

Redefining *Escherichia coli* σ^{70} Promoter Elements: -15 Motif as a Complement of the -10 Motif[†]

Marko Djordjevic*

Institute of Physiology and Biochemistry, Faculty of Biology, The University of Belgrade, 11000 Belgrade, Serbia

Received 4 August 2011/Accepted 30 August 2011

Classical elements of σ^{70} bacterial promoters include the -35 element ($^{-35}\text{TTGACA}^{-30}$), the -10 element ($^{-12}\text{TATAAT}^{-7}$), and the extended -10 element ($^{-15}\text{TG}^{-14}$). Although the -35 element, the extended -10 element, and the upstream-most base in the -10 element (^{-12}T) interact with σ^{70} in double-stranded DNA (dsDNA) form, the downstream bases in the -10 motif ($^{-11}\text{ATAAT}^{-7}$) are responsible for σ^{70} -single-stranded DNA (ssDNA) interactions. In order to directly reflect this correspondence, an extension of the extended -10 element to a so-called -15 element ($^{-15}\text{TGnT}^{-12}$) has been recently proposed. I investigated here the sequence specificity of the proposed -15 element and its relationship to other promoter elements. I found a previously undetected significant conservation of ^{-13}G and a high degeneracy at ^{-15}T . I therefore defined the -15 element as a degenerate motif, which, together with the conserved stretch of sequence between -15 and -12 , allows treating this element analogously to -35 and -10 elements. Furthermore, the strength of the -15 element inversely correlates with the strengths of the -35 element and -10 element, whereas no such complementation between other promoter elements was found. Despite the direct involvement of -15 element in σ^{70} -dsDNA interactions, I found a significantly stronger tendency of this element to complement weak -10 elements that are involved in σ^{70} -ssDNA interactions. This finding is in contrast to the established view, according to which the -15 element provides a sufficient number of σ^{70} -dsDNA interactions, and suggests that the main parameter determining a functional promoter is the overall promoter strength.

Bacterial RNA polymerase is a central enzyme in cells, and initiation of transcription by bacterial RNA polymerase is a major point in gene expression regulation. Core RNA polymerase cannot by itself initiate transcription, so a complex between RNA polymerase core and a σ factor (called RNA polymerase holoenzyme) is formed, which is abbreviated as RNAP here for simplicity (10). Different σ factors interact with double-stranded DNA (dsDNA) and single-stranded DNA (ssDNA) in a sequence-specific manner and are responsible for transcription under different conditions (2). The present study concentrates on σ^{70} , which is the major σ factor in *Escherichia coli*, that is responsible for transcription of house-keeping genes (39).

Transcription initiation begins with RNAP binding to dsDNA, which is referred to as the closed complex formation (7). Subsequent to RNAP binding, the two strands of DNA are separated through thermal fluctuations that are facilitated by interactions of RNAP with ssDNA (8). The opening of two DNA strands results in the formation of an ~ 15 -bp transcription bubble, which typically extends from -11 to $+3$ (where $+1$ corresponds to the transcription start site) (3). After the open complex is formed, RNAP clears the promoter and enters the elongation, which leads to synthesis of RNA from DNA template (2).

The main elements that determine promoter recognition are

the -35 element ($^{-35}\text{TTGACA}^{-30}$, where the coordinates in the superscript are relative to the transcription start site), the -10 element ($^{-12}\text{TATAAT}^{-7}$), the spacer between these two elements, and the extended -10 element ($^{-15}\text{TG}^{-14}$) (22). The spacer ranges from 15 to 19 bp, with the most optimal value being 17 bp (41). Interactions of σ^{70} with dsDNA of the -35 element, the extended -10 element, and the -12 base of the -10 element result in the closed complex formation (35). On the other hand, the downstream bases of the -10 element (-11 to -7) interact with σ^{70} in ssDNA form (35) and are directly involved in the open complex formation.

In order to better relate the involvement of different promoter elements with the kinetic steps of transcription initiation (the closed and the open complex formation), it was recently proposed that the region from -15 to -7 be reorganized in the following way (22): region from -15 to -12 is connected in a new element that is defined as the -15 element. This element includes the extended -10 element, the most upstream base in the -10 element (base -12), and base -13 that is in between. Consequently, the -10 element is shortened for one base pair (to the region -11 to -7), which I here refer to as the short -10 element. In this way, the -35 and -15 elements are directly related to σ^{70} -dsDNA interactions, whereas the short -10 element is directly related to σ^{70} -ssDNA interactions.

This reorganization of the promoter elements has been proposed based entirely on biochemical arguments, i.e., on the interactions of the bases with relevant σ^{70} domains and on their involvement in closed and open complex formations. However, at the sequence level, the -10 motif and the extended -10 motif appear to be physically separated due to the absence of recognized conservation at position -13 ; note that the -15 element is currently defined as $^{-15}\text{TGnT}^{-12}$ (22).

* Mailing address: Institute of Physiology and Biochemistry, Faculty of Biology, University of Belgrade, Studentski trg 16, 11000 Belgrade, Serbia. Phone: (381 64) 234-8951. Fax: (381 11) 263-9882. E-mail: dmako@bio.bg.ac.rs.

† Supplemental material for this article may be found at <http://jba.asm.org/>.

‡ Published ahead of print on 9 September 2011.

Therefore, detecting conservation at position -13 is desirable, since it would define a continuous conserved stretch of sequence corresponding to the -15 element.

A related question concerns the sequence specificity of the introduced -15 element. That is, the specificity of the extended -10 element is currently presented in a “binary” manner: as either the presence or absence of $^{-15}\text{TG}^{-14}$ one base upstream of the -10 element (see, for example, reference 32). Consequently, only a small fraction ($\sim 20\%$) of promoters is recognized as “extended -10 ” (32), while for other promoters the possible contribution of bases -15 and -14 to transcription initiation is not taken into account. This is in contrast to the -35 and -10 elements, which are recognized as highly degenerate (28), where mismatches from the promoter sequence allow graded decreases in promoter strength. This high degeneracy is quantitatively represented by appropriate weight matrices (46). It would therefore be desirable to find an equivalent description for the -15 element as well, which would allow treatment of all promoter elements in a unified manner and direct implementation of this description in promoter searches.

Another question is how the strength of the -15 element relates to the strengths of the -10 and -35 elements. The relationships between the strengths of the promoter elements may indicate a role that they play in promoter recognition. A classic view of the extended -10 element is that it supplements for the weak -35 element in order to allow sufficient interaction strength with dsDNA for the closed complex formation (24, 29, 32, 40, 53). It seems plausible to extrapolate this classic view to the -15 element, particularly since this element is defined so as to directly relate to σ^{70} -dsDNA interactions. On the other hand, it is not evident whether some minimal number of contacts with dsDNA is needed *per se* for promoter function, since the open complex can form (although very slowly) even in the absence of recognizable -35 and extended -10 elements (37). Therefore, investigating relationship of the -15 element to the strengths of the other promoter elements may provide clues about the physical mechanism of promoter recognition in bacteria.

I first investigated the sequence specificity and conservation of bases within and immediately upstream of the -15 element. To achieve this, I sought to “*de novo*” align experimentally established σ^{70} promoters. This alignment would furthermore allow accurate inference of the weight matrices for the promoter elements, which will provide much more accurate estimates of the motif strengths compared to mismatches to the consensus sequence. The inferred weight matrices will then be used to investigate relationships between the -15 element, the other promoter elements, and the overall promoter strength. The relevance of these findings to the mechanism of promoter recognition, a newly established mix-and-match model of promoter recognition (22), and bioinformatic search of promoters are discussed.

MATERIALS AND METHODS

-10 motif alignment. To perform unsupervised motif alignment, I used a Gibbs sampler (48). The Gibbs sampler implements a version of the Gibbs search algorithm (25), which is used to find mutually similar motifs in a given set of DNA sequences. Only the DNA strand defined by the direction of transcription was searched, since both the -10 box motifs and the -35 box are not palindrome

symmetric. The search was done with the initial assumption that one motif element is present in each DNA segment; however, in the end of the Gibbs sampler search, individual motif elements are added in or taken out, in a single pass of the algorithm, depending upon whether or not their inclusion improves the value of the alignment score. The last step allows excluding from the alignment sequences that do not contain -10 box motifs, e.g., due to database mis-assignments. The search resulted in the identification of 322 aligned -10 boxes, which were used in the further analysis. The alignment of the -10 box automatically yielded the alignment of the -15 box, since this element consists of the upstream-most base in the -10 element and three more bases upstream of the -10 element.

-35 motif alignment. The aligned -10 elements were used as an anchor to align the -35 elements. To identify -35 elements, I selected sequences that span 16 to 25 bp from the upstream most base in -10 elements, which is consistent with the length of the -35 element (6 bp) and possible spacer lengths from 15 to 19 bp (41). I next used a Gibbs search (48) to locate 6-bp motifs within these 10-bp sequences. I initially assumed one motif within each sequence, while in the last pass of the Gibbs search individual elements were added in or taken out, depending on whether or not it improved the alignment score. Finally, I used the aligned motifs to build a weight matrix for the -35 element and searched each 10-bp sequence with this matrix to ensure that I located the highest-scoring motif within each sequence. These highest-scoring motifs composed the final alignment of -35 motifs. The alignment of the promoter elements is given in File S1 in the supplemental material.

Construction of weight matrices. I started from a collection of aligned sequences. For each position in the alignment, I determined the frequency with which each of the four bases occurs. I also determined the background base frequency by sampling the frequency at which each of the four bases occurs in *E. coli* intergenic sequences; background frequencies were sampled from intergenic sequences, since transcription start sites are located within them. Weight matrix elements $w_{i,\alpha}$ that correspond to base α present at position i in the motif are calculated as shown in the following equation (20):

$$w_{i,\alpha} = \log\left(\frac{nv_{i,\alpha} + p_\alpha}{p_\alpha(n + 1)}\right),$$

where n is the total number of motifs in the alignment from which the weight matrix is inferred, $v_{i,\alpha}$ is number of times that base α appears at position i divided by n , and p_α is the background frequency of base α . One should note that the addition of p_α in the numerator of the logarithm corresponds to so-called pseudocounts, which become important for small data sets. Note that for large n (as is the case for the data set used here) the above calculation approximately reduces to the log ratio of base frequency and background frequency.

A similar expression is used for weights corresponding to different spacer lengths: $w_i = \log(v_i/0.2)$, where w_i is the weight corresponding to the spacer of length i ($i \in \{15, \dots, 19\}$), while equiprobable background frequencies (0.2) were taken. All of the weight matrices are included in File S2 in the supplemental material, while the spacer weights are shown in Fig. 1B.

Finally, I used the property that one can add (or subtract) a provisional constant from each column of the weight matrix (see, for example, references 9 and 42), which corresponds to shifting zero of the weight matrix scores. Accordingly, I subtracted from each column of the weight matrix a constant equal to the weight matrix element corresponding to the consensus base at a given position. This choice is convenient for easier reference (see, for example, reference 15), since the weight matrix score of the consensus sequence then corresponds to zero.

Significance of base overrepresentations. To assess base overrepresentation in the regions corresponding to -35 and -15 element, I first calculated the corresponding frequency matrices. Entries in the frequency matrix are given by the following equation:

$$f_{i,\alpha} = \frac{nv_{i,\alpha} + p_\alpha}{(n + 1)p_\alpha},$$

where the addition of p_α in the numerator corresponds to pseudocounts (see above). Note that within the limit of a large n (which approximately holds in this case), the table entries $f_{i,\alpha}$ correspond to the ratio between the base frequency and the background frequency: $v_{i,\alpha}/p_\alpha$.

To assess whether a given base at a given position in the motif is significantly overrepresented over background, I calculated the standard deviation of $f_{i,\alpha}$ ($\delta f_{i,\alpha}$) by assuming a Poisson distribution. The standard deviation was estimated as follows:

$$\delta f_{i,\alpha} = \frac{\sqrt{nv_{i,\alpha} + p_\alpha}}{(n + 1)p_\alpha}$$

Finally, the bases for which frequencies in the alignment are larger than background probabilities outside of the 95% confidence intervals (i.e., for which $f_{i,\alpha} - 1$ is larger than $1.96\delta f_{i,\alpha}$) are marked as being overrepresented.

Specificity at a given position in motif. To estimate the specificity (level of conservation) for a given position in a motif, I started from the weight matrix (see above) for this motif. The specificity at position i in the motif (s_i) is then calculated as follows (9):

$$s_i = \sum_{\alpha=1}^4 (w_{i,\alpha} - \bar{w}_i)^2,$$

$$\text{where } \bar{w}_i = \sum_{\alpha=1}^4 w_{i,\alpha}.$$

Note that when no base is overrepresented, all weight matrix elements at position i ($w_{i,\alpha}$) become mutually equal, and equal to their mean \bar{w}_i , so that the specificity s_i becomes zero.

Correlation coefficients and their significance. Correlation coefficients were determined by using a MATLAB (Mathworks) routine. The same MATLAB function allows calculating P values of the obtained correlation coefficients. Briefly, the routine is based on randomly permuting the points in the data set. The correlation coefficient for each random permutation is calculated, and the statistical significance of the difference between the original correlation coefficient and the correlation coefficients in the permuted data set is estimated by using a Student t test.

Difference between the means of the score distributions. To assess the P value for the difference between the means of the score distributions, I used a Student t test implemented through the appropriate MATLAB (Mathworks) routine.

RESULTS

Promoter alignment. Accurate alignment of promoter elements is a nontrivial bioinformatic task, which is largely complicated by weaker conservation of the -35 element and its variable distance to the -10 element (23). Consequently, most studies up to now either did not align -35 element or are based on earlier models of promoter specificity (23, 32, 41, 51). An exception to this is the alignment determined by Shultzaberger et al. (45), which, however, exhibits notable qualitative differences compared to the -35 element alignment presented here (this is discussed below). In addition, it is nontrivial to produce an alignment with sufficient accuracy for analyzing the -15 element, given a weaker conservation of this element compared to both -10 and -35 elements. For example, the most recent comprehensive alignment of promoter elements (45) did not detect conservation at the -15 and -14 positions, despite clear evidence of a contribution of this region to transcription initiation and its involvement in interaction with RNA polymerase (22).

To evade biases in alignment, I started here directly from experimentally determined transcription start sites in the genome (14). As described in Materials and Methods, I used a Gibbs search algorithm for unsupervised alignment of promoter elements, which in the end was improved through supervised searching by weight matrices defined through the Gibbs algorithm. The approach was to first align the -10 element and to consequently use this element as an anchor to align the -35 element. Alignment of other relevant elements (spacer and -15 element) is directly determined once the -10 element and the -35 element are aligned.

To align -10 elements, I used the assembly of transcription

TABLE 1. Significance of base overrepresentation for the -35 region^a

Base	Ratio of base frequencies at position:							
	-37	-36	-35	-34	-33	-32	-31	-30
A	1.00	0.99	0.78	0.27	0.94	1.36	1.29	1.32
T	1.14	0.75	2.14	2.51	0.55	0.80	0.58	0.96
C	0.76	1.12	0.49	0.90	0	1.14	1.80	0.62
G	1.05	1.21	0.30	0.06	2.68	0.65	0.37	1.01

^a The columns in the table correspond to the six bases in the -35 region, including two additional upstream bases. The four rows in the table correspond to four bases. For each base at each position, the ratio of the frequency of base appearance in the alignment and the background base frequency was calculated. Values for bases that are significantly overrepresented over background (see Materials and Methods) are marked in boldface.

start sites from RegulonDB database (14). This assembly includes both experimentally verified promoters and computational predictions and corresponds to both σ^{70} and alternative σ factors. For this alignment I selected only experimentally verified σ^{70} transcription start sites, i.e., I disregarded all transcription start sites that are either not experimentally validated or corresponded to alternative σ factors. This selection results in a total of 342 σ^{70} transcription start sites, and I used the obtained start sites in order to extract DNA segments that corresponded to positions -17 to -2 relative to the transcription start sites. These positions were chosen bearing in mind that the position of the -10 element can deviate for 5 bp relative to its canonical position (-12 to -7) (17).

To identify the 6-bp -10 elements within the selected DNA segments, I used the Gibbs sampler (25, 48). The algorithm allowed me to perform an unsupervised search, i.e., I used no prior information on the sequence specificity of the -10 box (see Materials and Methods). Some of the initial 342 segments were found not to contain a recognizable -10 box (possibly due to database mis-assignments); consequently, the search resulted in the identification of 322 aligned -10 boxes, which were used in further analysis.

To identify -35 elements, I started from the aligned -10 elements and selected DNA segments that correspond to range from 16 to 25 bp from the upstream-most base in the aligned -10 box. This range is based on the fact that the -35 element is 6 bp long and the spacer length between the -35 and -10 elements is 15 to 19 bp. I again used the Gibbs sampler to search for 6-bp overrepresented motifs within these segments. The search resulted in a motif with the consensus sequence GTTGAC; this motif is evidently shifted for 1 bp relative to the established consensus of -35 element (TTGACA). This shift is not surprising given that (i) the downstream-most base of the -35 element shows relatively low conservation (Table 1, which shows the significance of base overrepresentation for the -35 region), (ii) there is a fairly good conservation of the base-pair immediately upstream of -35 element (see Table 1), and (iii) it is common that a Gibbs search results in motifs that are shifted relative to their optimal alignment (25).

I therefore manually shifted the alignment obtained by Gibbs search for 1 bp, so that it coincided with the established consensus, and constructed a weight matrix for such a realigned -35 motif. To ensure that the optimal alignment is

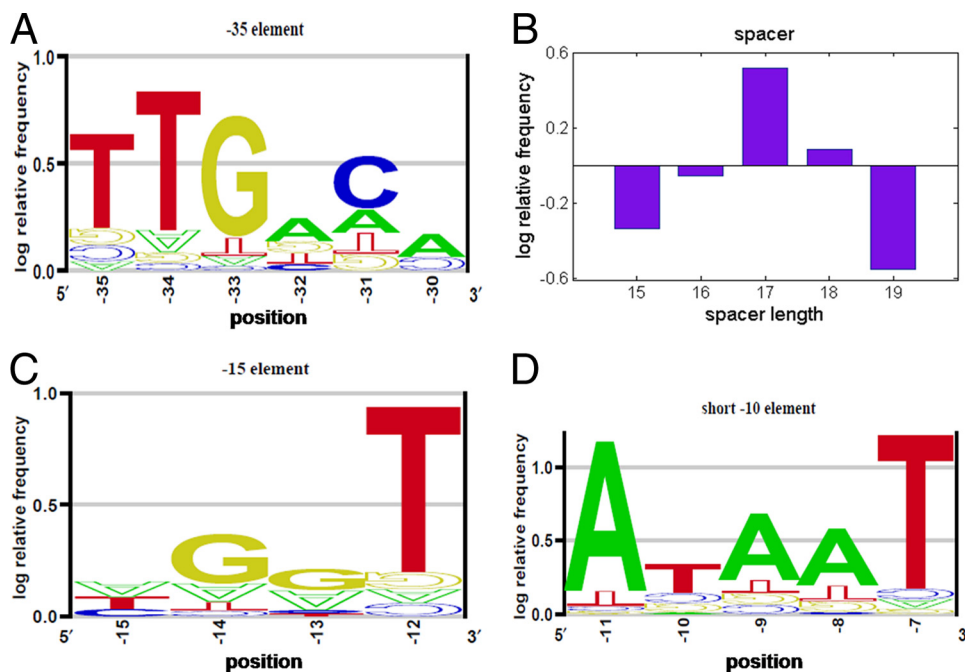


FIG. 1. Specificity of promoter elements. Sequence logos were generated by enoLOGOS (52) and correspond to the specificities of the -35 element (A), the -15 element (C), and the short -10 element (D). The heights of the bases are proportional to the logarithms of the ratio of base frequencies in the alignment and background base frequencies. Since the logarithm of the ratio is taken, for bases that are underrepresented relative to the background frequency the log ratio takes negative value. This is graphically represented as “upside-down” bases. (B) For spacers, log ratio of the frequency with which each spacer length appears in the population of promoters and an equiprobable background distribution (0.2 for each of the five spacer lengths) is shown.

indeed selected, I identified the motif with the highest weight matrix score on each of the original segments. The motifs are thus the final alignment for the -35 elements. Once -10 element and -35 element were aligned, it was straightforward to sample the distribution of the spacer lengths. Similarly, once the -10 element was aligned, the -15 element spans from 3 bases upstream of the -10 element to the upstream-most base of the -10 element.

Weight matrices that correspond to different promoter elements can be inferred directly from the obtained alignment (see Materials and Methods). The weight matrices allow a much more accurate measure of promoter element strength compared to the number of mismatches to the consensus sequence (which are often used to estimate promoter strength) (46, 47). This is because different positions within the element, as well as different base substitutions at a given position, can have a very different effect on promoter transcription activity. For example, while positions -11 and -7 within the -10 element are crucial for transcription activity, position -10 within the same element has much less importance (13, 27). These effects cannot be taken into account by counting mismatches to the consensus sequence but are accounted for by weight matrices. Weight matrices for all of the promoter elements obtained in our alignment are explicitly given in File S2 in the supplemental material and are used in the analysis below.

Specificity of the -35 element and the short -10 element.

The specificities of the aligned promoter elements are shown in Fig. 1. Note that, instead of the -10 element, the short -10 element is shown due to the reorganization of the promoter

elements discussed above. The figure, generated by EnoLogos (52), shows the ratio of the base frequencies in the alignment, relative to the background base frequencies. Note that the logarithm of the ratio is taken, so that if a base is underrepresented relative to the background frequency, the log ratio takes a negative value. Negative log ratios are graphically represented as “upside-down” bases. For spacers (Fig. 1B), log ratios are also presented, where the background distribution is equiprobable (i.e., 0.2 for each of the five spacer variants).

Table 1 shows an overrepresentation over background of six bases that belong to the -35 element (positions -35 to -30), including two additional bases upstream of this element (positions -37 and -36). The overrepresentation of -35 element bases obtained from this alignment (Table 1 and Fig. 1) is consistent with the available data on interactions between σ^{70} and -35 element (6): the largest overrepresentation was obtained for bases -35 , -34 , -33 , and -31 , which are bound to σ subunit residues with hydrogen bonds; the overrepresentation is notably smaller for bases -32 and -30 , which interact with σ with weaker van der Waals interactions. Finally, there was a statistically significant overrepresentation of G at position -36 (see Table 1). This might seem unexpected, since position -36 is not part of the -35 element; however, this conservation is consistent with the interaction data that indicate van der Waals interactions between the -36 and σ^{70} residues (6). I included one additional upstream base in Table 1 (position -37), where there was no significant overrepresentation. This is consistent with an absence of known physical interactions between σ^{70} and bases at this position.

Note that a recent alignment of -35 elements (45) shows

TABLE 2. Significance of base overrepresentation for the short -10 region^a

Base	Ratio of base frequencies at position:				
	-11	-10	-9	-8	-7
A	3.17	1.05	2.14	2.01	0.10
T	0.29	1.56	0.60	0.38	3.21
C	0.04	0.56	0.60	1.05	0.22
G	0.03	0.63	0.42	0.43	0.06

^a The columns in the table correspond to the five bases in the short -10 region. Table entries were calculated as described in Table 1, footnote a.

notable discrepancies with the -35 element alignment presented here. Specifically, in that study, base C at position -31 is significantly less conserved compared to base A at -32 in the present study; this is inconsistent with the available data on interactions between σ^{70} and -35 element, which indicate that base -31 interacts with σ^{70} through hydrogen bonds, whereas interactions with position -32 involve weaker van der Waals interactions. Furthermore, in the previous study (45), bases A and T show greater conservation compared to bases C and A at positions -31 and -30, which is inconsistent with both the interaction data (6) and the -35 element consensus (-³⁵TTGACA⁻³⁰) established through previous studies (22). This finding contrasts with alignment here, in which the consensus -³¹C and -³⁰A are clearly distinguished from the other bases at positions -31 and -30.

Finally, Table 2 shows the specificity of positions -11 to -7, which correspond to short -10 element. The largest conservation corresponds to positions -11 and -7, which were shown in a number of studies to be of special importance for σ^{70} -ssDNA interactions (see, for example, references 13 and 27). On the other hand, mutations at position -10 showed no notable effect on σ^{70} -ssDNA binding (13), a finding consistent with the smallest base overrepresentation at this position.

Specificity of the -15 element. I next evaluated the specificity of the -15 element. In promoters of Gram-positive bacteria, two bases upstream of the extended -10 element (positions -16 and -17) are also conserved (the reported consensus of the extended -10 element is -¹⁷TRTG⁻¹⁴) (18, 50). I first sought to determine whether such conservation also existed for σ^{70} in *E. coli*, since this provides a possibility for further extension of the -15 element. However, no such conservation was detected, i.e., the two bases with the highest overrepresentation at these positions are -¹⁷T and -¹⁶C, which are both not significantly overrepresented (Table 3).

TABLE 3. Significance of base overrepresentation for the -15 region^a

Base	Ratio of base frequencies at position:					
	-17	-16	-15	-14	-13	-12
A	0.92	0.83	0.71	0.64	0.60	0.09
T	1.15	0.97	1.18	0.85	1.05	2.69
C	0.99	1.15	1.14	0.86	1.07	0.65
G	0.92	1.11	1.01	1.80	1.40	0.33

^a The columns in the table correspond to the four bases in the -15 region, including two additional upstream bases which were added to test for possible overrepresentation upstream of the -15 region. Table entries were calculated as described in Table 1, footnote a.

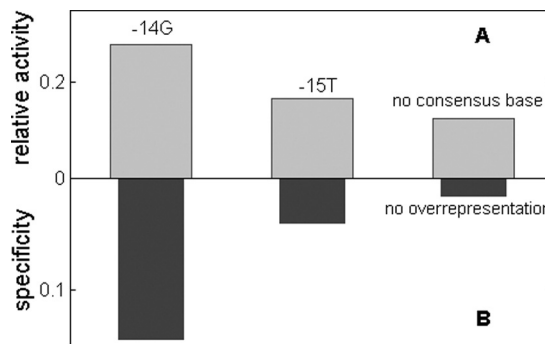


FIG. 2. Comparison of relative transcription activities and specificities at positions -14 and -15. (A) The first and second bars from the left show the averages of the transcription activities for all of the mutations where G and T remain, respectively, at positions -14 and -15. The third bar shows the average of the relative transcription activities where neither T nor G are present at positions -15 and -14. The transcription activities are measured relative to consensus -¹⁵TG⁻¹⁴, and the data are taken from an earlier report (32). (B) The three bars (from left to right) correspond to the specificities at positions -14, -15, and -16. The specificities were calculated from the inferred -15 element weight matrix as described in Materials and Methods. The figure shows that there is a significant contribution of -14G to transcription activity, despite the absence of -15T, and that the measured pattern of transcription activities is consistent with the pattern of calculated specificities.

I then examined the specificity of the binding positions within -15 motif. I first noted a high degeneracy at position -15, where bases T and C are similarly overrepresented (1.18 and 1.14 relative to the background frequencies). Therefore, it is more appropriate to represent the extended -10 motif with a weight matrix, or qualitatively with a degenerate consensus, than with a consensus sequence. Next, I noted a conservation of base G at position -13, which appears at a frequency 1.4 times greater than the background frequency (see Table 3). This overrepresentation is statistically highly significant ($P \approx 10^{-3}$). I also noted that the conservation of the base at position -13 was larger than the conservation at position -15, which is a canonical base within -15 motif (the T in TG). Conservation of base -13 at this position was not reported up to now. Actually, the consensus sequence for the extended -10 motif is presented in the literature as TGn, where the "n" at position -13 indicates no conservation (22).

To further investigate degeneracy of the -15 element, I reanalyzed data published previously (32), which presented the relative transcription activities for all mutations of the extended -10 consensus (-¹⁵TG⁻¹⁴) for four selected promoters. I calculated the following three averages and graphically represented them as bars in Fig. 2A: (i) the average over a total of 12 transcription activities corresponding to all mutations in four promoters where G remains at position -14, (ii) the average over total of 12 transcription activities corresponding to all mutations in four promoters where T remains at position -15, and (iii) the average of the relative transcription activities for all of the remaining mutations, where neither T nor G is, respectively, at positions -15 and -14 (32). By comparing the first and the third bar from left in Fig. 2A, it could be seen that G at position -14 provides a significant contribution to transcription activity, despite the absence of T at position -15. I

also noted that, compared to the G at position -14 , the presence of T at position -15 provided a notably smaller contribution to transcription activity (compare the first and second bars in Fig. 2A).

Furthermore, in Fig. 2B, I used the inferred weight matrix for the -15 element to calculate specificities (9) (see Materials and Methods)—which provides a measure of base conservation—for positions -14 , -15 , and -16 . The specificity at position -16 was included for comparison, since there is no base overrepresentation at this position (see Table 3). I noted that the specificity at position -15 was significantly smaller compared to the specificity at position -14 , which is consistent with significantly larger contribution of G at position -14 to the transcription activity (see Fig. 2A).

Relation between the -15 element and other promoter elements. I determined weight matrices for each of the promoter elements by using the alignments described above. Each of the weight matrices was then used to calculate the strengths of the promoter elements obtained in the alignment. One should note that the weight matrix score for a certain element provides an estimate of this element's contribution to the overall promoter strength (8). Zeros for the weight matrix scores (element strengths) are defined so that maximal weight matrix score corresponds to zero; consequently, the strengths of all sequences that deviate from the consensus are negative, and a stronger element corresponds to a larger (less-negative) weight matrix score. In addition, I assigned weights to four different spacer lengths (see Fig. 1B and Materials and Methods), which were used in the estimate of the overall promoter strength. That is, the optimal spacer length (17 bp) has the largest weight and consequently the largest contribution to the promoter strength, while smaller weights are assigned to suboptimal spacer lengths (e.g., 15 and 19 bp). I used these weight matrices to calculate the strengths of the -35 elements, the -15 elements, and the short -10 elements and to estimate the overall promoter strength. Note that the term "element strength" used in the present study corresponds to the estimated strength calculated by using the weight matrices.

The estimated strengths of the -15 promoter elements were next plotted against the corresponding strengths of the -35 element and the short -10 element. These relations are graphically presented in Fig. 3A and B, where Pearson correlation coefficients (referred to as simply "correlation coefficients" here) are also indicated. The strengths of both the short -10 element and the -35 element are negatively correlated with the strength of the -15 element. Therefore, there is a global tendency to have a stronger -15 element when a weaker -10 element or a weaker -35 element is present. Moreover, the negative correlation is stronger for -10 elements than for -35 elements. In the case of the -35 element the correlation coefficient was -0.10 , which is marginally significant ($P = 0.06$); the correlation in the case of the short -10 element is -0.17 , which is highly statistically significant ($P = 2 \times 10^{-3}$). The stronger correlation in the case of -10 element seems surprising, since both the -35 and the -15 elements are involved in RNAP-dsDNA interactions, while the short -10 element is involved in the open complex formation through RNAP-dsDNA interactions. This issue is discussed further below.

I furthermore correlated -15 element strengths with the overall promoter strength in the absence of -15 element (Fig.

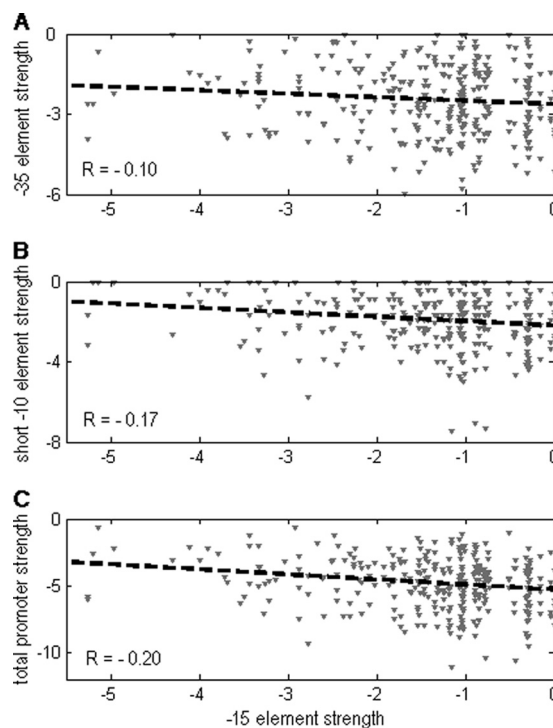


FIG. 3. Correlation of -15 element strength with other promoter strengths. Weight matrices inferred from the alignment were used to calculate the strengths of the -35 elements, the -15 elements, the short -10 elements, and the overall promoter strength. Correlations between the following strengths are shown: -15 element and -35 element (A), -15 element and short -10 element (B), and -15 element and overall promoter strength (C). Correlation coefficients are indicated in the figures. The figure shows negative correlations between the strength of the -15 element and the other calculated strengths, where the negative correlation is larger for the short -10 element compared to the -35 element, and largest for the overall promoter strength.

3C). The overall promoter strength in the absence of the -15 element was estimated as a sum of strengths that correspond to the -35 element, the strength of short -10 element, and the spacer weight. There was a highly significant negative correlation between -15 element strength and the overall promoter strength (correlation coefficient of -0.20 , with a P of 3×10^{-4}). This correlation is stronger than for the individual promoter elements. One should also note that the strength of σ^{70} -dsDNA interactions in the absence of the -15 element is simply given by the strength of the -35 element. Therefore, by comparing Fig. 3A and C, it is evident that a much stronger negative correlation is associated with the overall promoter strength than with σ^{70} -dsDNA interactions.

As described above, I analyzed the entire set of promoters and established that there is globally an inverse relationship between the strength of the -15 element on one side and -35 element, the -10 element, and the entire promoter strength on the other side. I now specifically concentrated on promoters with weak -35 and short -10 elements and compare their -15 element strengths with the -15 element strengths of promoters with strong -35 and -10 boxes. To do this, I selected the following six groups of promoters: (i) 20% of promoters with the weakest -35 elements, (ii) 20% of promoters with the

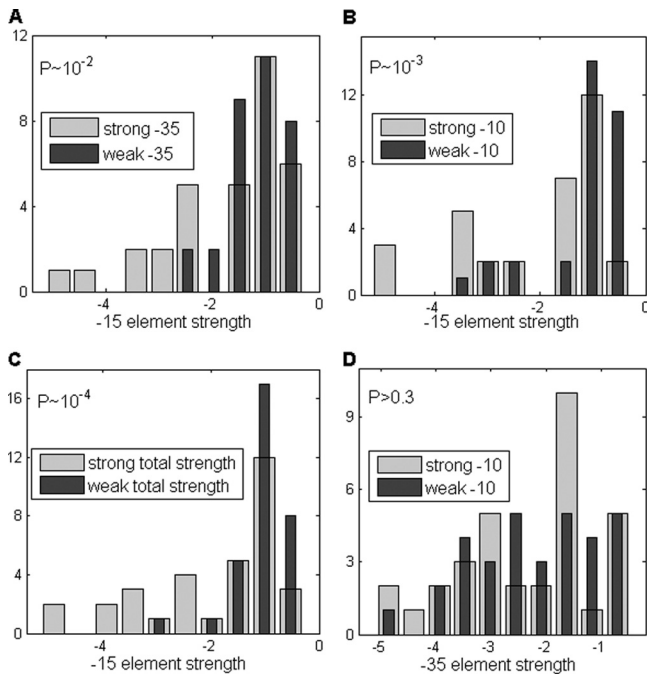


FIG. 4. Score distributions for weak and strong promoter elements. (A to C) Comparison of the -15 element score distributions are presented for promoters with weak (dark gray bars) and strong (light gray bars) -35 elements (A), promoters with weak and strong -10 elements (B), and promoters with weak and strong overall promoter strength (C). (D) Distribution of -35 element scores for weak and strong -10 elements. *P* values corresponding to the differences between the means of the two distributions are indicated in the figures. Panels A, B, and C show that the -15 element has a tendency to rescue promoters where other elements are weak, where this tendency is stronger for the -10 element than for the -35 element, and the strongest for the overall promoter strength. In contrast to this, panel D shows that there is no tendency of the -35 element to rescue promoters with weak -10 elements.

strongest -35 elements, (iii) 20% of promoters with the weakest -10 elements, (iv) 20% of promoters with the strongest -10 elements, (v) 20% of promoters with the weakest overall promoter strength, and (vi) 20% of promoters with the strongest overall promoter strength. As described above, the overall promoter strength was calculated without the -15 element so as not to bias the correlation between the overall promoter strength and the -15 element strength.

I next calculated the distribution of the -15 element scores for all six groups of promoters defined above. In Fig. 4A to C, the distributions of -15 element scores are compared for (i) promoters with the strongest and weakest -35 elements (the first and second groups defined above) (Fig. 4A), (ii) promoters with the strongest and weakest -10 elements (the third and fourth groups defined above) (Fig. 4B), and (iii) promoters with the strongest and weakest overall promoter strength (the fifth and the sixth groups defined above) (Fig. 4C).

In Fig. 4A, the distribution of the -15 scores for promoters with the weak -35 element is notably shifted toward higher scores compared to the distribution of the -15 scores for promoters with the strong -35 elements. The difference between the means of the two distributions (for the promoters with strong and weak -35 elements) is statistically significant,

with a *P* value of 10^{-2} . Similarly, Fig. 4B and C show that the distribution of -15 element scores is significantly shifted toward the stronger scores for both promoters with weak -10 elements and weak overall promoter strength (*P* values of 10^{-3} and 2×10^{-4}). Therefore, promoters characterized by weaknesses in either of the promoter elements (-35 and -10) or the overall promoter strength are indeed characterized by enrichment of the strong -15 elements.

Furthermore, when distributions shown in Fig. 4A and B are compared, it is clear that there is a significantly larger shift between the distributions for the -10 element compared to the -35 element. Therefore, -15 elements have a significantly greater tendency to rescue promoters with weak -10 elements compared to promoters with weak -35 elements; this result is consistent with the global tendency shown in Fig. 3. Finally, Fig. 4C shows that the shift between the distributions of the -15 element scores is most pronounced when comparing promoters with weak and strong overall promoter strengths. Comparison of Fig. 4A and C shows that there is a much stronger tendency of -15 motif to rescue promoters with a weak overall promoter strength than promoters with weak σ^{70} -dsDNA interactions (i.e., weak -35 element).

Relations between other promoter elements. I next analyzed the relationship between other promoter elements (the -35 element, the -10 element, and spacer lengths), so that they could be compared to the behavior of the -15 elements. Accordingly, I used the inferred weight matrices (see above) to calculate the strengths of the -35 and -10 elements for all aligned promoters. I first noted an absence of negative correlation between the strengths of the -35 elements and -10 elements (see Fig. S1 in the supplemental material). This result is in contrast to the negative correlations that were obtained for the -15 element (Fig. 3). Furthermore, similar to what was observed in Fig. 4A to C, I next concentrated specifically on promoters with weak and strong -10 elements and sought to determine whether there is a significant difference in -35 element strengths associated with these promoters. As can be seen in Fig. 4D, there is no significant difference in the distribution of -35 element scores for these two groups of promoters, so promoters with weak -10 elements are not supplemented with strong -35 elements; this result is consistent with the absence of a global negative correlation noted in Fig. S1 in the supplemental material.

I further investigated the relationships between spacer length and the strengths of other promoter elements. Because of the negative correlation between the strengths of the -15 element and the other two promoter elements (-35 and short -10), one might expect a similar behavior for spacer lengths, so that suboptimal spacer lengths (15 and 19 bp) are enriched with stronger -15 motifs. Furthermore, an earlier study (32) reported that the presence of the extended -10 motif (TG) upstream of the -10 element is associated with a larger spacer length, which was interpreted by necessity to allow more space for interactions of σ^{70} with TG motif. This observation might be extrapolated to the expectation that stronger -15 elements should be associated with longer spacer lengths.

To test these two possibilities, I divided the promoters in groups according to their spacer length. For each of these groups, I calculated and consequently plotted (i) the mean value of the -15 element strength (Fig. 5A), (ii) the mean

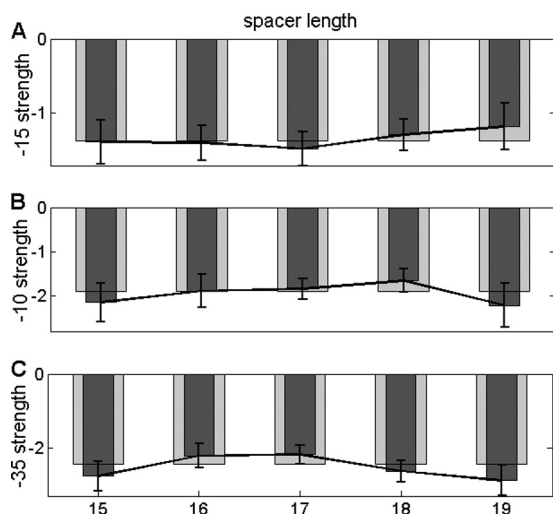


FIG. 5. Dependence of promoter element strengths on spacer length. The relationship between the spacer length and -15 element strength (A), -10 element strength (B), and -35 element strength (C) was evaluated. The bars correspond to the mean strength of a promoter element for (i) all of the promoters with a given spacer length (dark gray bars) and (ii) all of the promoters irrespective of the spacer length (light gray bars). The error bars correspond to the 95% confidence intervals. The figure indicates no significant dependence of the element strength on the spacer length.

value of the -10 element strength (Fig. 5B), and (iii) the mean value of the -35 element strength (Fig. 5C). However, there was no significant relation between the spacer length and the strengths of the three motifs. In Fig. 5A, a slight tendency for longer spacer lengths (18 and 19) to be associated with stronger -15 elements is evident, which is reminiscent of those earlier findings (32). However, this tendency is statistically insignificant. In Fig. 5B no tendency whatsoever can be seen, while in Fig. 5C there is a weak tendency for stronger -35 elements to be associated with more optimal spacer lengths, but this tendency is only marginally statistically significant.

DISCUSSION

An extension of the extended -10 element to two additional downstream bases to form the -15 element was proposed by Hook-Barnard and Hinton (22), who were motivated by previous detailed analyses of the extended -10 element (5, 32). The introduction of the -15 element was entirely based on biochemical arguments, i.e., on the fact that this element groups together the bases within and upstream of the -10 element that interact with σ^{70} in dsDNA form (22). The results reported here show that it is also useful to introduce the -15 element from the point of genomic analysis of the promoter sequences for the following reasons: (i) there exists a continuous conserved stretch of sequence corresponding to -15 element, and (ii) the -15 element was redefined here as a degenerate motif (weight matrix), so that it is described in a manner analogous to that of the -35 and -10 elements. These two points are discussed below in more detail.

I showed here that the stretch of conserved sequence that corresponds to the -15 element includes a previously unrecognized conservation of G at position -13 . Although the im-

portance of the G at position -13 was demonstrated for *Thermus aquaticus* σ^A promoters (12) (note that σ^A in *T. aquaticus* corresponds to σ^{70} in *E. coli*), the importance of position -13 for *E. coli* σ^{70} transcription was, to my knowledge, not reported until now. Understanding the exact functional role of the conserved G at position -13 therefore opens up a new area for experimental work.

The lack of research elucidating the role of position -13 is in contrast to the detailed mutational analysis of positions -17 to -14 (5, 32). Emphasis on these positions in *E. coli* was influenced by the conservation of this region in *Bacillus subtilis* (18, 33) (consensus $^{-17}\text{TRTG}^{-14}$). On the other hand, I detected no significant overrepresentation of bases at positions -17 and -16 in *E. coli*. This finding is consistent with the absence of known contacts between the σ subunit and bases at positions -17 and -16 (22). The absence of conservation at these two positions also indicates that there is no flexible spacer between the -15 element and the short -10 element, so that the -15 element is physically strongly linked to the short -10 element.

As the second point, I defined here the -15 element as a degenerate motif, which is in contrast to the previously used classification of extended -10 promoters based on a binary presence or absence of $^{-15}\text{TG}^{-14}$ (22). Related to this, some promoter detection methods (19, 23, 38, 44) include in weight matrices regions that flank "core" -10 and -35 elements; by doing so, these methods automatically take into account a degenerate description of the extended -10 region (as well as of other flanking bases). Since in the existing promoter alignments (45) bases -15 to -13 do not appear as significantly overrepresented, the -15 element weight matrix that I derived here can be incorporated into these (and other) promoter detection methods (30, 41, 45) to improve their specificity. Furthermore, since, to my best knowledge, the alignment presented here is the first to accurately detect both the -35 element and the -15 element (see Results), the weight matrices for other promoter elements derived here can be used to further improve accuracy of promoter detection.

In addition to the sequence analysis presented in the previous section, the degeneracy of the extended -10 (and consequently the -15) element is also supported by biochemical evidence. Specifically, based on (re)analyses of earlier experimental data (32), it follows that the presence of G at position -14 substantially contributes to transcription activity, even in the absence of T at position -15 . Conversely, it was shown that the presence of T at position -15 in Plac, even in the absence of G at position -14 , influences transcription levels (26, 34). However, taking into account earlier findings (5, 32), it follows that a G at position -14 provides a significantly larger contribution to transcription activity compared to a T at position -15 , which is consistent with a significantly larger conservation (specificity) at position -14 obtained in the alignment presented here. The larger contribution of G at position -14 to transcription activity is also consistent with genetic (1, 4, 43) and structural (36) analyses, which suggest that there is a direct interaction between residues of σ^{70} region 3 and position -14 .

An established view of the extended -10 element is that its main role is to complement weaknesses in the -35 box, so that it ensures a sufficient number of σ^{70} -dsDNA contacts to form a closed complex in the case of a weak -35 box (24, 29, 40, 53).

One might expect to extrapolate this view to the -15 element, particularly since this element is defined to entirely correspond to σ^{70} -dsDNA interactions. Surprisingly, I determined that the -15 element is much more related to weaknesses in the -10 element than to deficient -35 elements; consequently, the -15 element has a much greater tendency to complement promoters with weak overall promoter strength than promoters with weak σ^{70} -dsDNA interactions. This last conclusion suggests that the main parameter determining a functional promoter is the overall promoter strength, rather than the minimal number of σ^{70} -dsDNA interactions. Since the activity of some of the promoters used in this analysis may depend on transcription activators (14), an interesting future question will be to determine whether the primary role of these activators is to also complement the overall promoter strength.

Although contrary to the established view, the significant role of the -15 element in complementing weak -10 elements appears not to be inconsistent with existing evidence. Specifically, it was observed earlier (32) that promoters containing $^{-15}\text{TG}^{-14}$ are able to tolerate weaker -10 elements. Furthermore, both gapAP1 and Pminor promoters have weak -10 elements, and the presence of a strong -15 element is required for promoter function (21, 49). On the other hand, mutations that improve the -10 element were shown to abolish the requirement for a strong -15 element, therefore demonstrating the direct role of the -15 element in complementing weak -10 element strength.

I also found that, in contrast to the -15 element, the other promoter elements (-10, spacer, and -35) showed no mutual complementation of their strengths. It is useful to put this result in the context of recently proposed "mix and match" model of promoter recognition (22, 31). According to this model, different promoter elements mix with each other and match each other's strengths in order to provide a sufficient number of σ^{70} -DNA interactions. The results presented here allow further refinement of this model, so that the -15 element is defined as a "matcher," i.e., an element with a specific role in complementing the weaknesses in other promoter elements. On the other hand, the -10 element, the spacer, and the -35 element may be called "mixers" since their strengths—while uncoupled to each other—clearly mix to result in a required level of transcription activity.

This discussion raises an issue regarding distinction between the -15 element and the short -10 element. I here established that there is a continuous stretch of conserved sequence spanning from positions -15 to -7 and that there is a significant complementation between the strengths of the -15 element and the short -10 element. Because of this, it may be tempting to connect these two elements in one "super" -10 element spanning positions -15 to -7. There are, however, two arguments against the formation of such a long element. First, such an element would not be a single DNA recognition element, since it mixes bases (-15 to -12) that interact with σ^{70} in dsDNA form (binding element) and bases that interact with σ^{70} in ssDNA form (melting element). This is the original biochemical argument (22) for introducing the -15 element. Second, I found that the -15 element has a unique role as a matcher, i.e., an element that complements the strengths of both the -35 element and the short -10 element. On the other hand, there is no mutual complementation between the

-35 and short -10 elements. Therefore, such a "super" -10 element would be heterogeneous both biochemically and functionally, and I think that its separation into the -15 element and the short -10 element is more useful.

Finally, it would be interesting to extend the analysis presented here to UP elements, which (in the same way as -15 elements) interact with only dsDNA but appear to exist in only a small fraction of promoters (~5%) (11). Applying the terminology used above, such an analysis would allow us to determine whether UP elements belong to "mixers" (their strength uncoupled to strengths of other elements) or "matchers" (so that, similarly to the -15 element, they complement other elements). If the latter were true, then a system in which major promoter elements (the -10 element, the -35 element, and the spacer) mix in order to achieve a required level of transcription activity, while nonessential promoter elements match their strengths, might emerge. However, a reliable analysis of UP elements is a very challenging task, given both their reported rare occurrence in promoters (11) and the apparently complex rules through which they contribute to transcription activity (16).

ACKNOWLEDGMENTS

This study was supported by a Marie Curie International Reintegration Grant within the 7th European Community Framework Programme (PIRG08-GA-2010-276996) and by the Ministry of Science and Technological Development of the Republic of Serbia under project ON173052.

REFERENCES

1. Barne, K. A., J. A. Bown, S. J. Busby, and S. D. Minchin. 1997. Region 2.5 of the *Escherichia coli* RNA polymerase σ^{70} subunit is responsible for the recognition of the "extended -10" motif at promoters. *EMBO J.* **16**:4034-4040.
2. Borukhov, S., and E. Nudler. 2003. RNA polymerase holoenzyme: structure, function, and biological implications. *Curr. Opin. Microbiol.* **6**:93-100.
3. Borukhov, S., and K. Severinov. 2002. Role of the RNA polymerase sigma subunit in transcription initiation. *Res. Microbiol.* **153**:557-562.
4. Bown, J. A., et al. 1999. Organization of open complexes at *Escherichia coli* promoters: location of promoter DNA sites close to region 2.5 of the σ^{70} subunit of RNA polymerase. *J. Biol. Chem.* **274**:2263-2270.
5. Burr, T., J. Mitchell, A. Kolb, S. Minchin, and S. Busby. 2000. DNA sequence elements located immediately upstream of the -10 hexamer in *Escherichia coli* promoters: a systematic study. *Nucleic Acids Res.* **28**:1864.
6. Campbell, E. A., et al. 2002. Structure of the bacterial RNA polymerase promoter specificity (σ) subunit. *Mol. Cell* **9**:527-539.
7. DeHaseth, P. L., M. L. Zupancic, and M. T. Record, Jr. 1998. RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. *J. Bacteriol.* **180**:3019-3025.
8. Djordjevic, M., and R. Bundschuh. 2008. Formation of the open complex by bacterial RNA polymerase: a quantitative model. *Biophys. J.* **94**:4233-4248.
9. Djordjevic, M., A. M. Sengupta, and B. I. Shraiman. 2003. A biophysical approach to transcription factor binding site discovery. *Genome Res.* **13**:2381-2390.
10. Ebright, R. H. 2000. RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J. Mol. Biol.* **304**:687-698.
11. Estrem, S. T., T. Gaal, W. Ross, and R. L. Gourse. 1998. Identification of an UP element consensus sequence for bacterial promoters. *Proc. Natl. Acad. Sci. U. S. A.* **95**:9761-9766.
12. Feklistov, A., et al. 2006. A basal promoter element recognized by free RNA polymerase [σ] subunit determines promoter recognition by RNA polymerase holoenzyme. *Mol. Cell* **23**:97-107.
13. Fenton, M. S., and J. D. Gralla. 2001. Function of the bacterial TATAAT -10 element as single-stranded DNA during RNA polymerase isomerization. *Proc. Natl. Acad. Sci. U. S. A.* **98**:9020-9025.
14. Gama-Castro, S., et al. 2011. RegulonDB version 7.0: transcriptional regulation of *Escherichia coli* K-12 integrated within genetic sensory response units (Gensor units). *Nucleic Acids Res.* **39**:D98.
15. Gerland, U., J. D. Moroz, and T. Hwa. 2002. Physical constraints and functional characteristics of transcription factor-DNA interaction. *Proc. Natl. Acad. Sci. U. S. A.* **99**:12015-12020.

16. Gourse, R. L., W. Ross, and T. Gaal. 2000. Ups and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Mol. Microbiol.* **37**:687–695.
17. Harley, C. B., and R. P. Reynolds. 1987. Analysis of *Escherichia coli* promoter sequences. *Nucleic Acids Res.* **15**:2343–2361.
18. Helmann, J. D. 1995. Compilation and analysis of *Bacillus subtilis* A-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA. *Nucleic Acids Res.* **23**:2351.
19. Hertz, G. Z., and G. D. Stormo. 1996. *Escherichia coli* promoter sequences: analysis and prediction. *Methods Enzymol.* **273**:30–42.
20. Hertz, G. Z., and G. D. Stormo. 1999. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics* **15**:563–577.
21. Hook-Barnard, I., X. B. Johnson, and D. M. Hinton. 2006. *Escherichia coli* RNA polymerase recognition of a σ^{70} -dependent promoter requiring a -35 DNA element and an extended -10 TGn motif. *J. Bacteriol.* **188**:8352–8359.
22. Hook-Barnard, I. G., and D. M. Hinton. 2007. Transcription initiation by mix and match elements: flexibility for polymerase binding to bacterial promoters. *Gene Regul. Systems Biol.* **1**:275.
23. Huerta, A. M., and J. Collado-Vides. 2003. Sigma 70 promoters in *Escherichia coli*: specific transcription in dense regions of overlapping promoter-like signals. *J. Mol. Biol.* **333**:261–278.
24. Kumar, A., et al. 1993. The -35 -recognition region of *Escherichia coli* σ^{70} is inessential for initiation of transcription at an \llcorner extended minus 10 \gg promoter. *J. Mol. Biol.* **232**:406–418.
25. Lawrence, C. E., et al. 1993. Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science* **262**:208–214.
26. Liu, M., S. Garges, and S. Adhya. 2004. lacP1 promoter with an extended -10 motif: pleiotropic effects of cyclic AMP protein at different steps of transcription initiation. *J. Biol. Chem.* **279**:54552–54557.
27. Matlock, D. L., and T. Heyduk. 2000. Sequence determinants for the recognition of the fork junction DNA containing the -10 region of promoter DNA by *Escherichia coli* RNA polymerase. *Biochemistry* **39**:12274–12283.
28. Michalowski, C. B., M. D. Short, and J. W. Little. 2004. Sequence tolerance of the phage lambda PRM promoter: implications for evolution of gene regulatory circuitry. *J. Bacteriol.* **186**:7988–7999.
29. Minakhin, L., and K. Severinov. 2003. On the role of the *Escherichia coli* RNA polymerase 70 region 4.2 and subunit C-terminal domains in promoter complex formation on the extended -10 galP1 promoter. *J. Biol. Chem.* **278**:29710.
30. Mironov, A. A., N. P. Vinokurova, and M. S. Gelfand. 2000. Software for analyzing bacterial genomes. *Mol. Biol.* **34**:253–262. (In Russian.)
31. Miroslavova, N. S., and S. J. Busby. 2006. Investigations of the modular structure of bacterial promoters. *Biochem. Soc. Symp.* **2006**:1–10.
32. Mitchell, J. E., D. Zheng, S. J. W. Busby, and S. D. Minchin. 2003. Identification and analysis of “extended -10 ” promoters in *Escherichia coli*. *Nucleic Acids Res.* **31**:4689.
33. Moran, C. P., et al. 1982. Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. *Mol. Gen. Genet.* **186**:339–346.
34. Munson, L. M., W. Mandecki, M. H. Caruthers, and W. S. Reznikoff. 1984. Oligonucleotide mutagenesis of the lacPUV5 promoter. *Nucleic Acids Res.* **12**:4011–4017.
35. Murakami, K. S., and S. A. Darst. 2003. Bacterial RNA polymerases: the whole story. *Curr. Opin. Struct. Biol.* **13**:31–39.
36. Murakami, K. S., S. Masuda, E. A. Campbell, O. Muzzin, and S. A. Darst. 2002. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* **296**:1285–1290.
37. Niedziela-Majka, A., and T. Heyduk. 2005. *Escherichia coli* RNA polymerase contacts outside the -10 promoter element are not essential for promoter melting. *J. Biol. Chem.* **280**:38219.
38. Ozoline, O. N., and A. A. Deev. 2006. Predicting antisense RNAs in the genomes of *Escherichia coli* and *Salmonella typhimurium* using promoter-search algorithm PlatProm. *J. Bioinform. Comput. Biol.* **4**:443–454.
39. Paget, M. S. B., and J. D. Helmann. 2003. The sigma 70 family of sigma factors. *Gen. Biol.* **4**:203–208.
40. Ponnambalam, S., C. Webster, A. Bingham, and S. Busby. 1986. Transcription initiation at the *Escherichia coli* galactose operon promoters in the absence of the normal -35 region sequences. *J. Biol. Chem.* **261**:16043.
41. Robison, K., A. McGuire, and G. Church. 1998. A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete *Escherichia coli* K-12 genome. *J. Mol. Biol.* **284**:241–254.
42. Roulet, E., et al. 2002. High-throughput SELEX-SAGE method for quantitative modeling of transcription-factor binding sites. *Nat. Biotechnol.* **20**:831–835.
43. Sanderson, A., J. E. Mitchell, S. D. Minchin, and S. J. Busby. 2003. Substitutions in the *Escherichia coli* RNA polymerase sigma70 factor that affect recognition of extended -10 elements at promoters. *FEBS Lett.* **544**:199–205.
44. Shavkunov, K. S., I. S. Masulis, M. N. Tutukina, A. A. Deev, and O. N. Ozoline. 2009. Gains and unexpected lessons from genome-scale promoter mapping. *Nucleic Acids Res.* **37**:4919–4931.
45. Shultzaberger, R. K., Z. Chen, K. A. Lewis, and T. D. Schneider. 2007. Anatomy of *Escherichia coli* σ^{70} promoters. *Nucleic Acids Res.* **35**:771–788.
46. Stormo, G. D. 2000. DNA binding sites: representation and discovery. *Bioinformatics* **16**:16–23.
47. Stormo, G. D., and D. S. Fields. 1998. Specificity, free energy and information content in protein-DNA interactions. *Trends Biochem. Sci.* **23**:109–113.
48. Thompson, W., E. C. Rouchka, and C. E. Lawrence. 2003. Gibbs recursive sampler: finding transcription factor binding sites. *Nucleic Acids Res.* **31**:3580–3585.
49. Thouvenot, B., B. Charpentier, and C. Branlant. 2004. The strong efficiency of the *Escherichia coli* gapA P1 promoter depends on a complex combination of functional determinants. *Biochem. J.* **383**:371–382.
50. Voskuil, M. I., and G. H. Chambliss. 2002. The TRTGn motif stabilizes the transcription initiation open complex. *J. Mol. Biol.* **322**:521–532.
51. Wang, H., and C. J. Benham. 2006. Promoter prediction and annotation of microbial genomes based on DNA sequence and structural responses to superhelical stress. *BMC Bioinform.* **7**:248.
52. Workman, C. T., et al. 2005. enoLOGOS: a versatile web tool for energy normalized sequence logos. *Nucleic Acids Res.* **33**:W389–W392.
53. Young, B. A., T. M. Gruber, and C. A. Gross. 2004. Minimal machinery of RNA polymerase holoenzyme sufficient for promoter melting. *Science* **303**:1382–1384.