Cell, Vol. 38, 175-182, August 1984, Copyright © 1984 by MIT

0092-8674/84/080175-08 $02.00/0

Nucleotide Sequence of the Heat Shock Regulatory Gene of E. coli Suggests Its Protein Product May Be a Transcription Factor

Robert Landick,[†§] Vicki Vaughn,* Elizabeth T. Lau,* Ruth A. VanBogelen,* James W. Erickson,[‡] and Frederick C. Neidhardt*
*Department of Microbiology and Immunology
[†]Department of Biological Chemistry
University of Michigan Medical School
Ann Arbor, Michigan 48109
[‡]Department of Bacteriology
University of Wisconsin
Madison, Wisconsin 53706

Summary

We have sequenced a cloned segment of E. coli chromosomal DNA that includes the heat shock regulatory gene *htpR.* This segment contains an 852 nucleotide open reading frame bounded by transcriptional and translational signals. Both in vivo and in vitro the cloned segment produces a single protein that migrates in gels with the cellular protein (F33.4) implicated as the *htpR* product. Properties of a cloned fragment of the coding sequence truncated at the promoter-distal end are consistent with this assignment. The *htpR* gene product appears homologous to the sigma factor of RNA polymerase, and the two proteins are predicted to have similar secondary structure. In addition, two regions of the predicted *htpR* product resemble protein-DNA contact points conserved in known DNA-binding proteins.

Introduction

The response of Escherichia coli to a shift-up in growth temperature resembles in many respects the response of other microbes and of animal and plant cells (reviewed by Schlesinger et al., 1982). Almost immediately after an elevation in temperature there is an accelerated synthesis of more than a dozen polypeptides. The higher the temperature, the more pronounced is the preferential synthesis of these heat shock proteins, and after 1 hr of incubation at 50°C they become virtually the only translational products. How this response is brought about is not fully understood in any organism, but in E. coli it clearly depends on the protein product of a gene, *htpR* (Neidhardt and VanBogelen, 1981) or *hin* (Yamamori and Yura, 1982), which exerts, overall, a positive control on the unlinked operons of the heat shock regulon.

This bacterial regulatory gene has been identified and cloned (Neidhardt et al., 1983). In minicells and maxicells the cloned gene on plasmid pFN97 produces a protein, which by its migration on two-dimensional polyacrylamide gels appears to be identical in size (33 kd) and isoelectric point with a protein, F33.4, normally present in E. coli but deficient in an *htpR* mutant (Neidhardt et al., 1983). We present evidence for the sequence of the coding region of the *htpR* gene, its entire termination region, and a 193 nucleotide segment upstream of the translational initiation codon. In addition, we show that this DNA directs the synthesis of F33.4 in an in vitro transcription-translation system. The predicted sequence of this polypeptide resembles the predicted sequence for a portion of the E. coli sigma subunit of RNA polymerase, and two segments show similarity to protein-DNA contact points conserved in known DNA-binding proteins. These findings are consistent with the idea that the HtpR protein activates heat shock genes, possibly by interacting both with DNA and with RNA polymerase or sigma factor.

[§] Present address: Department of Biological Sciences, Stanford University, Stanford, California 94305.

Results

Choice of DNA and Sequencing Strategy

A partial restriction map of the E. coli chromosomal region inserted in two plasmids used for sequencing studies is shown in Figure 1. Plasmid pFN82 does not complement the *htpR* mutation in strain K165, but does yield wild-type recombinants in a *recA*$^+$ background; it produces no detectable insert-specific proteins in minicells or maxicells. The second plasmid, pFN97, readily complements the *htpR* mutation in a *recA* background, and it directs the synthesis of a protein indistinguishable from the normal cellular protein F33.4, the postulated product of the *htpR* gene. From these and other results, and from the size of F33.4 (approximately 33 kd), it had been concluded (Neidhardt et al., 1983) that the *htpR* gene is located wholely within the pFN97 insert at the end in common with the insert of pFN82, and that the *htpR* gene is transcribed in the direction from the Hind III site to the Pvu II sites (from left to right as shown in Figure 1). These plasmids were therefore used to prepare DNA restriction fragments for sequencing, using the strategy shown in Figure 1. A third plasmid, pFN92, contains all but the 15 amino acid codons of *htpR* located after the Pvu II site (Figures 1 and 4) and was used to verify the *htpR* reading frame.

Nucleotide Sequence of *htpR*

We determined the DNA sequence of an 1125 base pair region containing the *htpR* gene (Figure 2). This region contained a single open reading frame of sufficient length to encode the *htpR* gene product. The 852 base pairs of this reading frame encode a protein of 284 amino acid residues, beginning 126 base pairs after the Hind III site. The molecular weight (32,381) and charge (nearly equal numbers of acidic and basic residues) are consistent with the observed mobility of F33.4 in gels (Neidhardt et al., 1983). The amino acid composition predicted by the DNA sequence is summarized in Table 1.

DNA Sequences Flanking the *htpR* Gene

Twelve base pairs upstream of the presumed ATG start codon of this open reading frame is the sequence GGAGG. This sequence should provide a good ribosome binding site for the *htpR* gene by the criteria of Shine and

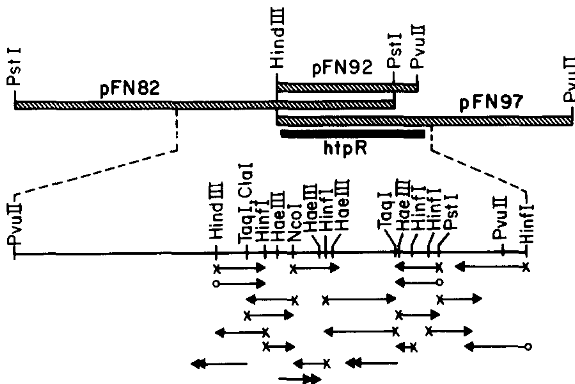

Figure 1. Sequencing Strategy for the E. coli K12 *htpR* Gene

The upper part of the figure shows the DNA cloned into plasmids used for sequencing, and indicates the location of the *htpR* gene. Plasmid pFN97 was derived from pOX15 as described in Neidhardt et al. (1983), and both plasmids were the source of DNA for sequencing this region.

The lower part of the figure shows the restriction sites used in sequencing. The length and direction of sequencing of each fragment are shown by the arrows. X indicates that the 5′ end was labeled; O indicates the 3′ end was labeled in the Maxam-Gilbert procedure. Double-headed arrows indicate that the sequencing was done by dideoxyribonucleotide triphosphate chain termination.

Dalgarno (1974). In addition, at least seven recognizable promoter-like sequences exist in the 200 base pairs preceding the *htpR* gene. The −35 regions for these possible promoters begin 187, 150, 136, 127, 122, 116, and 45 base pairs preceding the start codon and have spacings between the −35 and −10 regions of 19, 17, 21, 18, 18, 17, and 16 base pairs, respectively (Figure 2). One of these might be the sole transcription initiation signal in vivo; alternatively, one or more from a set of functional promoters might be utilized under different growth conditions. Determination of the in vivo transcription initiation site(s) under different growth conditions should clarify this issue.

The flanking sequence from −58 to −24 upstream of the *htpR* gene contains a region of dyad symmetry that, if transcribed, would produce a potential RNA hairpin with a predicted stability of −15.6 kcal/mole (Figure 2). In addition, nucleotides −69 to −62 constitute an octanucleotide "boxA" sequence (consensus: C/TGCTCTT(T)A). This sequence has been implicated in the activity of N as an antiterminator in lambda (Friedman and Gottesman, 1983) and may be the site of action of the *nusA* gene product. This sequence is found just upstream of the phage lambda *nutL*, *nutR*, *qut*, phage 21*nut*R, and phage P22*nut*L regions, the E. coli *trp* t terminator and *rrnB* leader region, both of which are NusA-dependent termination sites, as well as in the conserved region of the *rrnA*, *rrnD*, *rrnE*, and *rrnX* operons (Friedman and Gottesman, 1983). The "boxA" sequence usually precedes a potential RNA hairpin in an arrangement similar to that observed for *htpR* (Figure 2). Transcription initiation at six of the seven potential promoters would produce an *htpR* RNA leader region that would include both the "boxA" sequence and the RNA hairpin. Whether or not these structures are important for transcriptional regulation of *htpR* is not known.

A second "boxA" sequence appears four nucleotides after the TAA translational stop codon at the end of the gene. It is followed closely by a sequence resembling a traditional transcription termination signal—a potential GC-rich hairpin of predicted stability −16.7 kcal/mole, followed by a run of 6 AT base pairs. Termination of transcription by RNA polymerase at this site was observed during in vitro transcription of DNA restriction fragments containing this region (data not shown). Approximately 80 nucleotides from this termination site is the −35 region of the promoter for *livJ*, the gene for the leucine, isoleucine, valine binding protein (a periplasmic component of the E. coli high-affinity branched chain amino acid transport system; R. Landick and D. Oxender, unpublished data).

Evidence for the Assignment of the *htpR* Reading Frame

To verify the reading frame assignment shown in Figure 2 we examined the proteins produced in a standard Zubay transcription-translation system (Zubay et al., 1970) programmed with pFN92 or pFN97. Nucleotides 805–810 of the *htpR* gene constitute a Pvu II site that is the end of the inserted fragment of pFN92 (Figure 1 and 2). This Pvu II cut deletes the *livJ* fragment, the *htpR* termination region, and the 15 C-terminal amino acid codons of the *htpR* gene. In pFN92, this Pvu II cut is the site of splicing into the pBR322 Pvu II site. The nucleotide sequence of pFN92 in the region surrounding this fusion is shown in Figure 3. If the reading frame assignment shown in Figure 2 is correct, translation of the pFN92 sequence is terminated at a stop signal six codons into the pBR322 sequence, resulting in a fusion protein 9 amino acids shorter than the wild-type HtpR coded by pFN97. If either of the other two reading frames was correct, pFN92 would code for a fusion protein longer than the wild-type HtpR. As shown in Figure 4, in vitro transcription-translation of pFN92 produced a smaller protein product than did pFN97. Analysis by two-dimensional gel electrophoresis (not shown) demonstrated that the product of pFN97 coincided with the previously established location of the cellular protein F33.4 (Neidhardt et al., 1983). These results are consistent with the previous finding that pFN92 and pFN97 direct the synthesis in minicells and maxicells of two proteins that, as judged by two-dimensional gel electrophoresis, differ in size by approximately 1 kd (Neidhardt et al., 1983). The smaller protein coded by pFN92 is biologically inactive. These results provide strong evidence that the *htpR* reading frame shown in Figure 2 is correct.

Codon preference analysis (Gribskov et al., 1984) revealed that the assigned reading frame utilizes preferred codons, the others do not (analysis not shown).

On these bases we believe the protein sequence shown in Figure 2 is correct.

Comparison of Nucleotide Sequence of *htpR* and *rpoD*

Although the *htpR* gene product is only 45% the size of the sigma subunit of RNA polymerase (Burton et al., 1981),

```
                                                                                     -119
      CCCCGTTGTCTCGTTCCCGGTATTTCATCTCTATGTCACATTTTGTGCGTAATTTATTCACAAGCTTGCATTGA

                                                                                      -40
ACTTGTGGATAAAATCACGGTCTGATAAAACAGTGAATGATAACCTCGTTGCTCTTAAGCTCTGGCACAGTTGTTGCTA
                                                                  box A        hairpin
                                        1                                          30
CCACTGAAGCGCCAGAAGATATCGATTGGGAGGATTTGG ATG GCT GAC AAA ATG CAA AGT TTG GCT TTG
                                 S.D.    Met Ala Asp Lys Met Gln Ser Leu Ala Leu
                                                                                   90
GCC CCA GTT GGA GGC CTA GAT TCC TAC ATC CGG GCA GCT AAC GCG TGG CCG ATG TTG TCG
Ala Pro Val Gly Gly Leu Asp Ser Tyr Ile Arg Ala Ala Asn Ala Trp Pro Met Leu Ser
                                                                                  150
GCT GAC GAG GAG CGG GCG CTG GCT GAA AAG CTG CAT TAC CAT GGC GAT CTG GAA GCA GCT
Ala Asp Glu Glu Arg Ala Leu Ala Glu Lys Leu His Tyr His Gly Asp Leu Glu Ala Ala
                                                                                  210
AAA ACG CTG ATC CTG TCT CAC CTG CGG TTT GTT GTT CAT ATT GCT CGT AAT TAT GCG GGC
Lys Thr Leu Ile Leu Ser His Leu Arg Phe Val Val His Ile Ala Arg Asn Tyr Ala Gly
                                                                                  270
TAT GGC CTG CCA CAG GCG GAT TTG ATT CAG GAA GGT AAC ATC GGC CTG ATG AAA GCA GTG
Tyr Gly Leu Pro Gln Ala Asp Leu Ile Gln Glu Gly Asn Ile Gly Leu Met Lys Ala Val
                                                 Segment A                        330
CGC CGT TTC AAC CCG GAA GTG GGT GTG CGC CTG GTC TCC TTC GCC GTT CAC TGG ATC AAA
Arg Arg Phe Asn Pro Glu Val Gly Val Arg Leu Val Ser Phe Ala Val His Trp Ile Lys
                                                                                  390
GCA GAG ATC CAC GAA TAC GTT CTG CGT AAC TGG CGT ATC GTC AAA GTT GCG ACC ACC AAA
Ala Glu Ile His Glu Tyr Val Leu Arg Asn Trp Arg Ile Val Lys Val Ala The Thr Lys
                                                                                  450
GCG CAG CGC AAA CTG TTC TTC AAC CTG CGT AAA ACC AAG CAG CGT CTG GGC TGG TTT AAC
Ala Gln Arg Lys Leu Phe Phe Asn Leu Arg Lys Thr Lys Gln Arg Leu Gly Trp Phe Asn
                                                                                  510
CAG GAT GAA GTC GAA ATG GTG GCC CGT GAA CTG GGC GTA ACC AGC AAA GAC GTA CGT GAG
Gln Asp Glu Val Glu Met Val Ala Arg Glu Leu Gly Val Thr Ser Lys Asp Val Arg Glu
                                            Segment B                             570
ATG GAA TCA CGT ATG GCG GCA CAG GAC ATG ACC TTT GAC CTG TCT TCC GAC GAC GAT TCC
Met Glu Ser Arg Met Ala Ala Gln Asp Met Thr Phe Asp Leu Ser Ser Asp Asp Asp Ser
                                                                                  630
GAC AGC CAG CCG ATG GCT CCG GTG CTC TAT CTG CAG GAT AAA TCA TCT AAC TTT GCC GAC
Asp Ser Gln Pro Met Ala Pro Val Leu Tyr Leu Gln Asp Lys Ser Ser Asn Phe Ala Asp
                                                                                  690
GGC ATT GAA GAT GAT AAC TGG GAA GAG CAG GCG GCA AAC CGT CTG ACC GAC GCG ATG CAG
Gly Ile Glu Asp Asp Asn Trp Glu Glu Gln Ala Ala Asn Arg Leu Thr Asp Ala Met Gln
                                                                                  750
GGT CTG GAC GAA CGC AGC CAG GAC ATC ATC CGT GCG CGC TGG CTG GAC GAA GAC AAC AAG
Gly Leu Asp Glu Arg Ser Gln Asp Ile Ile Arg Ala Arg Trp Leu Asp Glu Asp Asn Lys
                                                                                  810
TCC ACG TTG CAG GAA CTG GCT GAC CGT TAC GGC GTT TCC GCT GAG CGT GTA CGC CAG CTG
Ser Thr Leu Gln Glu Leu Ala Asp Arg Tyr Gly Val Ser Ala Glu Arg Val Arg Gln Leu
                                    Segment C                                     874
GAA AAG AAC GCG ATG AAA AAA TTG CGT GCT GCC ATT GAA GCG TAA TTTCCGCTATTAAGCAGAG
Glu Lys Asn Ala Met Lys Lys Leu Arg Ala Ala Ile Glu Ala END       box A
                                                                                  953
AACCCTAGATGAGAGTCCGGGGTTTTTGTTTTTGGGCCTCTGTAATAATCAATTTCCCCTCCGGCAAAACGCCAATCC
 termination                                                                     1032
CCACGCAGATTGTTAATAAACTGTCAAAATAGCTATTCCAATATCATAAAAATCGGGATATGTTTTAGCAGAGTATGCT
                                   "-35"                      "-10"      livJ
```

Figure 2. Nucleotide Sequence of the E. coli K12 *htpR* Region

Nucleotides are numbered from the start of the translation initiation codon. Underlined segments are as follows: boxA, putative Nus recognition sites; hairpin, region of dyad symmetry that would produce a potential RNA hairpin if transcribed; S.D., Shine-Dalgarno sequence complementary to the 3′ end of 16S rRNA; Segment A, 14 amino acid segment identical with amino acids 403–416 in sigma factor (see Figure 6); Segments B and C, regions strongly resembling the DNA-protein contact points in known DNA-binding proteins, with highly conserved amino acids boxed; termination, a region of dyad symmetry followed by a T-rich region; "−35" and "−10," the RNA polymerase sites located upstream of the initiation of translation of *livJ*.

some evidence suggests that sigma factor antagonizes the function of the *htpR* protein (Yamamori et al., 1982). It therefore seemed of interest to examine the two genes for regions of homology. Seven regions were discovered that consist of 15 or more bases with >85% identity (Figure 5). All seven are in the coding region of *rpoD*, but one is in the leader region of *htpR*. The regions do not appear sequentially in the two genes. Nevertheless, this degree of similarity strongly suggests an evolutionary relationship between *htpR* and *rpoD*.

Comparison of Amino Acid Sequence and Composition of Sigma Factor and HtpR

Analysis of amino acid sequences revealed that the predicted sequence for the *htpR* gene product is similar to one portion of the predicted sequence of sigma factor

(Figure 6). This similarity includes a 14 amino acid segment of identity, and overall there is 24% complete matching and 43% total matching (identities plus conservative replacements) between HtpR and the amino acid residues 326–613 of sigma factor.

Similarities in composition are numerous. Both contain higher than average percentages of charged amino acid residues, especially acidic amino acid residues, and both contain nearly 2.5 times more methionine than the average protein (Burton et al., 1981). Both sequences contain less cysteine (HtpR has none), glycine, and aromatic amino acid residues than the average protein, and both sequences contain clusters of positive or negative charge. Particularly interesting are the clusters of acidic residues found in the HtpR sequence at positions 32–34, 183–191, 213–215, and 246–248 (Figure 2) and throughout the N-terminal portion of sigma. Like sigma factor, HtpR contains major regions of high positive charge (residues 109–145, net +10) and high negative charge (residues 179–220, net −12).

The secondary structure was predicted from the amino acid sequence of the *htpR* gene product by the method of Chou and Fasman (1978). This analysis suggests a protein of reasonably high α-helical (55%–60%) and low β-sheet (8%–15%) conformation. For sigma factor the α-helical content is 50%–55% and the β-sheet 10%–15% (Burton et al., 1981). Both proteins contain long, continuous sections of α-helix. Of interest is the fact that the two proteins have similar predicted secondary structure for 20 amino acid residues on each side of the 14 amino acid segment of identity—a coil region of 10–12 amino acids bounded by β-sheets.

Table 1. Amino Acid Composition of HtpR Predicted from DNA Sequence

Amino Acid	Residues	Mole %
Alanine	34	12.0
Arginine	23	8.1
Asparagine	12	4.2
Aspartate	23	8.1
Cysteine	0	0
Glutamate	22	7.7
Glutamine	14	4.9
Glycine	13	4.6
Histidine	6	2.1
Isoleucine	12	4.2
Leucine	29	10.2
Lysine	16	5.6
Methionine	11	3.9
Phenylalanine	8	2.8
Proline	6	2.1
Serine	16	5.6
Threonine	8	2.8
Tryptophan	6	2.1
Tyrosine	7	2.5
Valine	18	6.3
Total	284	100

Similarity to DNA-Binding Proteins

We also examined the *htpR* gene product sequence for regions similar to the consensus sequence for the two α-helices containing the main DNA–protein contact points in known DNA-binding proteins (Sauer et al., 1982; Matthews et al., 1982). The *htpR* protein sequence contains two regions that strongly resemble this consensus sequence (segments B and C in Figure 2). Both contain the pattern Ala-N-N-N-Gly-N-N-N-N-N-Val(Ile) that is virtually invariant among DNA-binding proteins. The first region (amino acids 154–173) contains residues at 13 positions that are found at the same position in other DNA-binding proteins, and the second region (amino acids 253–272) contains matching residues at 14 positions.

Chou–Fasman analysis (see above) predicted that both sequences (B and C in Figure 2) form two α-helices connected by a turn at the conserved glycine—a structural motif observed in the DNA contact points of known repressor structures (Matthews et al., 1982).

Discussion

The protein product of the *htpR* gene may be essential at all temperatures (Carol Gross and Alan Grossman, personal communication), and is certainly required for E. coli cells to exhibit a normal heat shock response (Neidhardt and VanBogelen, 1981; Yamamori and Yura, 1982). This conclusion was originally based on the properties of a mutant with a nonsense mutation in *htpR* in a strain carrying a temperature-conditional suppressor, and has

A. Normal sequence (pFN97)

```
- CGC CAG*CTG GAA AAG AAC GCG ATG AAA AAA TTG CGT GCT GCC ATT GAA GCG TAA -
- arg gln leu glu lys asn ala met lys lys leu arg ala ala ile glu ala end
```

B. Sequence deduced for pFN92

```
- CGC CAG*ctg cct cgc gcg ttt cgg tga -
- arg gln leu pro arg ala phe arg end
```

Figure 3. Nucleotide Sequence of the C-Terminal End of the *htpR* Gene in Plasmid pFN92

(A) The sequence of the C-terminal end of the *htpR* gene and its normal product is shown with an asterisk to mark the cut at the Pvu II site (CAGCTG).
(B) The sequence resulting from splicing the main portion of the *htpR* gene to the pBR322 vector at the Pvu II site to form pFN92. The lowercase base letters indicate pBR322 sequence (Sutcliffe, 1978).

been supported by subsequent studies. For example, the *dnaK* gene product, itself one of E. coli's heat shock proteins, appears to antagonize the action of *htpR* (Tilly et al., 1983). Mutants defective in *dnaK* fail to exhibit the dampening of heat shock induction that normally occurs after 10 min of high level synthesis. From this observation it was possible to imagine that the role of the *dnaK* protein was to repress synthesis of heat shock proteins, and that HtpR might function merely to remove or inactivate this repressor. The same authors showed, however, that *htpR* function is still required for a heat shock response even in a *dnaK* defective mutant—a double mutant, *dnaK* and *htpR,* cannot produce a heat shock response.

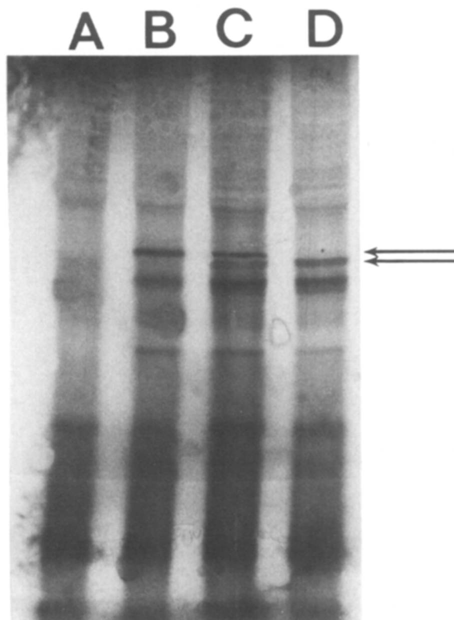

Figure 4. In Vitro Synthesis of Protein Products from Plasmids pFN92 and pFN97

The upper arrow indicates the protein produced from the *htpR* gene on pFN97, as verified by two-dimensional gel electrophoresis. The lower arrow indicates the slightly smaller product produced from the truncated *htpR* gene of pFN92 (see Figure 4). Lane A, no added DNA; Lane B, reaction with pFN97 DNA; lane D, reaction with pFN92 DNA; lane C, mixture of reaction samples B and D.

Although the overall effect of the *htpR* protein is positive, it is still possible that it exerts its effect by countering some negative factor (other than the *dnaK* gene product) or factors. It could, for example, act similarly to the *recA* protein, which is thought to cleave the general repressor of SOS operons, LexA (Little and Mount, 1982). Support for this idea has come recently from the work of Gross and her collaborators (T. A. Baker, A. D. Grossman, and C. A. Gross, submitted), who have demonstrated that *htpR* mutant cells are defective in a proteolytic activity. There have been no reports, however, pointing to the existence of a general repressor of heat shock operons, and therefore one must consider the possibility that if HtpR does act by releasing repression it might have to inactivate as many as 15 different repressors.

On the other hand, there is a distinct possibility that HtpR directly activates heat shock genes upon a shift-up in temperature. It could do this in a number of ways consistent with current information. It might bind to the promoter regions of heat shock genes, much as the cAMP–CAP protein binds to promoters of many catabolic operons (reviewed in Botsford, 1981). This possibility is consistent with our observation that HtpR has regions similar to key protein–DNA contact points of known DNA-binding proteins (Figure 2). In this respect HtpR resembles proteins such as CAP and lambda CI, two proteins responsible for transcriptional activation (Steitz et al., 1982; Gussin et al., 1983).

Alternatively, HtpR might activate heat shock genes by modifying promoter recognition by RNA polymerase. Yamamori et al. (1982) have reported that synthesis of heat shock proteins is markedly enhanced when cellular amounts of sigma are reduced severalfold, and that this effect is brought about by increased transcription from heat shock genes (as is the normal heat shock response in E. coli). Conversely, a mutant that overproduces sigma produced significantly reduced amounts of heat shock

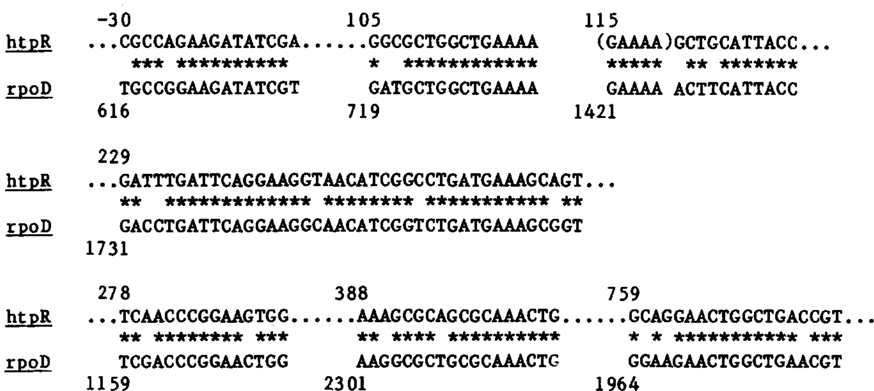

Figure 5. Comparison of the DNA Sequence of *htpR* and *rpoD*

The seven segments of the *htpR* gene consisting of 15 or more bases having 85% identity with regions of the *rpoD* gene (Burton et al., 1981) are shown in sequential order as they appear in the *htpR* gene. Identical bases are indicated by the asterisks. Bases in *htpR* are numbered as in Figure 2; bases in *rpoD* are numbered as in Burton et al. (1981).

```
htpR    1 MADKMQSLALAPVGGLDSYIRAANAWPMLSADEERALAEKLHYHGDLE--AAKTLILSHL
            *=*  *       =            =* *  * *  **         = = **   =
rpoD  326 WSEKLHDVSEEVHRALQKLQQIEEE-TGLTIEQVKDINRRMSIGEAKARRAKKEMVEANL

htpR   59 RFVVHIARNYAGYGLPQADLIQEGNIGLMKAVRRFNPEVGVRLVSFAVHWIKAEIHEYVL
          = =* ==* =   ==   ============== *=    = *  **=  ==*  =   *
rpoD  385 RLVISIAKKYTNRGLQFLDLIQEGNIGLMKAVDKFEYRRGYKFSTYATWWIRQAITRSIA

htpR  119 RNWRIVKVATTKAQRKLFFNLRKTKQRLGWFNQDEV-EMVARELGVTSKDVREMESRMAA
             = ***           = = **= =     *   = *=  * *    *= *  **=
rpoD  445 DQARTIRIPVHMIETINKLN-RISRQMLQEMGREPTPEELAERMLMPEDKIRKV-LKIAK

htpR  178 QDMTFDLSSDDDSDSQPMAPVLYLQDKSSNFADGIEDDNWEEQAANRLTDAMQGLDERSQ
            ** *    == ==       ** = *      **    =   =    = * ==  =
rpoD  503 EPISMETPIGDDEDSHLGD---FIEDTTLELP--LDSATTESLRAATH-DVLAGLTAREA

htpR  238 DIIRARW-LDEDNKSTLQELADRYGVSAERVRQLEKNAMKKLRAAIEA(284)
           **= =* *=     == =*   * =* ==*==*=  =**===
rpoD  557 KVLRMRFGIDMNTDYTLEEVGKQFDVTRERIRQIEAKALRKLRHPSRSEVLRSFLDD(613)
```

Figure 6. Similarity between the Amino Acid Sequence Predicted from the DNA Sequences of the *htpR* and *rpo* Genes

Comparison of the amino acid sequences of the *htpR* gene product (Figure 2) and the sigma subunit of *E. coli* RNA polymerase (Burton et al., 1981). Identical amino acid residues are marked (=); conservative replacements are marked (*). Conservative replacements are defined as being within the groups: (D,E), (K,R), (S,T), (F,Y,W), and (I,L,V,M). Gaps have been inserted to improve the alignment, and are marked by hyphens within the sequences. The one-letter amino acid codes are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

proteins. These findings raise the interesting possibility that the heat shock regulon might be controlled by the reprogramming of RNA polymerase. Alternative sigma-like transcription factors are known in other bacterial species such as Bacillus subtilis in which they have been shown to control transcription of specific sets of genes during sporulation and bacteriophage synthesis, and replacement of one species of sigma factor by another may be facilitated by a sigma-dissociating factor (reviewed in Losick and Pero, 1981). The minor and developmentally specific sigma factors are smaller than the major form (e.g., 28, 29, and 37 kd compared with 55 kd). A gene for one of them, encoded in the phage SP01, has recently been cloned and sequenced. The predicted structure of its product resembles the E. coli sigma factor in having a large content of α-helix (Costanzo and Pero, 1983).

Recently direct genetic evidence has been obtained of an interaction between sigma factor and HtpR. First, mutations that suppress the *rpoD800* allele (which produces a temperature-sensitive sigma factor) have been found to map in *htpR* (Alan Grossman, Yan-Ning Zhu, Carol Gross, Gail Christie, Joseph Helig, and Richard Calendar, unpublished data). In addition, the *rpoD800* allele suppresses the *htpR165* mutation; i.e., the double mutant has a near normal heat shock response (Alan Grossman, Yan-Ning Zhu, Tania Baker, and Carol Gross, unpublished data).

In this context the *htpR* sequence reported here is of considerable interest. Clearly, the polypeptide predicted from the *htpR* DNA sequence appears homologous to that predicted for sigma from the *rpoD* DNA sequence (Figure 6). Taken with the other similarities between the two proteins (see Results), this apparent homology suggests that HtpR and sigma factor may share some function in transcription initiation. The region of greatest similarity (segment A in Figure 2) might represent a common domain with which HtpR and sigma factor contact the core RNA polymerase.

Could all of the properties ascribed to the *htpR* gene product be involved in its role in the heat shock response? It is possible that HtpR acts both as a DNA-binding protein and as a reprogrammer of RNA polymerase. One possibility, which also incorporates the reported defect in proteolysis in *htpR* mutants, is that HtpR binds to promoter regions of heat shock genes and, upon a shift-up in temperature, becomes active as a protease that modifies sigma in some way that brings about transcriptional activation.

The aspects of HtpR structure described here will be useful in direct tests of HtpR function through the use of additional mutants and cell free systems.

Experimental Procedures

Construction of Plasmids and M13 Clones

The construction of pFN82, pFN92, pFN97, and pOX15 has been described in Neidhardt et al. (1983). Fragments from pOX15 were used for sequencing by the method of Maxam and Gilbert (1980). The 5′ upstream region was cloned by ligating the Pvu II–Cla I fragment from pFN82 into the Sma I–Acc I sites of the single-stranded phage vector M13mp8 of Messing et al. (1981). The Hae III fragments were cloned by electroeluting the Hind III–Pst I fragment from an agarose gel as described in Maniatis et al. (1982) and cutting this fragment with Hae III before ligating into the Sma I site of the M13mp8. T4 DNA ligase and Klenow polymerase were from New England Biolabs. Restriction enzymes were from New England Biolabs, Bethesda Laboratories, and International Biotechnologies, Inc.

DNA Sequencing

For sequencing by chemical cleavage, DNA fragments were either 5′-end-labeled by T4 polynucleotide kinase (New England Nuclear) catalyzed reaction with γ-^{32}P-ATP (Amersham) or 3′-end-labeled by Micrococcus

luteus DNA polymerase (P-L Biochemicals) catalyzed reaction with the appropriate α-^{32}P-deoxyribonucleoside triphosphate (Amersham). The labeled fragments were sequenced by the method of Maxam and Gilbert (1980), except that A>G-specific cleavage was generated by modification of the DNA fragment in 18 mM EDTA, pH 4.0, for 5 min at 80°C in a total volume of 22 μl for subsequent piperidine cleavage as described by Maxam and Gilbert (1980).

Sequencing by dideoxynucleoside triphosphate chain termination was done as described by Sanger et al. (1980). The codon preference plot of Gribskov et al. (1984) was found useful in confirming the accuracy of the DNA sequence.

Sequence Analysis and Comparison

Analysis of the DNA sequence of the *htpR* region and its comparison with that of *rpoD* were carried out using programs described by Fristensky et al. (1982).

The amino acid sequences were aligned using the program GAP (Devereux et al., 1984) combined with a data matrix derived from Dayhoff (1978). GAP uses the algorithm of Needleman and Wunsch (1970) modified to allow the imposition of a gap length penalty. We used the method of Fitch and Smith (1983) to determine the best alignment. The sequence comparison shown in Figure 6 is changed from that derived by the computer by the deletion of a gap inserted between amino acids 518 and 519 of sigma and by increasing the length of the gap between residues 521 and 522 to 3.

In Vitro Transcription and Translation

In vitro synthesis of protein was done using the cell-free system described in Zubay et al. (1970). E. coli strain K165 (the *htpR* mutant described in Neidhardt and VanBogelen, 1981) was grown at 30°C and used for preparation of the S-30 fraction. The reaction was carried out at 37°C for 40-60 min, and samples were run on SDS-polyacrylamide gels (11.5%) as described by O'Farrell (1975). Synthesis of the protein products of *htpR* and *bla* was directed from pFN97.

Acknowledgments

This work was supported by National Institutes of Health grant GM 17892 (F. C. N.). The authors would like to thank Wesley Dunnick and Michael Mowatt for assistance with sequence analysis, and Alan Schauer for assistance with computer analysis of sequences and for many helpful discussions. Costa Georgopoulos, Carol Gross, Roger Hendrix, and Alvin Markowitz were generous in providing advance copies of manuscripts and related information.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 18, 1984; revised May 23, 1984

References

Botsford, J. L. (1981). Cyclic nucleotides in procaryotes. Microbiol. Rev. *45*, 620-642.

Burton, Z., Burgess, R. R., Lin, J., Moore, D., Holder, S., and Gross, C. A. (1981). The nucleotide sequence of the cloned *rpoD* gene for the RNA polymerase sigma subunit from *E. coli* K-12. Nucl. Acids Res. *9*, 2889-2903.

Chou, P. Y., and Fasman, G. D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. *47*, 45-148.

Costanzo, M., and Pero, J. (1983). Structure of a *Bacillus subtilis* bacteriophage SP01 gene encoding a RNA polymerase sigma factor. Proc. Nat. Acad. Sci. USA *80*, 1236-1240.

Dayhoff, M. O. (1978). Atlas of Protein Sequence and Structure, *5*, Suppl. 3. (Silver Springs, Maryland: National Biomedical Research Foundation).

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence programs for the VAX. Nucl. Acids Res. *12*, 387-395.

Fitch, W. M., and Smith, T. F. (1983). Optimal sequence alignments. Proc. Nat. Acad. Sci. USA *80*, 1382-1386.

Friedman, D. I., and Gottesman, M. (1983). Lytic mode of lambda development. In Lambda II, Hendrix, Roberts, Stahl, and Weisberg, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 21-51.

Fristensky, B., Lis, J., and Wu, R. (1982). Portable microcomputer software for nucleotide sequence analysis. Nucl. Acids Res. *10*, 6451-6463.

Gribskov, M., Devereux, J., and Burgess, R. (1984). The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression. Nucl. Acids Res. *12*, 539-549.

Gussin, G. N., Johnson, A. D., Pabo, C. O., and Sauer, R. T. (1983). Repressor and cro protein: structure, function, and role in lysogenization. In Lambda II, Hendrix, Roberts, Stahl, and Weisberg, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 93-121.

Little, J. W., and Mount, D. W. (1982). The SOS regulatory system of Escherichia coli. Cell *29*, 11-22.

Losick, R., and Pero, J. (1981). Cascades of sigma factors. Cell *25*, 582-584.

Maniatis, T., Fritsch, E. F., and Sambrook, J., eds. (1982). Molecular Cloning, A Laboratory Manual. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Matthews, B. W., Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., and Takeda, Y. (1982). Cro repressor protein and its interaction with DNA. Cold Spring Harbor Symp. Quant. Biol. *47*, 427-433.

Maxam, A. M., and Gilbert, W. (1980). Sequencing end-labeled DNA with base specific chemical cleavages. Meth. Enzymol. *65*, 449-560.

Messing, J., Crea, R., and Seeberg, R. (1981). A system of shotgun DNA sequencing. Nucl. Acids Res. *9*, 309-321.

Needleman, S. B., and Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. *48*, 443-453.

Neidhardt, F. C., and VanBogelen, R. A. (1981). Positive regulatory gene for temperature-controlled proteins in *Escherichia coli*. Biochem. Biophys. Res. Commun. *100*, 894-900.

Neidhardt, F. C., VanBogelen, R. A., and Lau, E. T. (1983). Molecular cloning and expression of a gene that controls the high-temperature regulon of *Escherichia coli*. J. Bacteriol. *153*, 597-603.

O'Farrell, P. H. (1975). High-resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. *250*, 4007-4021.

Pedersen, S., Reeh, S. V., Parker, J., Watson, R. J., Friesen, J. D., and Fiil, N. P. (1976). Analysis of the proteins synthesized in ultraviolet light-irradiated *Escherichia coli* following infection with the bacteriophages λ*drif 18* and λ*dfus-3*. Mol. Gen. Genet. *144*, 339-344.

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., and Roe, B. (1980). Cloning in single stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. *143*, 161-178.

Sauer, R. T., Yocum, R. R., Doolittle, R. F., Lewis, M., and Pabo, C. O. (1982). Homology among DNA-binding proteins suggests use of a conserved super-secondary structure. Nature *298*, 447-451.

Schlesinger, M. J., Ashburner, M., and Tissieres, A., eds. (1982). Heat Shock from Bacteria to Man. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Shine, J., and Dalgarno, L. (1974). The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Nat. Acad. Sci. USA *71*, 1342-1346.

Steitz, T. A., Weber, I. T., and Matthew, J. B. (1982). Catabolite gene activator protein: structure, homology with other proteins, and cyclic AMP and DNA binding. Cold Spring Harbor Symp. Quant. Biol. *47*, 419-426.

Sutcliffe, J. G. (1978). Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. Cold Spring Harbor Symp. Quant. Biol. *43*, 77-90.

Tilly, K., McKittrick, N., Zylicz, M., and Georgopoulos, C. (1983). The *dnaK* protein modulates the heat-shock response of Escherichia coli. Cell *34*, 641-646.

Yamamori, T., Osawa, T., Tobe, T., Ito, K., and Yura, T. (1982). *Escherichia coli* gene (*hin*) controls transcription of heat-shock operons and cell growth

at high temperature. In Heat Shock from Bacteria to Man, Schlesinger, Ashburner, and Tissieres, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 131-137.

Yamamori, T., and Yura, T. (1982). Genetic control of heat-shock protein synthesis and its bearing on growth and thermal resistance in *Escherichia coli* K12. Proc. Nat. Acad. Sci. USA *79*, 860-864.

Zubay, G., Chambers, D. A., and Cheong, L. C. (1970). Cell-free studies on the regulation of the *lac* operon. In The Lactose Operon, Beckwith and Zipser, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 375-391.

Note Added in Proof

The nucleotide sequence of bases 42 through 48 in Figure 2 should read CAACCTG. This correction changes a single amino acid and does not alter any of our conclusions.