Acta Biochimica Polonica

Vol. 50 No. 4/2003

941–945

QUARTERLY

Communication

Non-random distribution of GATC sequences in regions of promoters stimulated by the SeqA protein of *Escherichia coli*✪

Barbara Strzelczyk[1], Monika Słomińska-Wojewódzka[1], Grzegorz Węgrzyn[1,2] and Alicja Węgrzyn[3]✉

[1]*Department of Molecular Biology, University of Gdańsk, Gdańsk, Poland;* [2]*Institute of Oceanology, Polish Academy of Sciences, Gdynia, Poland;* [3]*Laboratory of Molecular Biology (affiliated with the University of Gdańsk), Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Gdańsk, Poland*

Received: 31 October, 2003; revised: 28 November, 2003; accepted: 03 December, 2003

Key words: SeqA protein, transcription regulation, promoter sequence, GATC motif

The SeqA protein of *Escherichia coli* is not only the main negative regulator of DNA replication initiation but also a specific transcription factor. It binds to hemimethylated GATC sequences and, with somewhat different specificity, to fully methylated GATC regions. Recently, a microarray analysis was reported, in which transcriptomes of wild-type and Δ*seqA* strains were compared. Although in the *seqA* mutant the levels of some transcripts were significantly decreased while certain transcripts were evidently more abundant relative to wild-type bacteria, no correlation between the presence of GATC motifs in promoter sequences and transcription activity was found. However, here we show that when larger DNA fragments, encompassing positions from –250 to +250 relative to the transcription start site, are analyzed, some common features of GATC distribution near the promoters activated by SeqA can be demonstrated. Nevertheless, it seems that the GATC pattern is not the only determinant of SeqA-dependence of promoter activity.

In *Escherichia coli,* the product of the *seqA* gene is the main negative regulator of chromosome replication initiation (Lu *et al.*, 1994; von Freiesleben *et al.*, 1994; Slater *et al.*,

✪This work was supported by the State Committee for Scientific Research (KBN, Poland, grant No. 3 P04A 029 22 to A.W.). G.W. acknowledges financial support from the Foundation for Polish Science (subsidy 14/2000).

✉Corresponding author: Dr. Alicja Węgrzyn, Laboratory of Molecular Biology, University of Gdańsk, Kładki 24, 80-822 Gdańsk, Poland; tel.: (48 58) 346 3014; fax: (48 58) 301 0072; e-mail: wegrzyn@biotech.univ.gda.pl

1995; Boye *et al.*, 1996; Wold *et al.*, 1998). The SeqA protein is involved in sequestration of the *origin* of chromosome replication, *oriC*, in the hemimethylated state (Lu *et al.*, 1994; Slater *et al.*, 1995). This regulatory process is crucial for precise control of bacterial cell cycle.

SeqA binds specifically to hemimethylated GATC sequences located on oligonucleotide fragments but no specific binding to analogous fully methylated or unmethylated DNA was reported (Brendler & Austin, 1999; Kang *et al.*, 1999). However, when longer DNA fragments were employed, SeqA did bind a fully methylated DNA fragment bearing *oriC* strongly and specifically (Slater *et al.*, 1995). Moreover, hemimethylated *oriC* and non-*oriC* sequences were bound by SeqA with similar affinity (Slater *et al.*, 1995). Investigation of SeqA-binding sites on hemimethylated *oriC* revealed the same preferential binding to each side of the DnaA box R1, as was found for fully methylated *oriC* (Skarstad *et al.*, 2000). Thus, this binding has the same specificity to certain GATC sites irrespective of whether they are fully or hemimethylated. According to this statement, it was demonstrated that SeqA interacts specifically with certain bacteriophage λ DNA fragments containing GATC sequences both in the fully methylated and hemimethylated state (Słomińska *et al.*, 2001).

Perhaps surprisingly, recent studies indicated that apart from being the main negative regulator of DNA replication initiation, SeqA is also a specific transcription factor. Namely, this protein specifically stimulates transcription from bacteriophage λ p_R, p_I and p_{aQ} promoters (p_I and p_{aQ} require also the action of the λ-encoded CII protein) both *in vivo* and *in vitro*, whereas some other λ promoters (like p_L and p_E) appear to be SeqA-independent (Słomińska *et al.*, 2001; 2003a; 2003b). Very recently, effects of Δ*seqA* mutation on the expression of all *E. coli* genes were studied using microarray analysis (Lobner-Olsen *et al.*, 2003). Transcription of certain genes was either decreased or increased in the *seqA* mutant, while levels of transcripts of most genes were similar (less than two-fold difference) in the wild-type and Δ*seqA* strains.

Since SeqA binds GATC sequences, a correlation between the presence of such motifs in the promoter sequence and stimulation or repression of transcription by the *seqA* gene product was analyzed. Lobner-Olsen *et al.* (2003) did not find any common features of the promoters whose activities appeared to be significantly changed in the *seqA* mutant relative to wild-type bacteria. Therefore, those authors excluded the possibility that SeqA acts directly at promoter sequences, and suggested that the decreased or increased transcription of certain genes in the *seqA* mutant may result from changes in nucleoid organization. However, contrary to that conclusion, it was demonstrated experimentally that SeqA stimulates transcription from certain promoters both *in vivo* and *in vitro*, even when relatively short DNA templates were used in *in vitro* experiments (Słomińska *et al.*, 2001; 2003a; 2003b). These results strongly suggested that the SeqA protein is directly involved in the regulation of activity of certain promoters. Interestingly, analysis of distribution of GATC motifs near bacteriophage λ promoters, as well as the results of studies on SeqA–DNA interactions at these regions, strongly suggest that SeqA-binding sites downstream of a promoter may be important for the SeqA-mediated stimulation of promoter activity (Słomińska *et al.*, 2003a; 2003b). These sites can be located as far as 100–200 bp from the transcription start site (Fig. 1). Such regions were not analyzed by Lobner-Olsen *et al.* (2003). Therefore, here we analyzed the distribution of GATC motifs in regions of promoters (from position –250 to +250) of the *E. coli* genes activated and repressed in the *seqA* mutant (reported by Lobner-Olsen *et al.*, 2003), and in randomly selected genes whose transcription was similar in the wild-type and Δ*seqA* strains.

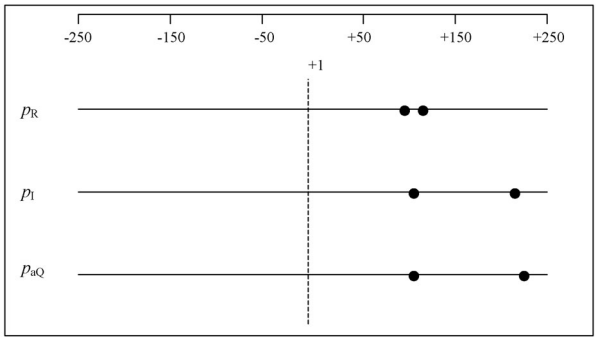

Figure 1. Distribution of SeqA-binding sites (dots) in regions of bacteriophage λ promoters (p_R, p_I and p_{aQ}) stimulated by the *seqA* gene product.

The binding sites were determined using electron microscopy; the data are taken from Słomińska *et al.* (2001, 2003a, 2003b). Nucleotide positions (from –250 to +250) are relative to transcription start points (+1; dotted vertical line) and are shown at the top of the panel.

MATERIALS AND METHODS

The list of genes whose transcription was either induced or repressed was taken from the work by Lobner-Olsen *et al.* (2003). Sequences of these genes (with adjacent regions), as well as of regions of randomly selected genes whose transcription was reported to be SeqA-independent (Lobner-Olsen *et al.*, 2003), were taken from the GenBank database (http://www.ncbi.nml.nih.gov). The sequences were aligned to show locations of GATC motifs relative to the transcription start site of each promoter.

RESULTS AND DISCUSSION

Results of our analysis of distribution of GATC motifs in regions of promoters of *E. coli* genes activated, repressed and unaffected in the *E. coli seqA* mutant are presented in Figs. 2 and 3. Although it is difficult to find any patterns of the GATC motifs that might by obligatory for SeqA-mediated regulation of promoter activity, the distribution of these sequences near promoters that are depressed in *seqA* mutants (thus, potentially activated by SeqA) seems to be non-random (Fig. 2A). In a relatively large fraction of these promoters (9 out of 19), one GATC motif occurs in a region near 100 (between 60 and 140) bp downstream of the promoter (promoters of genes: *cydD, gltB, trpE, holD, hisG, gapC, asnA, dld, osmE*), often followed by at least one more such sequence located either downstream or upstream of the transcription start site (all promoters listed above

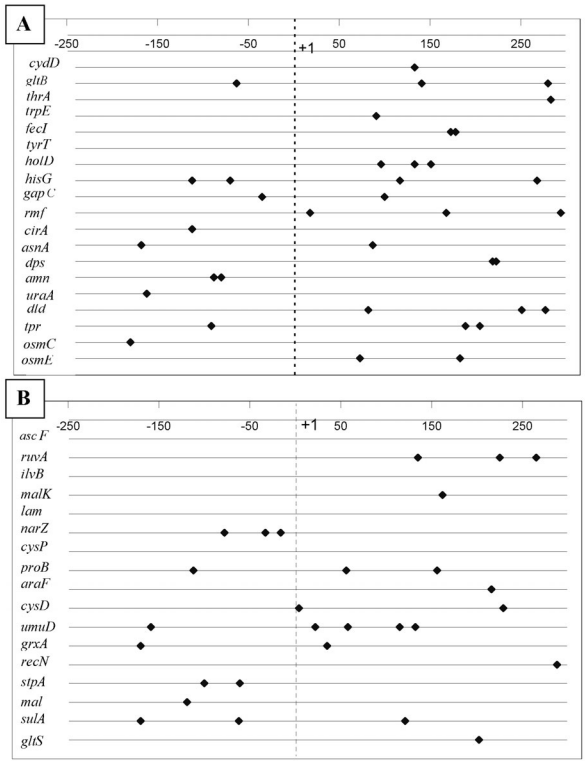

Figure 2. Distribution of GATC sequences (dots) in regions of promoters of genes for which decreased (panel A) or increased (panel B) mRNA levels in the Δ*seqA* mutant were demonstrated recently (Lobner-Olsen *et al.*, 2003).

Nucleotide positions (from –250 to +250) are relative to transcription start points (+1; dotted vertical line) and are shown at the top of each panel. Quantitative data on the levels of particular transcripts are given in Lobner-Olsen *et al.* (2003).

except those for genes *cydD, trpE* and *asnA*). A similar distribution of GATC motifs occurs downstream of the bacteriophage λ p_I promoter that is stimulated by SeqA (Słomińska *et al.*, 2003a). Moreover, there are two neighboring GATC sequences between positions +150 and +200 in the regions of three additional promoters stimulated by SeqA (promoters for *fecI, dps* and *tpr*). Such double GATC motifs were found downstream of the SeqA-activated λ p_R and p_{aQ} promoters (Słomińska *et al.*, 2001, 2003a, 2003b). These types of distribution of SeqA-binding sequences are significantly less frequent in regions of promoters that are repressed by SeqA (Fig. 2B) or are SeqA-independent (Fig. 3). Therefore, one may speculate that

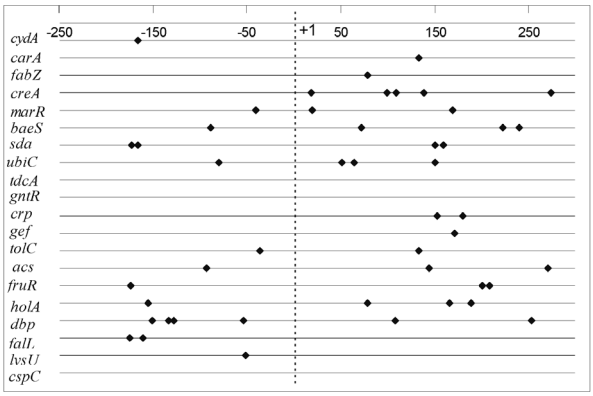

Figure 3. Distribution of GATC sequences (dots) in regions of randomly chosen promoters of genes for which mRNA levels in the Δ*seqA* mutant were found to be unaffected or changes relative to wild-type bacteria were less than two-fold (on the basis of the results reported by Lobner-Olsen *et al.*, 2003); in fact, most *E. coli* genes belong to this group.

Nucleotide positions (from –250 to +250) are relative to transcription start points (+1; dotted vertical line) and are shown at the top of the panel.

such patterns of GATC motifs are favorable for formation of SeqA-DNA complexes able to stimulate transcription from the nearby promoters. It is worth noting that these GATC motifs found near the SeqA-stimulated promoters generally do not follow the rules for SeqA binding proposed by Brendler & Austin (1999) on the basis of *in vitro* studies with oligonucleotides, namely the presence of at least two GATC sequences located on the same DNA side and separated by no more than 3 helical turns. However, it was demonstrated experimentally that when using longer DNA fragments, SeqA can bind to GATC regions distributed in a completely different manner than that suggested by Brendler & Austin (1999) as a requirement for SeqA–DNA interactions (Słomińska *et al.*, 2001; 2003a; 2003b). Interestingly, many promoters that are not significantly affected by SeqA have relatively many GATC motifs in their regions while others are devoid of such sequences.

It is plausible that the location of the GATC motif(s) relative to a promoter, described above, results in stimulation of transcription initiation by SeqA. Nevertheless, other promoters, which do not reveal characteristic patterns of GATC motifs, might be depressed in *seqA* mutants. There are several possible reasons that could explain this phenomenon. First, SeqA-dependent regulation of transcription due to changes in nucleoid structure, suggested by Lobner-Olsen *et al.* (2003), cannot be excluded. Second, the regulation of transcription of some genes by SeqA may be indirect, due to putative SeqA-mediated control of expression of other genes, whose products might be involved in transcription of those showed in Fig. 2A and 2B (note that in this figure, only genes whose transcript levels differed between the wild-type and Δ*seqA* strains at least two times are shown, and even minor changes in expression of genes coding for regulatory proteins may have significant effects on expression of other genes, controlled by these proteins). Third, although only a few examples of specific SeqA-mediated activation of promoters have been reported to date, it is clear that there are various mechanisms of regulation of transcription by this protein. For example, bacteriophage λ p_R promoter seems to be directly acti-

vated by SeqA (Słomińska *et al.*, 2001; 2003b), whereas *seqA* gene product stimulates the activity of p_I and p_{aQ} promoters by facilitating action of another transcription activator, the CII protein (Słomińska *et al.*, 2003a). Therefore, various locations of GATC motifs may be necessary for different mechanisms of SeqA-mediated transcription stimulation. We conclude that it is likely that for direct stimulation of a promoter by SeqA, the presence of a GATC motif (followed by another such motif) in the +100 region, or a couple of GATC sequences in a region between +100 and +200, is required. Activation of other promoters shown in Fig. 2A is perhaps indirect, for instance due to changes in nucleoid structure.

The results of the microarray analysis performed by Lobner-Olsen *et al.* (2003) have one more implication. It was demonstrated previously that the properties of cellular membranes are significantly changed in *seqA* mutants (Węgrzyn *et al.*, 1999). One possibility was that expression of genes coding for major membrane proteins, or genes whose products are involved in production or degradation of other membrane compounds, is affected in the absence of active SeqA. However, there are no such genes among those significantly stimulated or repressed by SeqA (Fig. 2). Therefore, other mechanism(s) must be responsible for the changes in cell membranes observed in *seqA* mutants.

REFERENCES

Boye E, Stokke T, Kleckner N, Skarstad K. (1996) *Proc Natl Acad Sci U S A.*; **93**: 12206–11.

Brendler T, Austin S. (1999) *EMBO J.*; **18**: 2304–10.

Kang S, Lee H, Han JS, Hwang DS. (1999) *J Biol Chem.*; **274**: 11463–8.

Lobner-Olsen A, Marinus MG, Hansen FG. (2003) *Proc Natl Acad Sci U S A.*; **100**: 4672–7.

Lu M, Campbell JL, Boye E, Kleckner N. (1994) *Cell.*; **77**: 413–26.

Skarstad K, Lueder G, Lurz R, Speck C, Messer W. (2000) *Mol Microbiol.*; **36**: 1319–26.

Slater S, Wold S, Lu M, Boye E, Skarstad K, Kleckner N. (1995) *Cell.*; **82**: 927–36.

Słominska M, Węgrzyn A, Konopa G, Skarstad K, Węgrzyn G. (2001) *Mol Microbiol.*; **40**: 1371–9.

Słomińska M, Konopa G, Ostrowska J, Kędzierska B, Węgrzyn G, Węgrzyn A. (2003a) *Mol Microbiol.*; **47**: 1669–79.

Słomińska M, Konopa G, Barańska S, Węgrzyn G, Węgrzyn A. (2003b) *J Mol Biol.*; **329**: 59–68.

von Freiesleben U, Rasmussen KV, Schaechter M. (1994) *Mol Microbiol.*; **14**: 763–72.

Węgrzyn A, Wróbel B, Węgrzyn G. (1999) *Mol Gen Genet.*; **261**: 762–9.

Wold S, Boye E, Slater S, Kleckner N, Skarstad K. (1998) *EMBO J.*; **17**: 4158–65.