was mixed with purified NtrC, RpoN holoenzyme or both, and the probe–test protein(s) mixtures were then directly subjected to PAGE. The two known NtrC target probes, *potG* and *ddpX*, formed

both NtrC–probe binary complexes and RpoN holoenzyme–probe binary complexes (Fig. 6; *potG* and *ddpX* panels, lanes 2 and 3). In both cases, two or three bands were detected, suggesting the presence of more than one promoter-like sequence on the *potG* and *ddpX* promoter probes. In the presence of both NtrC and the RpoN holoenzyme, the migration of the complex bands was significantly retarded, indicating the formation of RpoN holoenzyme–NtrC–probe ternary complexes (Fig. 6; *potG* and *ddpX* panels, lane 4). Next, we assessed seven spacer probes containing nine newly identified targets from the gel-shift assay under the same conditions as those employed for the two known targets. As in the case of *potG* and *ddpX* probes, the binding of NtrC alone (Fig. 6, lane 2) was confirmed by the disappearance of free probes. However, the expected probe–NtrC complex formed a smear band, likely due to the gradual dissociation of low-affinity NtrC during PAGE. In contrast, all seven DNA probes formed two or three detectable bands of RpoN holoenzyme–probe complexes (Fig. 6, lane 3). In the presence of both the RpoN holoenzyme and NtrC, the intensity of the free promoter probe clearly decreased for all seven probes, indicating an increase in DNA-binding intensity in the presence of both the RpoN holoenzyme and NtrC (Fig. 6, lanes 3 and 4). The

**Table 5.** Summary of regulatory targets of RpoN holoenzyme, NtrC and RpoN holoenzyme–NtrC

Binding sites and intensity of RpoN (Table 1), NtrC (Table 2) and RpoN–NtrC (Table 3) were combined (the dark orange shading shows the intensity 61–100%, medium orange shading shows 41–60% and pale orange shading shows 30–40%). The known regulatory targets of RpoN and NtrC in the RegulonDB database are shown in the RegulonDB columns, respectively (shown in yellow). RpoN targets detected by ChIP-chip analysis are shown in the ChIP-chip column (shown in blue). Columns D indicate the direction of transcription. Potential target genes or operons of RpoN and/or NtrC were predicted based on the adjacent genes and gene orientation (shown with green shading). The grey shading shows genes that are not potential targets. The targets analysed *in vitro* and *in vivo* are shown in red or blue, respectively.

No.	gSELEX peak type	Map position (bp)	RpoN holo+NtrC	RpoN holo	NtrC	Group	Regulon RpoN	Regulon NtrC	ChIP-chip	Operon	Left gene	D	Test protein	D	Right gene	Operon
1	A	42 372								<i>catI</i> TABCDE	<i>catI</i> T	<		<	<i>fixA</i>	<i>fixA</i> ABCX
2	A	77 346				II-A				<i>sgtR-sroA-thpA-thiPQ</i>	<i>sgtR</i>	<		<	<i>sgtS</i>	<i>sgtST-setA</i>
3	B	257 850									<i>frsA</i>	>		>	<i>crf</i>	<i>crf</i>
4	A	347 864								<i>prpR</i>	<i>prpR</i>	<		<	<i>prpB</i>	<i>prpBCDE</i>
5	A	367 650								<i>mhpR-lacI</i>	<i>mhpR</i>	<		<	<i>mhpA</i>	<i>mhpABCDFE</i>
6	B	371 336									<i>mhpC</i>	>		>	<i>mhpD</i>	<i>mhpDFE</i>
7	B	418 832									<i>phoR</i>	>		>	<i>brnQ</i>	<i>brnQ-proY</i>
8	B	433 872									<i>ribD</i>	>		>	<i>ribE</i>	<i>ribE-nusB-thiL-pgpA</i>
9	A	443 846								<i>panE-yajL</i>	<i>panE</i>	<		<	<i>yajQ</i>	<i>yajQ</i>
10	B	471 846									<i>mdlB</i>	>		>	<i>glnK</i>	<i>glnK-antB</i>
11	B	547 672									<i>fdxA</i>	>		>	<i>ybfF</i>	<i>ybfE-ybfF</i>
12	B	619 432								<i>fepC</i>	<i>fepC</i>	<		<	<i>fepG</i>	
13	A	655 760								<i>dcaC</i>	<i>dcaC</i>	<		<	<i>pagP</i>	<i>pagP</i>
14	B	688 560								<i>insH</i>	<i>insH</i>	<		<	<i>lnt</i>	
15	A	784 656								<i>ybgS</i>	<i>ybgS</i>	<		<	<i>aroG</i>	<i>aroG</i>
16	B	847 362								<i>glnHPQ</i>	<i>glnH</i>	<		<	<i>dps</i>	
17	B	874 568									<i>yliE</i>	>		>	<i>yliF</i>	<i>yliF</i>
18	A	882 830								<i>ybjG</i>	<i>ybjG</i>	<		<	<i>cmr</i>	<i>cmr</i>
19	B	891 170								<i>rfsA</i>	<i>rfsA</i>	>		>	<i>rimK</i>	<i>rimK-ybjN</i>
20	B	892 632								<i>ybjN</i>	<i>ybjN</i>	>		>	<i>potF</i>	<i>potFGHI</i>
21	B	894 130				I-B				<i>potF</i>	<i>potF</i>	>		>	<i>potG</i>	<i>potGHI</i>
22	A	1 073 268								<i>rutA</i>	<i>rutA</i>	<		<	<i>rutR</i>	<i>rutR</i>
23	B	1 177 842								<i>lolE</i>	<i>lolE</i>	>		>	<i>nagK</i>	<i>nagK-cobB</i>
24	B	1 191 232								<i>purB</i>	<i>purB</i>	<		<	<i>lplD</i>	
25	A	1 250 156								<i>dhaKLM</i>	<i>dhaK</i>	<		<	<i>dhaR</i>	<i>dhaR</i>

Continued



Table 5. Continued

26	B	1 308 556				<i>kch</i>	<i>kch</i>	<	<	<i>yciI</i>
27	A	1 366 070				<i>pspF</i>	<i>pspF</i>	<	>	<i>pspA</i> <i>pspABCDE</i>
28	B	1 527 534				<i>yncH</i>	<i>yncH</i>	>	>	<i>ydcD</i> <i>ydcD</i>
29	B	1 561 132	II-B			<b><i>ddpXABCDF</i></b>	<b><i>ddpX</i></b>	<	<	<i>dos</i>
30	B	1 613 766				<i>yneJ</i>	<i>yneJ</i>	>	>	<i>yneK</i> <i>yneK</i>
31	A	1 630 062				<i>ydfI</i>	<i>ydfI</i>	<	>	<i>ydfK</i> <i>ydfK</i>
32	B	1 653 158				<i>rspAB</i>	<i>rspA</i>	<	<	<i>ynfA</i>
33	B	1 678 972				<i>yegI</i>	<i>yegI</i>	>	>	<i>folM</i> <i>folM</i>
34	B	1 709 534	II-B			<i>rsxE</i>	<i>rsxE</i>	>	>	<i>nth</i> <i>nth</i>
35	A	1 830 436				<i>astCADBE</i>	<i>astC</i>	<	>	<i>xthA</i> <i>xthA</i>
36	B	1 863 654	I-B			<i>yeaE</i>	<i>yeaE</i>	<	<	<i>mipA</i>
37	B	1 905 652	II-B			<b><i>yobF-espC</i></b>	<b><i>yobF</i></b>	<	<	<i>yebO</i>
38	A	2 036 832				<i>yedWV</i>	<i>yedW</i>	<	>	<i>hiuH</i> <i>hiuH</i>
39	A	2 060 070				<i>nac</i>	<i>nac</i>	<	>	<i>asnV</i> <i>asnV</i>
40	B	2 184 766				<i>renA</i>	<i>renA</i>	>	>	<i>renB</i> <i>renB</i>
41	B	2 321 470				<i>atoC</i>	<i>atoC</i>	>	>	<i>atoD</i> <i>atoDAEB</i>
42	B	2 360 468				<i>yfaX-rhmD-yfaVU</i>	<i>yfaX</i>	<	<	<i>yfaY</i>
43	A	2 411 432				<i>yfbV</i>	<i>yfbV</i>	<	>	<i>ackA</i> <i>ackA-pia</i>
44	B	2 425 832				<i>argT-his/QMP</i>	<i>argT</i>	<	<	<i>ubiX</i>
45	B	2 429 072				<i>cvpA-purF-ubiX</i>	<i>cvpA</i>	<	<	<i>dedD</i>
46	A	2 458 968	I-A			<i>yfzZ</i>	<i>yfzZ</i>	<	>	<i>fadL</i> <i>fadL</i>
47	A	2 493 362				<i>yfjY</i>	<i>yfjY</i>	<	>	<i>lpxP</i> <i>lpxP</i>
48	B	2 520 564				<i>xapR</i>	<i>xapR</i>	<	<	<i>xapB</i>
49	B	2 522 072				<i>xapB</i>	<i>xapB</i>	<	<	<i>xapA</i>
50	A	2 529 354				<i>zipA</i>	<i>zipA</i>	<	>	<i>cysZ</i> <i>cysZ</i>
51	B	2 531 464				<i>cysK</i>	<i>cysK</i>	>	>	<i>ptsH</i> <i>ptsHI-arr</i>
52	B	2 599 140				<i>bcp</i>	<i>bcp</i>	>	>	<i>hyfA</i> <i>hyfABCDEFGHIJR-jocB</i>
53	B	2 689 364				<i>glmY</i>	<i>glmY</i>	<	<	<i>purL</i>
54	B	2 825 748				<i>norR</i>	<i>srlB</i>	>	>	<i>srlD</i> <i>srlD-gutM-srlR-gutQ</i>
55	A	2 830 336				<i>norR</i>	<i>norR</i>	<	>	<i>norV</i> <i>norVW</i>

Continued



simultaneous binding of the RpoN holoenzyme and NtrC was observed, based on the super shift of protein–DNA complexes from RpoN holoenzyme alone (lane 3) to the mixture of RpoN holoenzyme and NtrC (lane 4). These observations indicate the enhancement of RpoN holoenzyme binding to target promoters by NtrC. Neither the RpoN holoenzyme nor NtrC exhibited binding to a non-specific *lacUV5* promoter region used as an internal reference (Fig. 6, *lacUV5* panel).

### RT-qPCR assay *in vivo*

All seven promoter probes containing nine promoters were found to interact with both the RpoN holoenzyme and NtrC by the gel-shift assay *in vitro*. Next, we examined these interactions *in vivo* using an RT-qPCR assay. In addition to the seven probes employed in the gel-shift assay, we added an additional six probes in the *in vivo* assay, including *ntrC* (TF; under the control of the *glnA* promoter), *potI* (putrescine transporter; under the control of the *potF* promoter), *ddpF* (ABC-family transporter; under the control of *ddpX* promoter), *thiQ* (thiamine transporter; under the control of the *sgrR* promoter), *setA* (sugar efflux system; under the control of the *sgrS* promoter) and *cspC* (stress protein; under the control of the *yobF* promoter). All these genes were organized into their corresponding operons under the control of the NtrC regulator; the operons, including these six target genes, are indicated in Table 5.

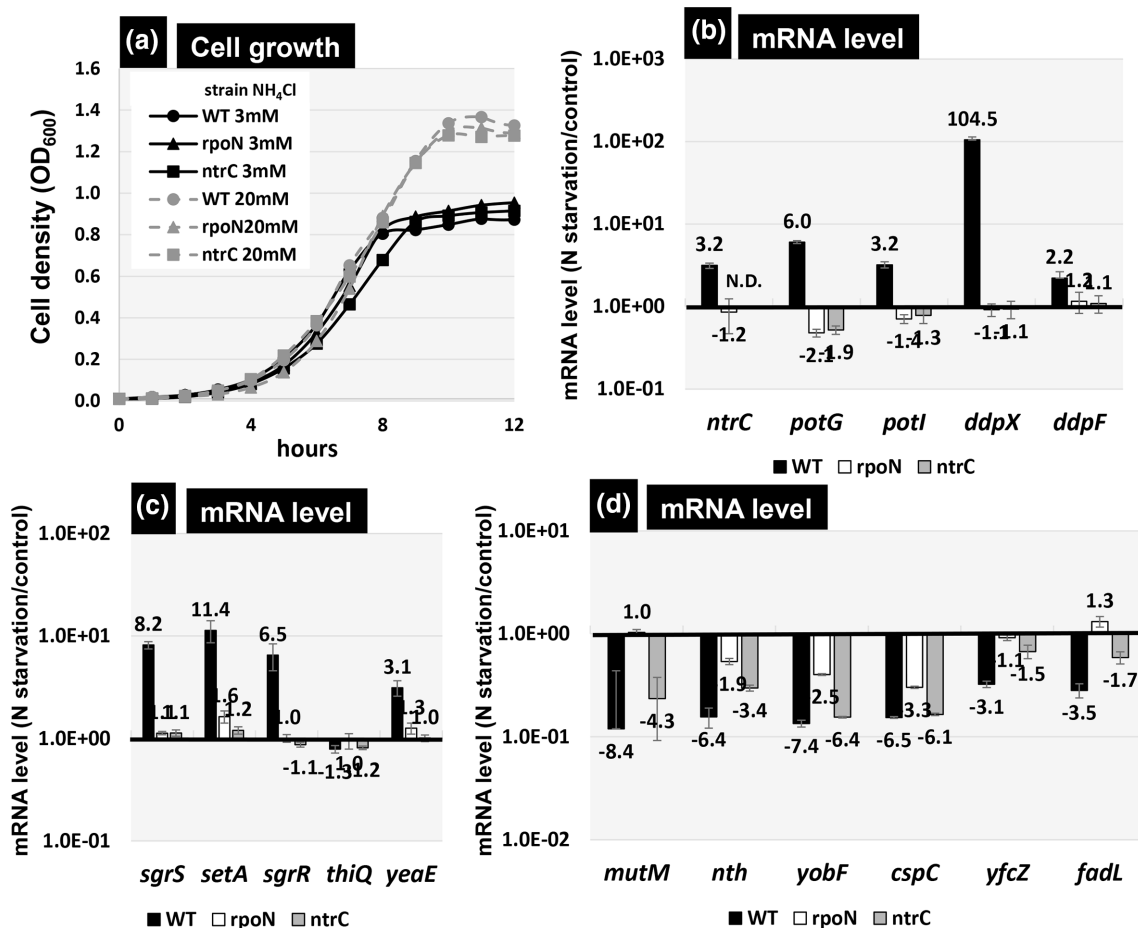
Intracellular levels of RpoN and NtrC are activated in response to nitrogen depletion [6, 47]. Gutnick minimal medium is widely used as a nitrogen source control medium [21, 34]. We observed the growth of the wild-type *E. coli* K-12 BW25113 strain and its *rpoN* and *ntrC* mutation strains. In the nitrogen-rich medium (supplemented with 20 mM NH<sub>4</sub>Cl), the cell density of all three strains reached OD<sub>600</sub> 1.3 (Fig. 7a) at 10 h after inoculation. However, in the nitrogen-limited medium (supplemented with 3 mM NH<sub>4</sub>Cl), the cell growth of all these strains was delayed at an OD<sub>600</sub> of 0.9, 8 h after inoculation. Based on this result, each of the wild-type, *rpoN*-deleted and *ntrC*-deleted strains were inoculated in nitrogen-limited Gutnick medium, and total RNA was prepared from the nitrogen replete phase (OD<sub>600</sub>=0.4) or depleted phase (OD<sub>600</sub>=0.9). The mRNA level was then measured for 17 representative target genes.

At present, the three representative promoters (*ntrC*, *potG* and *ddpX*) are known to be recognized by the RpoN sigma factor and activated by NtrC [12]. After repeated RT-qPCR analysis, the mRNA levels of *ntrC* were found to increase under nitrogen-depleted conditions (Fig. 7b, *ntrC* slot), in agreement with the proposed induction of NtrC by nitrogen depletion [12]. As expected, *ntrC* induction was not observed in the *rpoN*-defective mutant, and negligible levels of *ntrC* mRNA were detected in the *ntrC*-defective mutant. The *potG* gene is a member of the *potGHI* operon for the putrescine ABC transporter, and the *ddpX* gene forms the *ddpXABCDF* operon for D-alanyl-D-alanine dipeptidase and D-Ala-D-Ala transporter. The *potG* and *ddpX* promoters are known regulatory targets of the NtrC-dependent RpoN holoenzyme. Both promoters have been confirmed as regulatory targets

of the NtrC-dependent RpoN holoenzyme, via gSELEX screening and the gel-shift assay. To measure the expression of these two operons, the mRNA levels were determined for the first (*potG* and *ddpX*) and last (*potI* and *ddpF*) genes of each operon. The levels of *potG* and *potI* mRNA increased by 6.0- and 3.2-fold, respectively, while the *ddpX* and *ddpF* mRNA levels increased by 104.5- and 2.2-fold, respectively (Fig. 7b). Thus, we concluded that the *potGHI* and *ddpX-ABCDF* operons were markedly activated in the presence of both RpoN and NtrC.

Next, we examined the expression of newly identified target genes or operons of the NtrC-dependent RpoN holoenzyme (Fig. 7c). The *sgrST-setA* operon produces *sgrS* sRNA for the translation inhibition of *ptsG* mRNA, the SgrT inhibitor of glucose transporter PtsG and the sugar exporter SetA. mRNA for the *sgrST-setA* operon increased 8.2-fold, when detected with *sgrS*, and 11.4-fold, when detected with *setA*, in the wild-type *E. coli* K-12 strain under nitrogen depletion. We also analysed the expression of the *sgrR-sroA-thiBPQ* operon, which is located divergently from the *sgrST-setA* operon. SgrR is a transcription activator for the *sgrST-setA* operon, and *sroA* sRNA regulates the translation of ThiBPQ, a thiamine transporter. The transcription unit of *sgrR-sroA-thiBPQ* was inferred automatically computationally without experimental confirmation [50]. Under nitrogen-depleted conditions, the mRNA levels for the first gene *sgrR* increased 6.5-fold (Fig. 7c), while the level of *thiQ*, the last gene of this predicted operon, did not increase, implying that *thiQ* is not organized in this operon. The mRNA levels of the divergently organized *sgrST-setA* operon and the divergently transcribed *sgrR* gene did not increase in the *rpoN* and *ntrC* mutants. Thus, we concluded that the *sgrST-setA* operon and the divergently transcribed *sgrR* gene encoding the regulator of the *sgrST-setA* operon are under the control of the NtrC-dependent RpoN holoenzyme. The *yeaE* gene was identified as another novel target of the NtrC-dependent RpoN holoenzyme. YeaE (methylglyoxal reductase) converts methylglyoxal to hydroxyacetone. *yeaE* mRNA increased 3.1-fold under nitrogen depletion (Fig. 7c), but not in *rpoN* and *ntrC* mutants, indicating that the *yeaE* promoter was regulated by the NtrC-dependent RpoN holoenzyme.

In contrast, mRNA levels of the other seven newly identified target genes decreased in wild-type *E. coli* K-12 under nitrogen-depleted conditions (Fig. 7d). Both *mutM* mRNA and *nth* mRNA expression decreased, by 8.4-fold and 6.4-fold, respectively. MutM (DNA-formamidopyrimidine glycosylase) is a DNA glycosylase for redox-damaged purine nucleotides, whereas Nth (endonuclease III) is a DNA glycosylase and apurinic/apyrimidinic lyase for the repair of DNA damage. These proteins appear to be involved in the modulation of DNA structure and function, which is not necessary under nitrogen-depleted conditions. In the absence of RpoN sigma, however, this reduction in the mRNA levels of *mutM* and *nth* was not observed in the *rpoN*-deleted mutant (Fig. 7d). This finding supports the repressor role of the RpoN holoenzyme, irrespective of the presence or absence of NtrC.



**Fig. 7.** Influence of *rpoN* or *ntrC* on the transcription levels of the newly identified NtrC-dependent RpoN targets. Growth curve of *E. coli* wild-type BW25113, the *rpoN*-deleted mutant JW3169 and the *ntrC*-deleted mutant JW3839 in Gutnick medium supplemented with 0.2% glucose with 3 or 20 mM  $\text{NH}_4\text{Cl}$  as the sole nitrogen source (a). The mRNA levels of the known targets (b) and new targets (c, d) of the wild-type, *rpoN* mutation and *ntrC* mutation strains under nitrogen depletion. Total RNA was prepared from the wild-type, *rpoN* mutation and *ntrC* mutation strains in the exponential phase ( $\text{OD}_{600}$  0.4) and the stationary phase (20 min after growth stopped) in Gutnick medium supplemented with 0.2% glucose and 3 mM  $\text{NH}_4\text{Cl}$  and, subsequently, it was subjected to RT-qPCR analysis. RT-qPCR was repeated at least three times, and the mean values of the experiments are shown. The y-axis represents the relative level of mRNA of each NtrC-dependent RpoN target gene between the nitrogen replete and deplete conditions, with the ratio of 16S rRNA used as an internal control in each strain. ND indicates it was not detected as below the detection limit.

CspC is a member of the cold shock proteins [51] that has the ability to bind RNA and ssDNA for the modulation of DNA/RNA functions. For instance, CspC acts as a transcription anti-terminator. The *cspC* gene forms an operon with the currently uncharacterized *yobF* gene. Levels of *cspC* and *yobF* mRNA were found to decrease 6.5- and 7.4-fold, respectively, under nitrogen-depleted conditions. Reduction in the levels of *yobF* and *cspC* mRNA decreased in the *rpoN* mutant, while a negligible effect was observed in the *ntrC* mutant (Fig. 7d). The *fadL* (fatty acid uptake outer membrane channel) and *yfcZ* genes form a divergent transcription unit. The levels of *fadL* and *yfcZ* mRNA decreased 3.5- and 3.1-fold, respectively (Fig. 7d), supporting the repressor role of the RpoN holoenzyme. These results suggest that the RpoN holoenzyme functions as a repressor for a set of genes.

To confirm the repressive function of NtrC-dependent RpoN holoenzyme binding, we further carried out an *in vitro* competition assay for four targets between RpoN holoenzyme against RNAP containing the major sigma factor RpoD in the presence or absence of NtrC. At first, each of these test probes was mixed with purified RpoN holoenzyme with or without NtrC or RpoD holoenzyme, and the probe-test protein mixtures were then subjected to PAGE. The complexes of RpoN holoenzyme-probe, RpoN holoenzyme-NtrC-probe and RpoD holoenzyme-probe were observed as shift bands (Fig. S1a). Next, RpoD holoenzyme was added into the mixture in which the target probe formed complex with RpoN holoenzyme under the presence or absence of NtrC, and then subjected to PAGE. The probe pattern was similar to that in the presence of RpoN holoenzyme with or without NtrC (Fig.



S1a). To confirm the binding of each RNAP holoenzyme, we then performed Western blot analysis using antibodies against RpoN and RpoD. Using the RpoN antibody, the signal was observed for RpoN holoenzyme and the intensity was higher in the presence of NtrC, which suggests the binding of RpoN holoenzyme to the target promoter was supported by NtrC (Fig. S1b). In contrast, the signal of RpoD was detected in RpoD holoenzyme alone, but the intensity became low under the presence of RpoN holoenzyme, and almost disappeared with addition of NtrC. These results suggest that the binding of NtrC supported RpoN holoenzyme competes against RpoD holoenzyme bindings due to their ability to repress these promoters. Thus, we propose naming the group of promoters recognized by the RpoN holoenzyme alone for the repression of the target genes as the group of repressive promoters. The RpoN holoenzyme acts as a repressor for the transcription of a group of target genes.

### Physiological roles of the RpoN promoters: a class of repressive promoters

We previously identified a set of constitutive promoters recognized by RNAP alone containing each of the major sigma factors of RpoD and the minor sigma factors (stationary-phase sigma RpoS, heat-shock sigma RpoH, flagella-chemotaxis sigma RpoF and extra-cytoplasmid sigma RpoE) using the gSELEX system [26, 27]. Constitutive promoters are recognized and transcribed by each of these RNAP holoenzymes alone in the absence of supporting regulatory proteins. In the case of RpoN, however, the RNAP RpoN holoenzyme recognizes and binds to a set of promoters, detected here using gSELEX screening. For transcription initiation, TFs with an enhancing role have been proposed, including those in the NtrC family (eight species in *E. coli*) or TyrR family (four species) [6, 16]. The experimental confirmation of this enhancer role has only been obtained with a few enhancers, including NtrC [8] and PspF [52]. In this study, we identified the whole set of binding sites for the RpoN holoenzyme in the presence and absence of the enhancer NtrC.

A total of 71 binding sites were identified by the RpoN holoenzyme alone, of which 23 (32%) are listed in RegulonDB database (Table S2). In the presence of NtrC, a total of 108 binding sites were identified by the RpoN holoenzyme in the presence of NtrC, of which 28 (26%) binding sites are listed in RegulonDB [50]. In RegulonDB, a total of 101 targets were listed as RpoN-dependent promoters; however, only seven targets (*acrD*, *actP*, *aslB*, *astC*, *chaC*, *ddpX* and *emrD*) have been experimentally confirmed. All these targets were identified using *in vitro* gSELEX screening, indicating the degree of reliability of gSELEX screening. The majority of the other targets listed in RegulonDB were predicted by computational approaches relying on consensus sequences obtained from a small number of known promoters.

RpoN regulates not only genes involved in nitrogen metabolism but also other cellular functions, including metabolic pathways, dependent on different enhancers, such as formate catabolism (TyrR family; FhlA-dependent), acetoacetate

catabolism (NtrC family; AtoC-dependent), propionate catabolism (TyrR family; PrpR-dependent), phage shock response (NtrC family; PspF-dependent) and zinc response (NtrC family; ZraR-dependent) [6]. We further identified a novel set of genes under the control of NtrC-dependent RpoN holoenzyme involved in carbon source metabolism, such as the *sgrST-setA* operon (inhibition of PtsG glucose transporter and sugar efflux pump) and the *sgrR* gene (activation of the *sgrST-setA* operon) [53, 54], both activated by NtrC. The activation of glucose transporter inhibitors and sugar efflux pumps, together with the repression of a long-chain fatty acid transporter, may lead to a carbon/nitrogen imbalance by decreasing the carbon source influx and increasing the carbon source efflux. Upon accumulation, 2-oxoglutamate, the key metabolite of the intersection between carbon and nitrogen metabolism, binds to GlnB (nitrogen regulatory protein PII) and activates NtrB sensor kinase, which phosphorylates NtrC, thereby activating a number of NtrC-dependent genes. The expression of the *ntrBC* (*glnLG*) operon is auto-regulated by NtrB-phosphorylated NtrC, depending on nitrogen availability.

Several promoters recognized by the RpoN holoenzyme alone were identified as repressive promoters. The occupation of repressive promoters by the RpoN holoenzyme was predicted to inhibit the expression of the genes located downstream. In the absence of enhancers under steady-state stressless conditions, the repressive promoters must be occupied by the RNAP RpoN holoenzyme, to prevent transcription by nearby promoters that are recognized by the holoenzymes containing sigma factors other than RpoN. In fact, the binding of the NtrC-enhanced RpoN holoenzyme to the targets interfered with the binding to the promoter of the RpoD holoenzyme (Fig. S1). Promoter sequences recognized by other sigma factors overlapped with the newly identified repressive promoters, such as the RpoD promoter on the *yobF-cspC* promoter, and RpoH and RpoD promoters on the *mutM* promoter (listed in Regulon DB). Our observations of the repressive promoters are in good agreement with the proposed model of the repressor role for RpoN in transcription of some genes [18]. In *E. coli* K-12 cells grown in rich media (e.g. LB broth), three species of the sigma subunit are present: (i) RpoD, the most abundant and responsible for the transcription of growth-related genes; (ii) RpoN, the second most abundant; (iii) RpoF, responsible for the transcription of flagella-chemotaxis gene [36, 37]. To date, the mechanism underlying the presence of high levels of RpoN in growing *E. coli* cells has yet to be elucidated, even though nitrogen is not limiting. One possibility is that the RpoN holoenzyme exerts a repressor role to prevent the transcription of a certain group of genes by the most abundant RpoD holoenzyme. The presence of the RpoN holoenzyme on repressive promoters allows transcription to initiate quickly once enhancer proteins are induced under the corresponding stress conditions, such as NtrC activation under nitrogen depletion.

In this study, we identified the whole set of binding sites of RpoN holoenzyme, the enhancer NtrC and NtrC-dependent

RpoN holoenzyme using *in vitro* gSELEX-chip screening in the absence of other regulators. The identification of a set of promoters recognized by the RpoN holoenzyme in the presence and absence of NtrC enhancer, including the repressive promoters, provides insight into the regulation of the bacterial genome.

#### Funding information

This work was supported by MEXT Grants-in-Aid for Scientific Research (C) (19K06618) to T.S., (B) (18310133) and (C) (25430173) to A.I., and MEXT-Supported Programme for the Strategic Research Foundation at Private University to A.I.

#### Acknowledgements

We thank Nobuyuki Fujita, Ayako Kori and Kayoko Yamada for the expression and purification of RNAP proteins and NtrC, as well as the National BioResource Project, National Institute of Genetics, Japan, for providing *E. coli* K-12 BW25113 and single-gene deletion mutants [JW3169 ( $\Delta rpoN$ ) and JW3839 ( $\Delta ntrC$ )].

#### Author contributions

Conceptualization, T.S. and A.I.; methodology, T.S., S.F. and A.I.; formal analysis, T.S., S.F. and A.I.; investigation, T.S. and S.F.; resources, A.I. and T.S.; writing – original draft preparation, T.S. and A.I.; writing – review and editing, A.I. and T.S.; funding acquisition, T.S. and A.I.

#### Conflicts of interest

The authors declare that there are no conflicts of interest.

#### References

- Ishihama A. Molecular assembly and functional modulation of *Escherichia coli* RNA polymerase. *Adv Biophys* 1990;26:19–31.
- Ishihama A. Functional modulation of *Escherichia coli* RNA polymerase. *Annu Rev Microbiol* 2000;54:499–518.
- Hirschman J, Wong PK, Sei K, Keener J, Kustu S. Products of nitrogen regulatory genes ntrA and ntrC of enteric bacteria activate glnA transcription in vitro: evidence that the ntrA product is a sigma factor. *Proc Natl Acad Sci USA* 1985;82:7525–7529.
- Kustu SG, Santero E, Keener J, Popham D, Weiss D. Expression of sigma 54 (ntrA)-dependent genes is probably united by a common mechanism. *Microbiol Rev* 1989;53:367–376.
- Studholme DJ, Buck M. The biology of enhancer-dependent transcriptional regulation in bacteria: insights from genome sequences. *FEMS Microbiol Lett* 2000;186:1–9.
- Reitzer L, Schneider BL. Metabolic context and possible physiological themes of  $\sigma$ 54-dependent genes in *Escherichia coli*. *Microbiol Mol Biol Rev* 2001;65:422–444.
- Reitzer LJ. Sources of nitrogen and their utilization. In: *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd edn. Washington, DC: American Society for Microbiology, 1996. pp. 380–390.
- Weiss V, Claverie-Martin F, Magasanik B. Phosphorylation of nitrogen regulator I of *Escherichia coli* induces strong cooperative binding to DNA essential for activation of transcription. *Proc Natl Acad Sci USA* 1992;89:5088–5092.
- Reitzer LJ, Magasanik B. Transcription of glnA in *E. coli* is stimulated by activator bound to sites far from the promoter. *Cell* 1986;45:785–792.
- Wigneshweraraj SR, Chaney MK, Ishihama A, Buck M. Regulatory sequences in sigma 54 localise near the start of DNA melting. *J Mol Biol* 2001;306:681–701.
- Carlo SD, Chen B, Hoover TR, Kondrashkina E, Nogales E, et al. The structural basis for regulated assembly and function of the transcriptional activator NtrC. *Genes Dev* 2006;20:1485–1495.
- Zimmer DP, Soupene E, Lee HL, Wendisch VF, Khodursky AB, et al. Nitrogen regulatory protein C-controlled genes of *Escherichia coli*: scavenging as a defense against nitrogen limitation. *Proc Natl Acad Sci USA* 2000;97:14674–14679.
- Buck M, Miller S, Drummond M, Dixon R. Upstream activator sequences are present in the promoters of nitrogen fixation genes. *Nature* 1986;320:374–378.
- Buck M, Gallegos MT, Studholme DJ, Guo Y, Gralla JD. The bacterial enhancer-dependent sigma(54) (sigma(N)) transcription factor. *J Bacteriol* 2000;182:4129–4136.
- Danson AE, Jovanovic M, Buck M, Zhang X. Mechanisms of  $\sigma$ 54-dependent transcription initiation and regulation. *J Mol Biol* 2019;431:3960–3974.
- Ishihama A, Shimada T, Yamazaki Y. Transcription profile of *Escherichia coli*: genomic SELEX search for regulatory targets of transcription factors. *Nucleic Acids Res* 2016;44:2058–2074.
- Bonocora RP, Smith C, Lapierre P, Wade JT. Genome-scale mapping of *Escherichia coli*  $\sigma$ 54 reveals widespread, conserved intragenic binding. *PLoS Genet* 2015;11:e1005552.
- Schaefer J, Engl C, Zhang N, Lawton E, Buck M. Genome wide interactions of wild-type and activator bypass forms of  $\sigma$ 54. *Nucleic Acids Res* 2015;43:7280–7291.
- Ishihama A. Prokaryotic genome regulation: multifactor promoters, multitarget regulators and hierarchic networks. *FEMS Microbiol Rev* 2010;34:628–645.
- Ishihama A. Prokaryotic genome regulation: a revolutionary paradigm. *Proc Jpn Acad Ser B* 2012;88:485–508.
- Gyaneshwar P, Paliy O, McAuliffe J, Popham DL, Jordan MI, et al. Sulfur and nitrogen limitation in *Escherichia coli* K-12: specific homeostatic responses. *J Bacteriol* 2005;187:1074–1090.
- Brown DR, Barton G, Pan Z, Buck M, Wigneshweraraj S. Nitrogen stress response are coupled in *Escherichia coli*. *Nat Commun* 2014;5:4115.
- Shimada T, Fujita N, Maeda M, Ishihama A. Systematic search for the Cra-binding promoters using genomic SELEX system. *Genes Cells* 2005;10:907–918.
- Shimada T, Ogasawara H, Ishihama A. Genomic SELEX screening of regulatory targets of *Escherichia coli* transcription factors. *Meth Mol Biol* 2018;1837:49–69.
- Shimada T, Ogasawara H, Ishihama A. Single-target regulators form a minor group of transcription factors in *Escherichia coli* K-12. *Nucleic Acids Res* 2018;46:3921–3936.
- Shimada T, Yamazaki Y, Tanaka K, Ishihama A. The whole set of constitutive promoters recognized by RNA polymerase RpoD holoenzyme of *Escherichia coli*. *PLoS One* 2014;9:e90447.
- Shimada T, Tanaka K, Ishihama A. The whole set of the constitutive promoters recognized by four minor sigma subunits of *Escherichia coli* RNA polymerase. *PLoS One* 2017;12:e0179181.
- Ishihama A. Building a complete image of genome regulation in the model organism *Escherichia coli*. *J Gen Appl Microbiol* 2018;63:311–324.
- Jishage M, Ishihama A. Variation in RNA polymerase sigma subunit composition within different stocks of *Escherichia coli* W3110. *J Bacteriol* 1997;179:959–963.
- Maeda H, Ishihama A. Competition among seven *Escherichia coli* sigma subunits: relative binding affinity to the core RNA polymerase. *Nucleic Acids Res* 2000;28:3467–3503.
- Igarashi K, Ishihama A. Bipartite functional map of the *E. coli* RNA polymerase sigma subunit: involvement of the C-terminal region in transcription activation by cAMP-CRP. *Cell* 1991;65:1015–1022.
- Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* 2000;97:6640–6645.
- Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, et al. Construction of *Escherichia coli* in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2006;2:2006.0008.
- Gutnick D, Calvo JM, Klopotoski T, Ames BN. Compounds which serve as the source of carbon or nitrogen for *Salmonella typhimurium* LT-2. *J Bacteriol* 1969;1:215–219.

35. Fujita N, Ishihama A. Reconstitution of RNA polymerase. *Meth Enzymol* 1996;273:121–130.
36. Jishage M, Ishihama A. Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: intracellular levels of  $\sigma^{70}$  and  $\sigma^{32}$ . *J Bacteriol* 1995;177:6832–6835.
37. Jishage M, Iwata A, Ueda S, Ishihama A. Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: intracellular levels of four species of sigma subunit under various growth conditions. *J Bacteriol* 1996;178:5447–5451.
38. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 2009;37:W202–W208.
39. Shimada T, Hirao K, Kori A, Yamamoto K, Ishihama A. RutR is the uracil/thymine-sensing master regulator of a set of genes for synthesis and degradation of pyrimidines. *Mol Microbiol* 2007;66:744–757.
40. Anzai T, Imamura S, Ishihama A, Shimada T. Expanded roles of pyruvate-sensing PdhR in transcription regulation of the *Escherichia coli* genome: fatty acid catabolism and cell motility. *Microb Genom* 2020;6:mgen000442.
41. Singer BS, Shtatland T, Brown D, Gold L. Libraries for genomic SELEX. *Nucleic Acids Res* 1997;25:781–786.
42. Ellington AD, Szostak JW. In vitro selection of DNA molecules that bind specific ligands. *Nature* 1990;346:818–822.
43. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 1990;249:505–510.
44. Shimada T, Ishihama A, Busby SJW, Grainger DC. The *Escherichia coli* RutR transcription factor binds at targets within genes as well as intergenic regions. *Nucleic Acids Res* 2008;36:3950–3955.
45. Yamamoto K, Hirao K, Oshima T, Aiba H, Utsumi R, et al. Functional characterization *in vitro* of all two-component signal transduction systems from *Escherichia coli*. *J Biol Chem* 2005;280:1448–1456.
46. Reitzer LJ, Magasanik B. Isolation of the nitrogen assimilation regulator NRI the product of the glnG gene of *Escherichia coli*. *Proc Natl Acad Sci USA* 1983;80:5554–5558.
47. Reitzer L. Nitrogen assimilation and global regulation in *Escherichia coli*. *Annu Rev Microbiol* 2003;57:155–176.
48. Wigneshweraraj SR, Ishihama A, Buck M. In vitro roles of invariant helix-turn-helix motif residue R383 in  $\sigma^{54}$  ( $\sigma^N$ ). *Nucleic Acids Res* 2001;29:1163–1174.
49. Burrows PC, Severinov K, Ishihama A, Buck M, Wigneshweraraj SR. Mapping  $\sigma^{54}$ -RNA polymerase interactions at the -24 consensus promoter element. *J Biol Chem* 2003;278:29728–29743.
50. Santos-Zavaleta A, Salgado H, Gama-Castro S, Sanchez-Perez M, Gomez-Romero L, et al. RegulonDB v10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. *Nucleic Acids Res* 2019;47:D212–D220.
51. Lee SJ, Xie A, Jiang W, Etchegaray JP, Jones PG, et al. Family of the major cold-shock protein, CspA (CS7.4), of *Escherichia coli*, whose members show a high sequence similarity with the eukaryotic Y-box binding proteins. *Mol Microbiol* 1994;11:833–839.
52. Jovanovic G, Dworkin J, Model P. Autogenous control of PspF, a constitutively active enhancer-binding protein of *Escherichia coli*. *J Bacteriol* 1997;179:5232–5237.
53. Vanderpool CK, Gottesman S. Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. *Mol Microbiol* 2004;54:1076–1089.
54. Raina M, Storz G. SgrT, a small protein that packs a sweet punch. *J Bacteriol* 2017;199:e00130–17.

### Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at [microbiologyresearch.org](https://microbiologyresearch.org).