DNA sequence and transcription of the region upstream of the *E. coli gyr*B gene

Toshiro Adachi*, Kiyoshi Mizuuchi, Rolf Menzel and Martin Gellert

Laboratory of Molecular Biology, National Institute of Arthritis, Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20205, USA

Received 19 June 1984; Accepted 17 July 1984

ABSTRACT

We have determined the sequence of a 1498 base-pair region in E. coli that extends from within dnaN through recF and into the gyrB gene. An open reading frame of 1071 base pairs has been identified with the recF structural gene. By S1 mapping, we have located a transcription start point 31 base pairs upstream of gyrB. The amount of this transcript is much greater in cells that have been treated with novobiocin, a treatment which is known to induce greater synthesis of DNA gyrase.

INTRODUCTION

This paper is one in a series of studies of bacterial DNA superhelicity. DNA supercoiling in bacteria serves, among other things, to modulate the expression of many genes (for reviews see 1, 2), increasing transcription from some promoters while decreasing that from others. The superhelicity of bacterial DNA is itself controlled by a balance of the DNA-supercoiling activity of DNA gyrase and the DNA-relaxing activities of other topoisomerases, principally topoisomerase I (3,4,5). One factor contributing to this balance is the level of expression of DNA gyrase, which is strongly influenced by DNA supercoiling. When the template DNA is relaxed, expression of both the gyrA and gyrB genes is greatly increased, thus tending to restore a higher level of DNA supercoiling (6).

In order to study this mode of regulation in more detail, we have determined the DNA sequence upstream of the gyrB gene; this paper reports the sequence and some studies of transcription in this region. The gyrB gene, at 83 min on the E. coli map, is contained in a cluster of genes related to DNA metabolism; the gene order, in the counter-clockwise direction on the map, is dnaA dnaN recF gyrB. The DNA sequences of dnaA (7,8) and dnaN (8) have been reported, and a fragment around the start of gyrB was previously sequenced in this laboratory (9). The sequence reported in this paper extends into dnaN and connects with the sequence at the upstream end of gyrB. In the process,

an open reading frame of 1071 nucleotides has been noticed, which can plausibly be identified as corresponding to the RecF protein.

Analysis of transcripts by S1 mapping has been used to identify an RNA initiation site 31 base pairs upstream of gyrB. In vivo transcription from this promoter increases upon addition of novobiocin in a manner parallel to the induction of GyrB protein synthesis.

MATERIALS AND METHODS

Restriction enzymes were obtained from New England Biolabs or Bethesda Research Laboratories, polynucleotide kinase and deoxynucleotidyl terminal transferase from P-L Biochemicals, and calf intestinal phosphatase from Boehringer-Mannheim. Plasmid pMK64 is a derivative of pBR322 containing a 2.2 kb EcoR1 fragment which spans the recF gene and extends through most of dnaN on one side and just into gyrB on the other. Plasmid pMK47 contains the entire dnaN recF gyrB region and most of dnaA (10).

E. coli strain N99 galK strA su$^-$ and its derivative RM161 galK strA su$^-$ recA::Tn10 (pMK47) were used for the preparation of messenger RNA.

A restriction map (Fig. 1) of the region of interest was constructed by partially digesting 5'-end labeled DNA and sizing the labeled fragments by gel electrophoresis. The locations of restriction sites were later confirmed by sequencing. The restriction enzymes shown in Fig. 1 are those which were used to prepare fragments for sequencing.

DNA sequences were determined by the Maxam-Gilbert method (11). The fragments were labeled at their 5' ends using polynucleotide kinase and γ-[^{32}P]-ATP; two that were labeled at their 3' ends with ^{32}P-cordycepin triphosphate by the action of deoxynucleotidyl terminal transferase are indicated separately in Figure 1.

E. coli cells for the preparation of mRNA were grown in LB broth plus 0.05 M Tris-HCl (pH 7.5). Novobiocin, when used, was added to a final concentration of 2 mg/ml from a 100 mg/ml stock solution; growth was stopped 50 min later. Preparations of mRNA were made by the method of Hagen and Young (12). S1 mapping of transcripts was done by a modification (13) of the method of Berk and Sharp (14).

RESULTS

The sequence of 1498 bases reported here (Figure 2) spans a region whose left end lies within the dnaN gene and whose right end is within gyrB. The dnaN open reading frame, corresponding to a protein of 366 amino acids (in

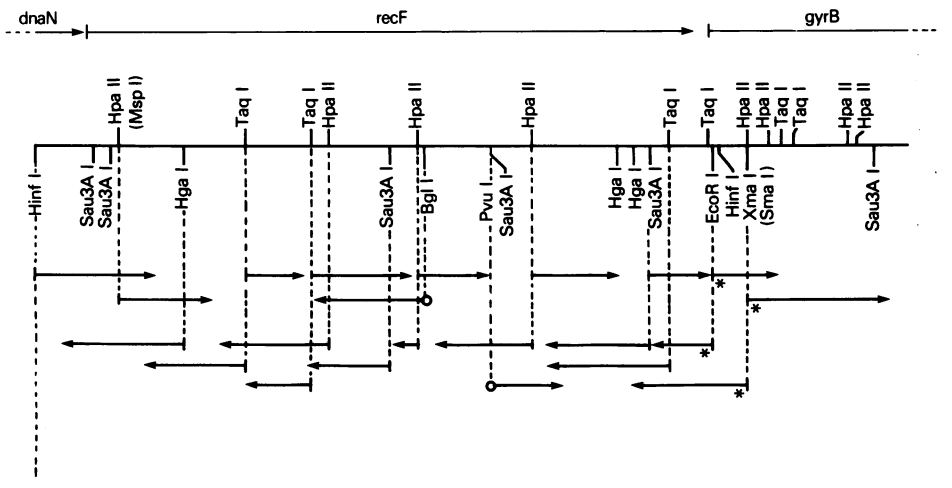

Figure 1. Restriction map of the region that was sequenced, extending from within dnaN on the left, through recF, and into gyrB on the right. The arrows indicate the direction and extent of the sequence determined on each fragment, with the tail of the arrow at the 5'-labelled end. An arrow with an open circle indicates 3'-labelling at that end. Fragments sequenced from plasmid pMK47 are marked with an asterisk.

agreement with the reported molecular weight of 37 to 45 kD (15,16,17)) has been identified in a sequence determined by Ohmori et al. (8). The open reading frame of dnaN in Ohmori's sequence ends with the TAA triplet at position 76-78 in our sequence. At the right end, the sequence of an open reading frame that starts at position 1180 agrees with the N-terminal amino-acid sequence of the GyrB protein, except that the terminal methionine is missing (9).

Independently of the present work, a sequence spanning most of the same region has recently been determined by Blanar et al. (18). The two sequences are in exact agreement in the area of overlap (0-1190 in our numbering). The region sequenced by Ohmori et al. (8) extends 384 bp beyond the end of dnaN; this sequence also agrees with ours where they overlap.

Between dnaN and gyrB, the only gene that has been identified is recF (19). RecF has also been localized to a 1.65 kb fragment spanning this region (20). Our sequence contains in this region a single long open reading frame of 1071 bp that starts at position 78, overlapping the dnaN stop codon, and ends at position 1148, 31 bp upstream of gyrB. This region contains no other open reading frame longer than 165 bp in either direction, even if GTG starts are allowed. Identification of the long open reading frame with recF

10
20
30
40
50
60
70
80
90
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400

has been made more definitively by Blanar et al. (18), who have found in maxicell experiments a protein band of 40 kD corresponding to recF, and have shown that recF function is inactivated by Tn 3 insertions throughout this region. Thus the sequence of the entire dnaA dnaN recF region is now known.

Figure 3 points out a feature of interest in the sequence: the rather high frequency of self-complementary sequences. Whether there is any relation between these possible structures and the poor expression of the recF gene (18) remains to be determined. Work to be presented elsewhere (R. Menzel and M. Gellert, in preparation) indicates that sequences upstream of the end of recF function as strong transcription terminators.

Four adjacent genes, dnaA, dnaN, recF, and gyrB are read in the same direction. The pattern of transcription in this block of genes is not yet completely worked out. Two transcripts that start upstream of dnaA extend through dnaN (7,21); there appears to be no promoter for dnaN itself (8). It is not known whether either of these transcripts covers recF, or whether there is a separate promoter for recF.

Because of our interest in the expression of DNA gyrase, we have mapped transcripts that extend into the gyrB gene. We previously identified an RNA species that begins 30 to 40 bases before gyrB (9). In the present work we have refined the mapping of this RNA start point, and have also investigated the induction of this RNA species when the cells are treated with novobiocin. The S1 mapping experiment of Figure 4 shows that cells containing plasmid pMK47 produce a major RNA species which starts at position 1149, or 31 bases before the initiation codon of gyrB. This position has an appropriate spacing from a Pribnow box sequence (TAAAAT, at 1136-1141 bp); it implies that the RNA starts with a T, which is relatively uncommon.

Figure 4 also shows that this RNA species is several-fold more abundant in cells treated with novobiocin. It is known that synthesis of both subunits of DNA gyrase is induced by conditions (addition of novobiocin or coumermycin A_1, or use of a temperature-sensitive gyrase mutant) which block DNA gyrase activity and thereby cause the relaxation of cellular DNA (6). The present results imply that this induction takes place at the level of transcription. In experiments to be published elsewhere (R. Menzel and M. Gellert, manuscript

Figure 2. Sequence of the E. coli region upstream of gyrB (and part of (gyrB). The dnaN open reading frame ends at bp 75, the open reading frame identified as recF extends from bp 78 to bp 1148, and the gyrB open reading frame starts at bp 1180.

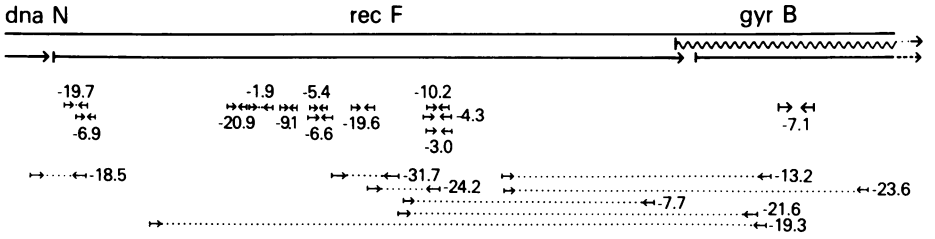

Figure 3. Potential intra-strand base pairing in the dnaN-recF-gyrB region. Self-complementary sequences are indicated by head-to-head pairs of short arrows, in some cases connected by dotted lines. The figures show the expected stabilizing free energy of the structure in Kcal/mol, calculated as described by Tinoco et al. (22). Open reading frames (long straight arrows) and the start of the gyrB transcript (wavy arrow) are also indicated.

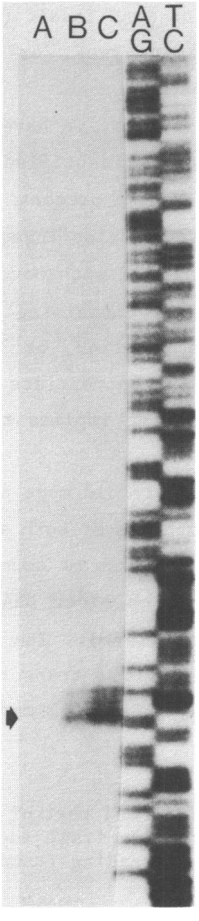

Figure 4. S1 mapping of the start of the gyrB transcript. RNA samples were as follows: lane A, RNA from strain N99; lane B, RNA from strain RM161 which carries plasmid pMK47 containing the dnaN recF gyrB region; lane C, RNA from novobiocin-treated RM161 (see Methods). The DNA probe was a single-stranded fragment labelled with ^{32}P at its 5' end at the XmaI site (position 1249) and extending to the PvuI site at bp 800. The last two lanes are the (A + G) and (C + T) Maxam-Gilbert sequencing reaction products of the same fragment. In determining the length of the protected DNA fragment in lanes A-C, one must correct for the fact that the sequencing products are identified with the 3'-terminal base which has been removed, and that they contain a 3'-phosphoryl group. Thus the S1-resistant DNA must be compared with a sequence marker 1.5 bases shorter than its apparent position (23).

in preparation), we have shown that the region immediately surrounding the promoter for this RNA species contains all the information necessary for transcription to be activated by DNA relaxation.

ACKNOWLEDGEMENT

We thank A. J. Clark and H. Ohmori for communicating results in advance of publication.

*Permanent address: Toyobo Co., Ltd., Katata Research Institute, Ohtsu, Shiga 520-02, Japan,

REFERENCES

1. Gellert, M. (1981) Annu. Rev. Biochem. 50, 879-910.
2. Cozzarelli, N.R. (1980) Science 207, 953-960.
3. Sternglanz, R., DiNardo, S., Voelkel, K.A., Nishimura, Y., Hirota, Y., Becherer, K., Zumstein, L., and Wang, J.C. (1981) Proc. Natl. Acad. Sci. USA 78, 2747-2751.
4. DiNardo, S., Voelkel, K.A., Sternglanz, R., Reynolds, A.E., and Wright, A. (1982) Cell 31, 43-51.
5. Pruss, G.J., Manes, S.H., and Drlica, K. (1982) Cell 31, 35-42.
6. Menzel, R., and Gellert, M. (1983) Cell 34, 105-113.
7. Hansen, E.B., Hansen, F.G., and von Meyenburg, K. (1982) Nucleic Acids Res. 10, 7373-7385.
8. Ohmori, H., Kimura, M., Nagata, T., and Sakakibara, Y. (1984) Gene 28, 159-170.
9. Gellert, M., Menzel, R., Mizuuchi, K., O'Dea, M.H., and Friedman, D.I. (1983) Cold Spring Harbor Symp. Quant. Biol. 47, 763-767.
10. Mizuuchi, K., Mizuuchi, M., O'Dea, M.H., and Gellert, M. (1984) J. Biol. Chem., in press.
11. Maxam, A.M., and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
12. Hagen, F., and Young, E.T. (1978) J. Virol. 26, 793-804.
13. Weaver, R.F., and Weissmann, C. (1979) Nucleic Acids Res. 7, 1175-1193.
14. Berk, A.J., and Sharp, P.A. (1977) Cell 12, 721-732.
15. Hansen, F.G., and von Meyenburg, K. (1979) Mol. Gen. Genet. 175, 135-144.
16. Yuasa, S., and Sakakibara, Y. (1980) Mol. Gen. Genet. 180, 267-273.
17. Johanson, J.O., and McHenry, C.S. (1980) J. Biol. Chem. 255, 10984-10990.
18. Blanar, M.A., Sandler, S.J., Armengod, M.-E., Ream, L.W., and Clark, A.J. (1984) Proc. Natl. Acad. Sci. USA, in press.
19. Ream, L.W., Margossian, L., Clark, A.J., Hansen, F.G., and von Meyenburg, K. (1980) Mol. Gen. Genet. 180, 115-121.
20. Ream, L.W., and Clark, A.J. (1983) Plasmid 10, 101-110.
21. Sakakibara, Y., Tsukano, H., and Sako, T. (1981) Gene 13, 47-55.
22. Tinoco, I., Jr., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M., and Gralla, J. (1973) Nature (London) New Biol. 246, 40-41.
23. Fisher, L.M., Mizuuchi, K., O'Dea, M.H., Ohmori, H., and Gellert, M. (1981) Proc. Natl. Acad. Sci. USA 78, 4165-4169.