JOURNAL OF BACTERIOLOGY, Apr. 1985, p. 271–275
0021-9193/85/040271-05$02.00/0

Vol. 162, No. 1

Regulatory Region of the Heat Shock-Inducible *capR* (*lon*) Gene: DNA and Protein Sequences

RANDALL C. GAYDA,[1]† PAUL E. STEPHENS,[2] RODNEY HEWICK,[3]‡ JOYCE M. SCHOEMAKER,[2]§ WILLIAM J. DREYER,[3] AND ALVIN MARKOVITZ[1]*

Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois 60637[1]; Department of Molecular Genetics, Celltech Ltd., Slough SL1 4DY, England[2]; and Division of Biology, California Institute of Technology, Pasadena, California 91109[3]

Received 22 August 1984/Accepted 4 January 1985

The CapR protein is an ATP hydrolysis-dependent protease as well as a DNA-stimulated ATPase and a nucleic acid-binding protein. The sequences of the 5′ end of the *capR* (*lon*) gene DNA and N-terminal end of the CapR protein were determined. The sequence of DNA that specifies the N-terminal portion of the CapR protein was identified by comparing the amino acid sequence of the CapR protein with the sequence predicted from the DNA. The DNA and protein sequences established that the mature protein is not processed from a precursor form. No sequence corresponding to an SOS box was found in the 5′ sequence of DNA. There were sequences that corresponded to a putative −35 and −10 region for RNA polymerase binding. The *capR* (*lon*) gene was recently identified as one of 17 heat shock genes in *Escherichia coli* that are positively regulated by the product of the *htpR* gene. A comparison of the 5′ DNA region of the *capR* gene with that of several other heat shock genes revealed possible consensus sequences.

The *capR* (*lon*) gene of *Escherichia coli* has been cloned (37, 45), and the protein has been purified to homogeneity (7, 44). The protein is a tetramer (6), with subunits of 94 kilodaltons, that has multiple enzymatic activities: an ATP hydrolysis-dependent protease activity (7, 8), a DNA-stimulated ATPase activity (5), and a nucleic acid-binding activity (7, 44). In addition, a defective CapR protein (CapR9) has been purified. The CapR9 protein retains the general nucleic acid affinity (7, 44) but has lost both the protease activity (6, 7) and the DNA-stimulated ATPase activity (5). *capR* mutants display several phenotypic alterations, including overproduction of capsular polysaccharide (20, 23, 24), increased sensitivity to UV and ionizing radiation (1, 20, 25, 42), inhibition of cell division resulting in filament formation (1, 12, 37, 42), reduced lysogenization by bacteriophages lambda and P1 (24), and reduced degradation of abnormal (3, 14, 15, 38) and normal (11, 28, 35, 36) polypeptides. These pleiotropic effects may be because the proteolytic function of the CapR protein either activates or inactivates substrates. Recently, the CapR protein has been shown to belong to the unique group of proteins that are heat shock inducible (13, 31).

In this report we present the sequence of the *Eco*RI-*Hin*dIII fragment (590 bases long) that contains the regulatory region of the *capR* gene and the N-terminal end of the CapR protein. The regulatory region is of interest for several reasons. Several of the phenotypic effects of mutation in the *capR* gene (filamentation and UV sensitivity) are observed after induction of the SOS response (1, 22). Is the *capR* gene therefore under SOS control? Can a heat shock consensus sequence be found? Are there any unique controling sequences or secondary DNA structures that may provide a clue to cellular regulation of the *capR* gene?

MATERIALS AND METHODS

Restriction mapping by end-labeling. The *Eco*RI-*Sal*I DNA fragment of the *capR* gene previously subcloned into plasmid pJMC22 (37) was used. pJMC22 contains a 2.1-kilobase (kb) *Eco*RI-*Sal*I fragment of the *capR* gene plus a 2.7-kb *Eco*RI-*Eco*RI fragment of ColE1 and 3.7-kb *Eco*RI-*Sal*I fragment of pBR322 (37). Evidence that the *Eco*RI-*Sal*I fragment of the *capR* gene contains a heat shock promoter was provided in a recent study (13). There is only one *Hin*dIII restriction site in pJMC22, in the *capR* DNA fragment at 0.6 kb from the *Eco*RI site. To map restriction sites from this *Hin*dIII site, the 5′ ends were labeled with [α-^{32}P]ATP (27). *Hin*dIII-digested pJMC22 end labeled with α-^{32}P was then cut with *Eco*RI and *Sal*I, and the two labeled bands of 0.6 (*Eco*RI-*Hin*dIII) and 1.5 (*Hin*dIII-*Sal*I) kb were isolated on a 4% polyacrylamide gel (26). A sample that included the end-labeled DNA fragment and 1 μg of unlabeled pBR322 DNA was incubated with 0.5 U of each restriction enzyme for 1, 5, or 10 min at 37°C. Restriction enzymes were inactivated by incubation at 65°C for 10 min. The digests were analyzed by radioautography after electrophoresis through a 6% polyacrylamide gel (26). Kodak XAR X-ray film was used at room temperature.

Restriction enzymes *Ava*I, *Ava*II, *Hin*fI, *Hae*III, *Pst*I, *Dde*I, *Hpa*II, *Hha*I, *Sau*3A, and *Taq*I were used as suggested by the manufacturer.

DNA sequencing. DNA sequence analysis was performed by the dideoxy chain termination method (34) in conjunction with M13 cloning vehicles (27). DNA fragments isolated from low-melting-point agarose gels were cloned into appropriate vectors. A *Hin*dIII-*Eco*RI fragment of 590 base pairs was cloned (see Fig. 2) into both M13 mp8 and mp9 to obtain DNA sequence data from both strands. In addition, a *Sau*3A-*Eco*RI and a *Sau*3A-*Hin*dIII fragment were cloned into vector M13 mp8 cut with *Bam*HI and *Eco*RI and vector M13 mp11 cut with *Bam*HI and *Hin*dIII, respectively.

* Corresponding author.
† Present address: Department of Microbiology, Louisiana State University, Baton Rouge, LA 70803.
‡ Present address: Genetics Institute, 225 Longwood Ave., Boston, MA 02115.
§ Present address: Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637.

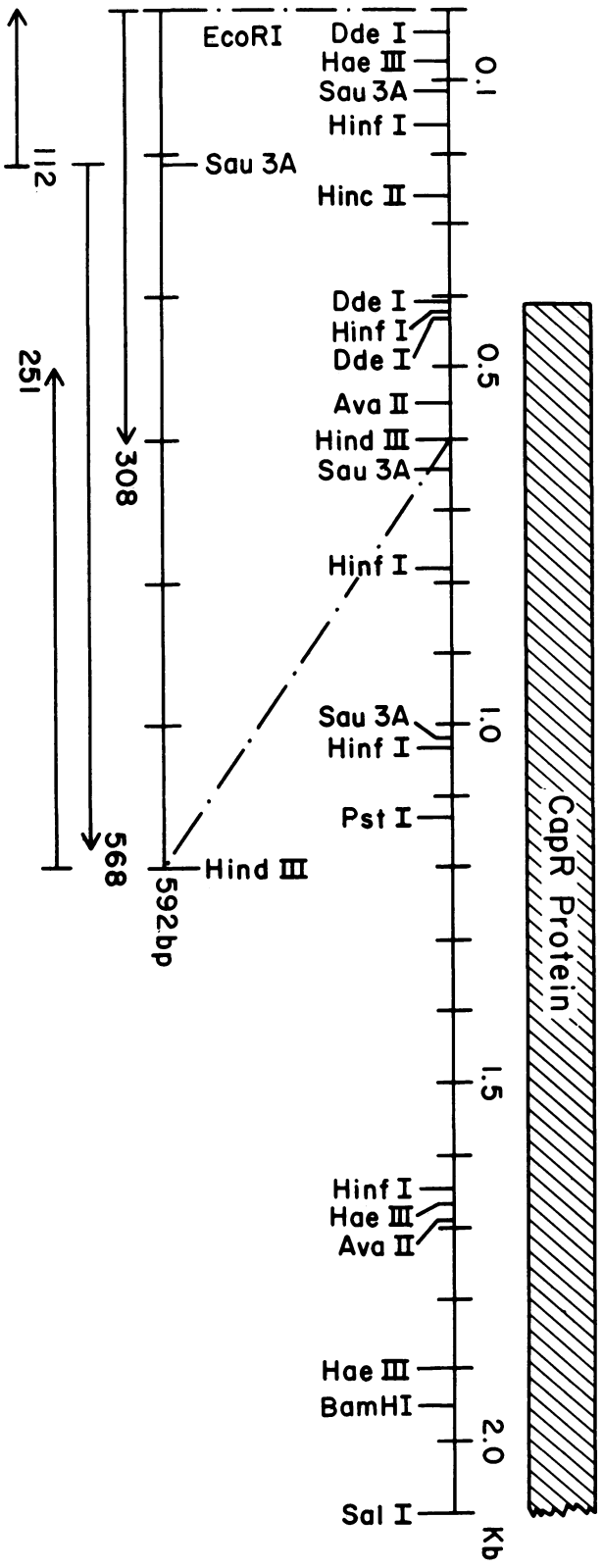

FIG. 1. Restriction endonuclease map of the *Eco*RI-*Sal*I DNA fragment of the *capR* (*lon*) gene and the sequencing strategy for the *Eco*RI-*Hin*dIII fragment (enlarged). The *Eco*RI-*Sal*I fragment specifies approximately 65% of the *capR* gene. The remainder of the gene is contained on a *Sal*I-*Pst*I fragment (37). The restriction map of the *Eco*RI to the *Hin*dIII site is based on the DNA sequence. The restriction map from *Hin*dIII to *Sal*I was determined from polyacrylamide gel analysis of end-labeled and restriction endonuclease-cut fragments. The DNA sequencing strategy is indicated by the direction of arrows below the enlarged *Eco*RI-*Hin*dIII fragment. The numbers listed near the arrows indicate distance (in base pairs) from the *Eco*RI site. Kb, Kilobases.

Recombinant phages were identified as colorless plaques on indicator plates containing 5-bromo-4-chloro-3-indolyl-β-galactoside after transfecting *E. coli* strain JM101 [Δ(*lac-pro*) *supE thi* F *traD36 proAB lac*I^{q}Z ΔM15] (27) as outlined by Sanger et al. (33).

An ambiguity in the DNA sequence between coordinates 40 and 60, due to compression on the polyacrylamide gel, was resolved by rerunning the samples on a polyacrylamide gel containing 25% formamide (A. Coulson, personal communication).

The nucleotide sequence was compiled and analyzed with the Staden computer programs (39). The *Eco*RI-*Hin*fI (0 to 154 base pairs) and the *Hin*fI-*Hin*dIII (421 to 592 base pairs) sequences (Fig. 1) were verified by the sequencing procedure of Maxam and Gilbert (26).

N-terminal protein sequencing. The CapR$^+$ and CapR9 protein preparations used for sequence determination were the purest preparations described (7). The protein samples were applied to the gas phase sequenator, and the amino acid sequences were determined (18). For determination of the N-terminal methionine of the CapR9 protein, the protein was first precipitated as described previously (41).

RESULTS

Previous studies identified the DNA region that contained the *capR* gene (23, 45) and localized the promoter region and direction of transcription and translation (37, 44). Those results were established as correct in the present study, as indicated in Fig. 1.

The restriction cleavage site map of the *Eco*RI-*Sal*I fragment of the *capR* gene and the sequencing strategy are shown in Fig. 1. It should be noted that the coding region of the CapR protein starts 418 bases from the *Eco*RI cleavage site.

The nucleotide sequence of the *Eco*RI-*Hin*dIII fragment and the amino acids of the N-terminal end of the CapR protein are shown in Fig. 2. The first 14 N-terminal amino acids of the CapR9 protein were also determined (Fig. 2). The same amino acid sequence was obtained as for the CapR protein except that amino acids 1, 12, and 14 were not identified due to technical difficulties. These amino acids correspond to the sequence predicted by the nucleotide sequence of the *capR*$^+$ gene DNA.

DISCUSSION

The N-terminal amino acid sequence of both the wild-type (CapR) and mutant (CapR9) proteins was the same as far as it was determined (Fig. 2). This finding is consistent with the fact that the *capR9* mutation mapped in the region of the DNA that specifies the C-terminus of the protein (37). Comparison of the actual N-terminal amino acid sequence of the CapR protein with all possible predicted amino acid sequences (determined from the nucleotide sequence) revealed the protein coding sequence (Fig. 2). From the amino acid and nucleotide sequences, it is clear that the mature CapR protein is not processed at the N-terminal end, i.e.,

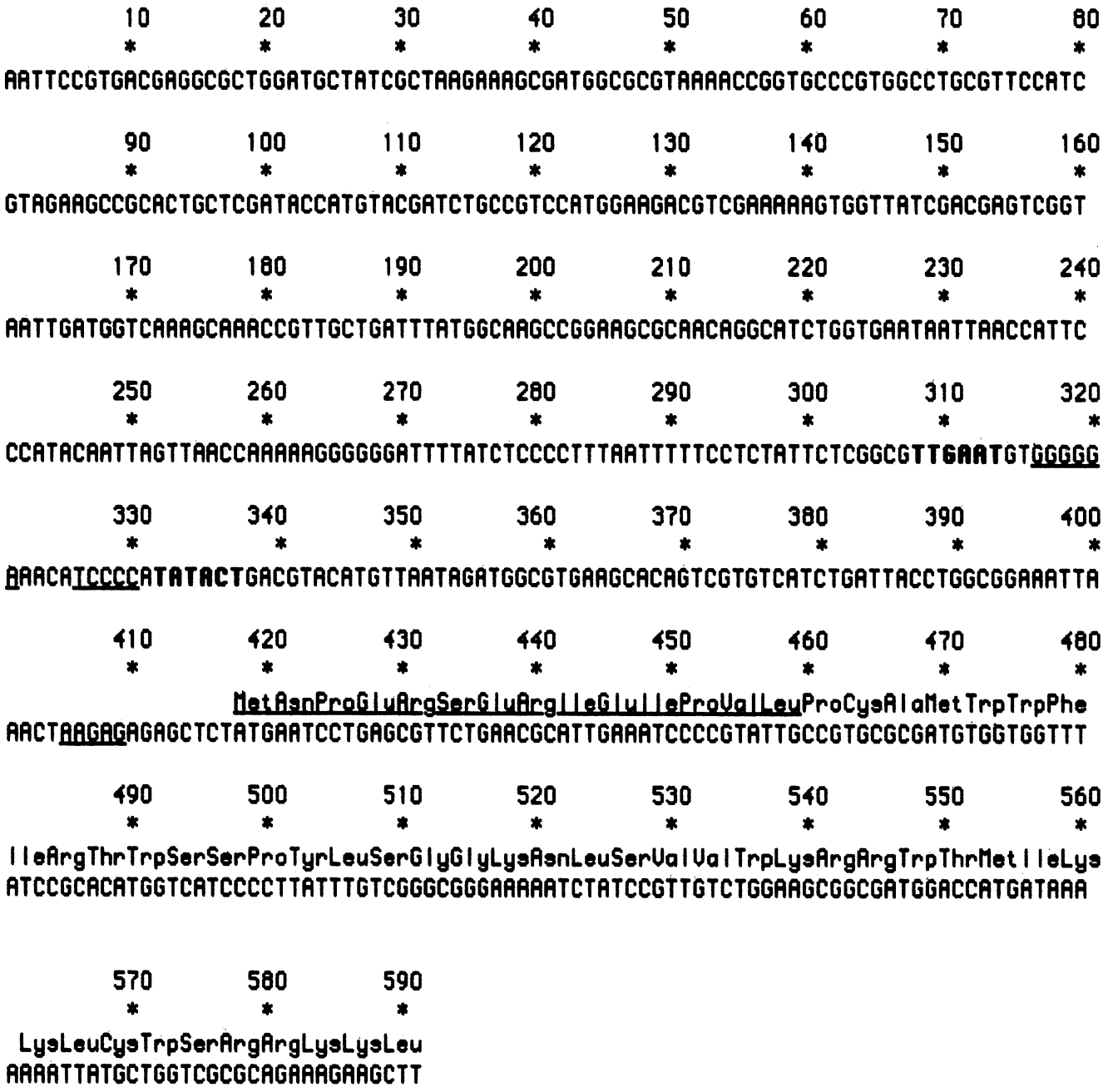

FIG. 2. Regulatory region of the *capR* (*lon*) gene: DNA and protein sequences. The amino acid sequence of the CapR9 protein determined by the sequenator is underlined. The same amino acid sequence was determined for the CapR protein, except that amino acids 1, 12, and 14 were not identified due to technical difficulties. The amino acids not underlined are predicted from the nucleotide sequence. The putative −35 (nucleotide 310) and −10 (nucleotide 335) RNA polymerase-binding sequences are in boldface type. The Shine and Dalgarno sequence (nucleotide 407) is underlined. Five consecutive G's and an A and a T and four consecutive C's between the −35 and −10 sequences that could form a stem and loop are also underlined.

there is no zymogen form as there is with many classic proteases. Zymogens serve the function of providing an essentially inactive precursor form as a mechanism of regulation. An alternative regulatory mechanism for the CapR protein may be its requirement for ATP hydrolysis (or other triphosphates or deoxytriphosphates) for its proteolytic activity (7; unpublished data). The RecA protein of *E. coli* also requires ATP to exhibit proteolytic activity, but in contrast to the CapR protein, ATP hydrolysis is not necessary (9). The RecA protein also does not have a zymogen form, although the mature protein has alanine at the N-terminus, indicating that the initiating methionine has been removed (19, 32).

Further inspection of the *capR* gene DNA sequence reveals a putative Shine and Dalgarno sequence, AAGAG, eight nucleotides 5′ to the ATG initiation codon of the gene (Fig. 2). Also, putative −35 and −10 sequences for RNA polymerase binding are present at position 310 and 335,

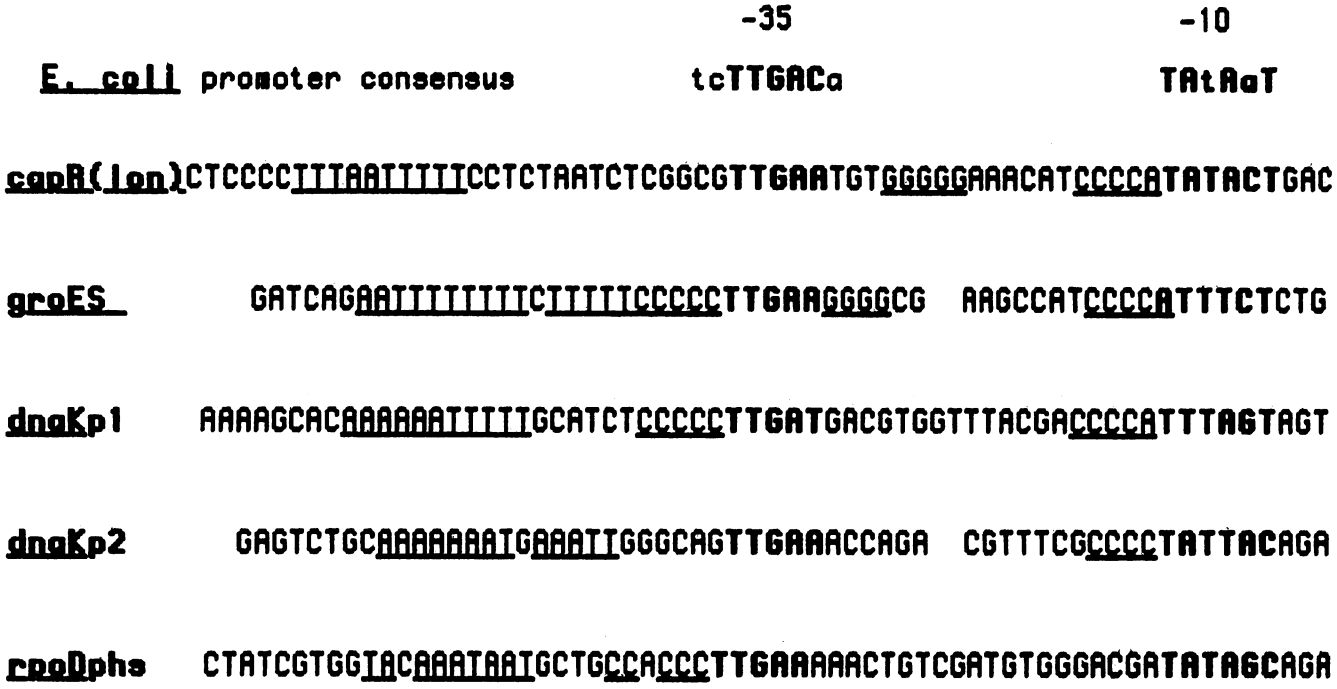

FIG. 3. Comparison of the DNA sequences around the −35 region of the known heat shock genes of *E. coli;* (30, 40). The *groES* DNA sequence was provided by R. Hendrix and C. Woolford (unpublished data), and the $dnaKp_1$ and $dnaKp_2$ sequences were provided by D. Cowing, J. Bardwell, E. Craig, and C. Gross (unpublished data). The *rpoD*phs DNA sequence has been published (4) and was determined to be a promoter by promoter cloning and S1 nuclease mapping (40). The −35 and −10 sequences (boldface type) for all putative promoters (except *rpoD*phs) were determined from comparison with the *E. coli* consensus sequences shown at the top of the figure (17). The A + T-rich regions 5′ to the −35 region are underlined, as are the multiple C's 5′ to the −35 and −10 regions. The multiple G's in the *capR* and *groES* genes are also underlined.

respectively (Fig. 2). An SOS consensus sequence [CTGT (9N) CAG (22)] is not present. An "approximated SOS sequence" starting at base 186 (Fig. 2) (CTGATTTATG GCAAG) is present. However, experimental conditions that induced two SOS proteins, the RecA and SulA proteins (nalidixic acid [35]), did not induce the CapR protein as determined by two-dimensional gel analysis (unpublished data). A cyclic AMP receptor protein consensus sequence (10) was not present in the regulatory region of the *capR* DNA.

Recent studies have determined that the *capR* gene is one of 17 heat shock genes (13, 30, 31). The DNA sequences of a few of these genes have been determined (Fig. 3). The sequences are limited to the putative RNA polymerase-binding regions. There are two general features that can be observed in the five sequences. First, there are either four or five C's immediately 5′ to the −10 region (*capR* and the second promoter of *dnaK* [$dnaKp_2$]), to the −35 region (the *rpoD* heat shock promoter [*rpoD*phs]) or to both the −10 and −35 regions (*groES* and $dnaKp_1$). Second, there is an A + T-rich sequence of 10 to 17 nucleotides 5′ to the −35 regions (underlined in Fig. 3). We also note that between the −35 and −10 regions, there are five (*capR*) and four (*groES*) consecutive G's that could form a stem and loop with repeats of the C's previously noted.

Perhaps some of the sequences underlined in Fig. 3 represent binding or recognition sites for the HtpR protein, a sigma factor that promotes transcription initiation at heat shock promoters (16, 21, 29, 43). Recently, the heat shock promoter of *rpoD* has been localized within the adjacent *dnaG* structural gene by promoter cloning and S1 nuclease mapping. Specifically, deletion analysis indicated that the C-rich region in the −45 to −35 region of the promoter was necessary for full heat shock activity (40).

The physiological role of the CapR protein in the heat shock response is not known. It probably is not critical for cell survival because neither we nor others have observed that *lon* mutant strains are particularly sensitive to growth at high temperature (see reference 31 for a discussion). Recent attempts to assess the contribution of the CapR protease to proteolysis in the heat shock response *vis à vis* the HtpR protein (which could also have a proteolytic role independent of the CapR protease) have led to different conclusions (2, 13). The CapR protein may be involved with degradation of the denatured proteins that result from exposure of cells to high temperature. It could also have a role in the recovery phase of the heat shock response by proteolyzing other heat shock proteins.

ACKNOWLEDGMENTS

We thank Spencer Emtage for providing DNA sequencing facilities at Celltech Ltd. and Robert Heinrikson and Roger Poorman for protein sequencing facilities at the University of Chicago. We thank Gerald Gordon for assistance with DNA sequencing in Chicago and Marc Charette and Gordon Henderson for purifying the CapR and CapR9 proteins. We thank Carol A. Gross, Roger Hendrix, Carol Woolford, and Frederick C. Neidhardt for providing the unpublished sequence data shown in Fig. 3 and for providing manuscripts before publication.

This work was supported by Public Health Service grant AI 06966 from the National Institute of Allergy and Infections Disease (to A.M.).

LITERATURE CITED

1. **Adler, H. I., and A. A. Hardigree.** 1964. Analysis of a gene controlling cell division and sensitivity to radiation in *Esche-*

richia coli. J. Bacteriol. **87**:720–726.
2. **Baker, T. A., A. D. Grossman, and C. A. Gross.** 1984. A gene regulating the heat shock response in *Escherichia coli* also affects proteolysis. Proc. Natl. Acad. Sci. U.S.A. **81**:6779–6783.
3. **Bukhari, A. I., and D. Zipser.** 1973. Mutants of *Escherichia coli* with a defect in the degradation of nonsense fragments. Nature (London) New Biol. **243**:238–241.
4. **Burton, Z. F., C. A. Gross, K. K. Watanabe, and R. R. Burgess.** 1983. The operon that encodes the sigma subunit of RNA polymerase also encodes ribosomal protein S21 and DNA primase in *E. coli* K12. Cell **32**:335–349.
5. **Charette, M. F., G. W. Henderson, L. L. Doane, and A. Markovitz.** 1984. DNA-stimulated ATPase activity of the Lon (CapR) protein. J. Bacteriol. **158**:195–201.
6. **Charette, M. F., G. W. Henderson, F. J. Kezdy, and A. Markovitz.** 1982. Molecular mechanism for dominance of a mutant allele of an ATP-dependent protease. J. Mol. Biol. **162**:503–510.
7. **Charette, M. F., G. W. Henderson, and A. Markovitz.** 1981. ATP hydrolysis-dependent protese activity of the *Ion(capR)* protein of *Escherichia coli* K-12. Proc. Natl. Acad. Sci. U.S.A. **78**:4728–4732.
8. **Chung, C. H., and A. L. Goldberg.** 1981. The product of the *Ion(capR)* gene in *E. coli* is the ATP-dependent protease La. Proc. Natl. Acad. Sci. U.S.A. **78**:4931–4935.
9. **Craig, N. L., and J. W. Roberts.** 1981. Function of nucleoside triphosphates and polynucleotide in *E. coli* RecA protein-directed cleavage of phage lambda repressor. J. Biol. Chem. **256**:8039–8044.
10. **de Crombrugghe, B., S. Busby, and H. Buc.** 1984. Cyclic AMP receptor protein: role in transcription activation. Science **224**:831–838.
11. **Gayda, R. C., H. Avni, P. E. Berg, and A. Markovitz.** 1979. Outer membrane protein a and other polypeptides regulate capsular polysaccharide synthesis in *E. coli* K-12. Mol. Gen. Genet. **175**:325–332.
12. **Gayda, R. C., L. T. Yamamoto, and A. Markovitz.** 1976. Second-site mutations in *capR* (*lon*) strains of *Escherichia coli* that prevent radiation sensitivity and allow bacteriophage lambda to lysogenize. J. Bacteriol. **127**:1208–1216.
13. **Goff, S. A., L. P. Casson, and A. L. Goldberg.** 1984. Heat shock regulatory gene *htpR* influences rates of protein degradation and expression of the *lon* gene in *E. coli*. Proc. Natl. Acad. Sci. U.S.A. **81**:6647–6651.
14. **Gottesman, S., and D. Zipser.** 1978. Deg phenotype of *Escherichia coli lon* mutants. J. Bacteriol. **133**:844–851.
15. **Grossman, A. D., R. R. Burgess, W. Walter, and C. A. Gross.** 1983. Mutations in the *lon* gene of *E. coli* K-12 phenotypically suppress a mutation in the sigma subunit of RNA polymerase. Cell **32**:151–159.
16. **Grossman, A. D., J. W. Erickson, and C. A. Gross.** 1984. The *htpR* gene product of *E. coli* is a sigma factor for heat-shock promoters. Cell **38**:383–390.
17. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *E. coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237–2255.
18. **Hewick, R. M., M. W. Hunkapiller, L. E. Hood, and W. J. Dreyer.** 1981. A gas-liquid solid phase peptide and protein sequenator. J. Biol. Chem. **256**:7990–7997.
19. **Horii, T., T. Ogawa, and H. Ogawa.** 1980. Organization of the recA gene of *E. coli*. Proc. Natl. Acad. Sci. U.S.A. **77**:313–317.
20. **Howard-Flanders, P., E. Simson, and K. Theriot.** 1964. A locus that controls filament formation and sensitivity to radiation in *Escherichia coli* K12. Genetics **49**:237–246.
21. **Landick, R., V. Vaughn, E. T. Lau, R. A. Van Bogelen, J. W. Erickson, and F. C. Neidhardt.** 1984. Nucleotide sequence of the heat shock regulatory gene of *E. coli* suggests its protein product may be a transcription factor. Cell **38**:175–182.
22. **Little, J. W., and D. W. Mount.** 1982. The SOS regulatory system of *Escherichia coli*. Cell **29**:11–22.
23. **Markovitz, A.** 1964. Regulatory mechanisms for synthesis of capsular polysaccharide in mucoid mutants of *E. coli* K12. Proc. Natl. Acad. Scid. U.S.A. **51**:239–246.
24. **Markovitz, A.** 1977. Genetics and regulation of capsular polysaccharide biosynthesis and radiation sensitivity, p. 415–462. *In* I. W. Sutherland (ed.), Surface carbohydrates of the procaryotic cell. Academic Press, Inc., New York.
25. **Markovitz, A., and N. Rosenbaum.** 1965. A regulator gene that is dominant on an episome and recessive on a chromosome. Proc. Natl. Acad. Sci. U.S.A. **54**:1084–1091.
26. **Maxam, A. M., and W. Gilbert.** 1979. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
27. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–77.
28. **Mizusawa, S., and S. Gottesman.** 1983. Protein degradation in *E. coli*: the *lon* gene controls the stability of the sulA protein. Proc. Natl. Acad. Sci. U.S.A. **80**:358–362.
29. **Neidhardt, F. C., R. A. VanBogelen, and E. T. Law.** 1983. Molecular cloning and expression of a gene that controls the high-temperature regulon of *E. coli*. J. Bacteriol. **153**:597–603.
30. **Neidhardt, F. C., R. A. VanBogelen, and V. Vaughn.** 1984. Genetics and regulation of heat shock proteins. Annu. Rev. Genet. **18**:295–329.
31. **Phillips, T. A., R. A. VanBogelen, and F. C. Neidhardt.** 1984. *lon* gene product of *Escherichia coli* is a heat-shock protein. J. Bacteriol. **159**:283–287.
32. **Sancar, A., C. Stachelek, W. Konigsberg, and W. D. Rupp.** 1980. Sequences of the recA gene and protein. Proc. Natl. Acad. Sci. U.S.A. **77**:2611–2615.
33. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**:161–178.
34. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
35. **Schoemaker, J. M., R. C. Gayda, and A. Markovitz.** 1984. Regulation of cell division in *Escherichia coli*: SOS induction and cellular location of the SulA protein, a key to *lon*-associated filamentation and death. J. Bacteriol. **158**:551–561.
36. **Schoemaker, J. M., G. W. Henderson, and A. Markovitz.** 1982. An *Escherichia coli* polypeptide controlled by the *lon* (*capR*) ATP hydrolysis-dependent protease and possibly involved in cell division. J. Bacteriol. **152**:919–923.
37. **Schoemaker, J. M., and A. Markovitz.** 1981. Identification of the gene *lon* (*capR*) product as a 94-kilodalton polypeptide by cloning and deletion analysis. J. Bacteriol. **147**:46–56.
38. **Shineberg, J. B., and D. Zipser.** 1973. The *lon* gene and degradation of β-galactosidase nonsense fragments. J. Bacteriol. **116**:1469–1471.
39. **Staden, R.** 1980. A new computer method for the storage and manipulation of DNA gel reading data. Nucleic Acids Res. **8**:3673–3694.
40. **Taylor, W. E., D. B. Straus, A. D. Grossman, Z. F. Burton, C. A. Gross, and R. R. Burgess.** 1984. Transcription from a heat-inducible promoter causes heat shock regulation of the sigma subunit of *E. coli* RNA polymerase. Cell **38**:371–381.
41. **Wessel, D., and U. L. Flügge.** 1984. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. Anal. Biochem. **138**:141–143.
42. **Witkin, E. M.** 1947. Inherited differences in sensitivity to radiation of *E. coli*. Genetics **32**:221–248.
43. **Yura, T., T. Tobe, K. Ito, and T. Osawa.** 1984. Heat shock regulatory gene (htpR) of *Escherichia coli* is required for growth at high temperature but is dispensable at low temperature. Proc. Natl. Acad. Sci. U.S.A. **81**:6803–6807.
44. **Zehnbauer, B. A., E. C. Foley, G. W. Henderson, and A. Markovitz.** 1981. Identification and purification of the *lon*$^+$(*capR*$^+$) gene product, a DNA-binding protein. Proc. Natl. Acad. Sci. U.S.A. **78**:2043–2047.
45. **Zehnbauer, B. A., and A. Markovitz.** 1980. Cloning of the gene *lon* (*capR*) of *Escherichia coli* K-12 and identification of polypeptides specified by the cloned deoxyribonucleic acid fragments. J. Bacteriol. **143**:853–863.