JOURNAL OF BACTERIOLOGY, June 1988, p. 2668–2675
0021-9193/88/062668-08$02.00/0

Vol. 170, No. 6

Sequence of the *dnaB* Gene of *Salmonella typhimurium*

ANNETTE WONG, LESLIE KEAN, AND RUSSELL MAURER*

Department of Molecular Biology and Microbiology, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106

Received 28 December 1987/Accepted 16 March 1988

The *dnaB* gene of *Escherichia coli* encodes a helicase that operates at replication forks of the bacterium and certain of its bacteriophages to produce separated strands suitable for subsequent use by primase and DNA polymerase III. Here, we present the sequence of the *dnaB* gene of *Salmonella typhimurium*, a functionally interchangeable analog of the *E. coli dnaB* gene. The DnaB proteins of these two organisms, inferred from the DNA sequences, are identical in length and in 93% of amino acid residues. Extended portions of the DnaB proteins are also similar to two phage-encoded DNA replication proteins: the gene 4 helicase-primase of coliphage T7 and, as reported previously (H. Backhaus and J. B. Petri, Gene 32:289–303, 1984), the gene 12 protein of *Salmonella* phage P22. In contrast, little similarity was found between DnaB and either the UvrD repair helicase or transcription termination factor Rho (an RNA-DNA helicase). These results identify *S. typhimurium* DnaB as a member of the DnaB family of proteins by structural, as well as functional, criteria and provide the basis for the eventual identification, by mutational studies, of residues in DnaB critical for its function.

A common feature of the DNA replication cycle of *Escherichia coli* and many of its bacteriophages is the formation, at an origin, of an apparatus capable of migrating into unreplicated DNA, unwinding this DNA if it is not already single stranded, and catalyzing or stimulating formation of RNA primers for complementary-strand synthesis (8, 30, 40, 41). The exact composition of such an apparatus varies for different DNA molecules, but a crucial component in many cases is the DnaB protein encoded in the *E. coli* chromosome. This protein, a hexamer of identical subunits (5, 51), utilizes the energy derived from ATP hydrolysis to drive strand separation, as well as its own translocation, in a 5′-to-3′ direction along the single strand to which it is bound (33). *E. coli* and its DnaB-dependent phages, λ and φX174, use elaborate controls to ensure that the few tens of molecules of DnaB available in a cell (58) are used efficiently and specifically for DNA replication. Elements of these controls include proteins which bind to the origin, proteins which transfer DnaB to the marked origin, and proteins which activate DnaB bound at the replication origin of phage λ (3, 4, 8, 14, 15, 22, 28, 40, 63). Some other phages, such as T4, T7, and *Salmonella* phage P22, do not depend on host *dnaB*. These phages use known or suspected phage-encoded helicases (48, 49, 54, 64). Once positioned at a replication fork, DnaB or an analogous protein may provide attachment sites for other required proteins, such as primase or DNA polymerase III (1, 7, 33, 65).

A full understanding of the operation of a DNA replication fork clearly requires knowledge of the structure of DnaB protein and the geometry of DnaB interactions with other proteins and with DNA. Part of the necessary knowledge may be obtained from comparisons of DnaB with its analogs. We have previously reported the existence in *Salmonella typhimurium* of a *dnaB* analog that can complement temperature-sensitive *dnaB* mutants of *E. coli* (39). The complementation tests were carried out by using phage λ recombinants that bore the *S. typhimurium dnaB* analog. One of these clones has been used as the starting point for physical localization of this gene and its complete DNA sequence determination.

* Corresponding author.

MATERIALS AND METHODS

General methods. *E. coli* RM84 [*dnaB22*(Ts) *sup*^{0}; derived from strain FA77 (42) by lysogenization with λ112 (37)] was used to test *dnaB* complementation by plasmid clones. Plasmids and phage M13 derivatives were propagated in strain JM103 (43). DNA preparations for transformation and sequencing have been described previously (13, 35). Enzymes were used under conditions suggested by their commercial suppliers. Sequencing on M13 clones was carried out by the dideoxy method (53). The entire sequence presented in Fig. 2 was determined on both strands. Sequence similarity between DnaB and other proteins was evaluated by using the PRTALN program (67).

Plasmid and M13 constructions. Plasmid pFF17 contains the 2.0-kilobase (kb) *Hin*dIII-*Eco*RI fragment from λRM113 (*dnaB*$^+$; 38, 39) in a pUC8 vector (59). The analogous plasmid containing *dnaB34*(Am) is pFF45 and was derived from λRM220 (38). To make deletions from the *Eco*RI side of pFF17, plasmid DNA was linearized with *Eco*RI, phenol extracted, ethanol precipitated, and suspended in Bal 31 buffer (Bethesda Research Laboratories, Inc.). This DNA (50 μg) was treated with 3 U of Bal 31 (Bethesda Research Laboratories) at 30°C. At intervals, samples were removed and digestion was halted by 20-fold dilution into ice-cold TE (10 mM Tris [pH 7.5], 1 mM EDTA) containing 10 mM EGTA [ethylene glycol-bis(β-aminoethyl ether)-*N*,*N*,*N*′,*N*′-tetraacetic acid]. Samples exhibiting suitable extents of digestion were identified by an analytical gel run. These samples were pooled, phenol extracted, ethanol precipitated, and suspended in T4 DNA polymerase buffer (Bethesda Research Laboratories). The digested ends of the fragments were made blunt in a two-step reaction with T4 DNA polymerase (1.7 U; Bethesda Research Laboratories). The first step was for 3 min at 37°C in the absence of nucleotides; this step removed any 3′ extensions, leaving 5′ extensions. Upon subsequent addition of nucleotides, the recessed 3′ ends were filled out. *Eco*RI linkers (GGAATT

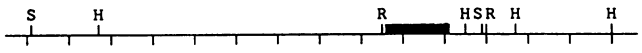

FIG. 1. Restriction map of the *S. typhimurium* genomic DNA insert in recombinant phage λRM113. The λ arms are not shown. The region identified as *dnaB* is shown by the heavy bar. S, *Sma*I; H, *Hin*dIII; R, *Eco*RI. Hatch marks below the line indicate 1-kb intervals.

CC; Bethesda Research Laboratories), phosphorylated with T4 DNA kinase, were added to the blunt-ended fragments by T4 DNA ligase. Following another phenol extraction and ethanol precipitation, the DNA was digested to completion with an excess of *Eco*RI and *Hin*dIII. The products of digestion were run on a 1.5% agarose gel in Tris-acetate buffer (35). Gel slices were obtained containing DNA in length increments of approximately 123 base pairs (bp) (as judged by reference to a 123-bp ladder [Bethesda Research Laboratories] in an adjacent lane), and the DNA was extracted from the slices by adsorption onto glass powder (60). The *Hin*dIII-*Eco*RI fragments thus obtained were cloned into pUC8 for complementation analysis and determination of the size of the insert. Selected fragments were recloned from pUC8 into M13mp19 (47) for sequencing. Deletions from the *Hin*dIII side were made in analogous fashion, except that *Hin*dIII linkers (GCAAGCTTGC; Bethesda Research Laboratories) were used and the recloning was into M13mp18 (47).

Fine-structure mapping of *dnaB34*(Am). The site of mutation in *dnaB34*(Am) was localized by heteroduplex mapping, a form of deletion mapping (56). A subset of the Bal 31-generated deletions was used. To use deletions entering from the *Eco*RI side, the pUC8 derivative containing the deleted *dnaB* fragment was linearized at the site of the deletion by digestion with *Eco*RI and heteroduplexed (56) with pFF45 linearized with *Hin*dIII. For deletions entering from the *Hin*dIII side, the procedure was similar, except that the reciprocal combination of enzymes was used. In some experiments, the deletion plasmid DNA was additionally digested with *Nde*I, an enzyme that cuts once in the vector portion of the plasmid; such treatment did not affect the basic design of the experiment but was helpful in reducing the background level of transformants in homoduplex (control) reactions. As described by Shortle (56), circular heteroduplex molecules have a gap in one strand corresponding to the material deleted in one of the plasmids. If the *dnaB34* mutation lies opposite the gap, then gap-filling upon transformation into *E. coli* produces exclusively mutant plasmids; but if the *dnaB34* mutation lies outside the gapped region, then mismatch repair following transformation produces a substantial fraction of DnaB$^+$ transformants. By identifying the largest deletion from each side that was able to produce DnaB$^+$ (i.e., temperature-resistant) transformants of strain RM84, *dnaB34* was mapped to the interval between nucleotides 427 through 561 (as numbered in Fig. 2). This region was sequenced by using suitable synthetic primers complementary to *dnaB*.

RESULTS

Physical localization of *dnaB*. The restriction map of λRM113 (*dnaB*$^+$) is shown in Fig. 1. The approximate location of *dnaB* within this clone was determined by a subcloning approach. Selected *Hin*dIII, *Eco*RI, and *Hin*dIII-*Eco*RI fragments from this phage were subcloned into the vector pUC8, and the subclones were tested for ability to complement the *E. coli dnaB* temperature-sensitive mutant strain, RM84. Of the fragments tested, only the 2.5-kb *Eco*RI fragment and the 2.0-kb *Hin*dIII-*Eco*RI fragment internal to it were able to complement. These results indicated that the entire *dnaB* gene was located on the 2.0-kb fragment.

Two sets of nested deletions, one set entering from each end of the 2.0-kb fragment, were used to further localize *dnaB* in complementation tests and to sequence the complete gene on both strands by overlapping the sequence determined across each deletion endpoint. The complementation analysis of the deletions showed that one end of *dnaB* is located between 11 and 157 bp from the center of the *Eco*RI recognition sequence in the original fragment. The other end of *dnaB* was positioned about 600 bp from the *Hin*dIII end of the original fragment by similar analysis.

***dnaB* sequence.** The nucleotide and derived amino acid sequence of the region identified as *dnaB* by complementation analysis, as well as some flanking sequences, are shown in Fig. 2. The sequence contains a single complete reading frame, 1,413 base pairs in length, encoding a protein of 471 amino acids, which we take to be DnaB. For ease of comparison with *E. coli dnaB*, we have adopted the *E. coli* numbering system for the *S. typhimurium dnaB* gene and protein. Since the N-terminal methionine of DnaB is removed in *E. coli*, the numbering system begins with the next codon (GCA, alanine; 44). Therefore, only 470 amino acids are shown in Fig. 2, even though the fate of the N-terminal methionine in *S. typhimurium* is unknown. That this reading frame encodes DnaB protein is confirmed by the sequence of a hydroxylamine-induced amber mutation which is caused by a C · G-to-T · A change in codon 149 (Gln-149 → Amber). Furthermore, the extent of this reading frame is consistent with the boundaries of *dnaB* established by the deletion analysis, and the derived amino acid sequence is closely similar to the previously determined amino acid sequence of the *E. coli* DnaB protein (itself derived by conceptual translation of the *E. coli dnaB* gene and confirmed by total amino acid composition and limited direct amino acid sequencing of the protein; 44). The differences between the *E. coli* and *S. typhimurium* DNA and protein sequences are also shown in Fig. 2.

Sequence upstream of *dnaB*. The transcriptional control signals for *dnaB* are unknown. Our deletion analysis suggests that no more than 99 bp of upstream sequence are needed for *dnaB* expression; the next-longer deletion from the upstream side enters the protein-coding sequence. In all the *dnaB* plasmids, *dnaB* is oriented oppositely to the vector *lacZ* gene, thus eliminating the *lac* promoter as a possible external source of *dnaB* transcription. In the *E. coli* sequence, Nakayama et al. (44) proposed the sequence TTGCCA . . . .(16 bases) . . . .TATCGT (starting 89 bp 5′ to the initiator ATG) as the likely *dnaB* promoter, since it bears the closest similarity to consensus bacterial promoters (25, 27). In the *S. typhimurium* sequence, the proposed −35 region is preserved at positions −74 to −69, but the −10 sequence is completely lost as part of an 18-bp deletion in *S. typhimurium* DNA. The sequence now found in the −10 position in *S. typhimurium* shows poor homology with the canonical sequence (AATGTG versus the canonical TATAAT). This observation raises some doubt about the identity of the *dnaB* promoter. We note that another sequence with good homology to consensus promoters can be found (both in *E. coli* and in *S. typhimurium*) 19 bp upstream of the sequence identified by Nakayama et al. (44). This sequence in *S. typhimurium* is TTGTGA . . . .(17 bases) . . . .CATGCT (CATGTT in *E. coli*). Whether either

```
Salmonella DNA sequence:                    -316  GATCGCCGACGCGAATTGCTCCACGCCGATGCCG
E. coli differences (starting from X):
-282  ACTTTACTGACCACACCCGCAGCTTCGGTTCCCAGGCCCGCAGGCAACGACGGGGGCGGATAGAGTCCGCTACGG

-207  ATATAGGTGTCGATGAAGTTGATACCAATGGCTTTGTTCTCAACCTGGATTTCGTGTTCCGCCGGTTCCGCTGGC

-132  GTAAACTCCACGGTCTGAAGCACTTCCGGACCACCATGCTTGTGAAATTCAATACGCGTTGCCATGCTTCCTCCA
                                  X G     G                T  T         T

-57   AAAGA------------------AATGTGGTAATCTTTCGACCCAATCACTATCTCGGTAACTCCATTCACTATG
      G  CGTATCGTCAGGGTCTGCTTC  A  A   AG          T  TT

1     AlaGlyAsnLysProPheAsnLysProGlnThrAspAlaArgAspArgAspProGlnValAlaGlyIleLysVal
1     GCAGGAAATAAACCCTTCAACAAACCACAGACTGATGCCCGCGACCGCGATCCGCAGGTTGCCGGGATAAAAGTG
                              AG   G    AC       A        A  A         C G
                            Gln   AlaGluPro   Glu                      Leu

26    ProProHisSerIleGluAlaGluGlnSerValLeuGlyGlyLeuMetLeuAspAsnGluArgTrpAspAspVal
76    CCGCCGCACTCGATTGAAGCGGAACAGTCGGTGTTGGGCGGTTTAATGCTGGATAACGAGCGCTGGGACGATGTG
        T           C        G                              A        A        T     A


51    AlaGluArgValValAlaGluAspPheTyrThrArgProHisArgHisIlePheThrGluMetGlyArgLeuGln
151   GCCGAGCGCGTGGTGGCGGAAGATTTCTATACCCGCCCGCATCGCCATATCTTTACGGAGATGGGGCGCTTGCAG
              T     A  A  C     T  C        A  C  T           T  A    C   T
                          Asp                                          Ala

76    GluSerGlySerProIleAspLeuIleThrLeuAlaGluSerLeuGluArgGlnGlyGlnLeuAspSerValGly
226   GAAAGCGGCAGTCCTATCGACCTGATTACGCTCGCGGAATCGCTGGAGCGGCAGGGCCAACTGGACAGCGTCGGC
              T  C        T        T  T              A        G     C  T        T


101   GlyPheAlaTyrLeuAlaGluLeuSerLysAsnThrProSerAlaAlaAsnIleSerAlaTyrAlaAspIleVal
301   GGTTTCGCCTATCTGGCGGAGCTGTCTAAAAACACGCCAAGCGCGGCGAATATCAGTGCTTATGCGGACATTGTG
           T  T         A        A     T        T     T  C         C             C


126   ArgGluArgAlaValValArgAspMetIleAlaValAlaHisGluIleAlaAspAlaGlyTyrAspProGlnGly
376   CGCGAACGCGCCGTGGTCCGCGATATGATCGCGGTGGCGCATGAAATTGCGGACGCCGGTTACGATCCACAGGGG
        T     T     T     T  G      T    T   A    G     C  A  T    TT     G
                          Glu       Ser       Asn         Glu       Phe      T=am34

151   ArgAsnSerAspGluLeuLeuAspLeuAlaGluSerArgValPheGlnIleAlaGluAsnArgAlaAsnLysAsp
451   CGCAATAGCGACGAACTGCTGGATCTGGCGGAGTCGCGCGTCTTCCAGATCGCGGAAAACCGGGCCAACAAAGAC
        T CC      A  T                 T  A  C         TA A  T  C     GT  T  G
         Thr    GluAsp                                  Lys          Ser

176   GluGlyProLysSerIleAspGlnIleLeuAspAlaThrValAlaArgIleGluGlnLeuPheGlnGlnProHis
526   GAAGGTCCGAAAAGCATCGACCAGATTCTCGACGCCACCGTGGCGCGTATTGAGCAGTTGTTCCAGCAACCGCAC
           G      G A      C G TG G         A                                T     G  A
                   Asn    AlaAspVal

201   AspGlyValThrGlyValAspThrGlyTyrGlnAspLeuAsnLysLysThrAlaGlyLeuGlnArgSerAspLeu
601   GATGGCGTTACAGGCGTGGATACCGGCTATCAGGATCTCAATAAAAAGACGGCAGGGTTACAGCGTTCGGATTTG
                C  G  AA C      T   G C         C      A  C  T  C  G    CG
                        Asn          Asp                                 Pro

226   IleIleValAlaAlaArgProSerMetGlyLysThrThrPheAlaMetAsnLeuCysGluAsnAlaAlaMetLeu
676   ATTATCGTCGCGGCGCGTCCCTCCATGGGTAAAACCACTTTTGCGATGAACCTCTGCGAAAATGCGGCGATGTTG
        C         C         G  G            A  A                 GT       C
                                                                  Val

251   GlnAspLysProValLeuIlePheSerLeuGluMetProGlyGluGlnIleMetMetArgMetLeuAlaSerLeu
751   CAAGATAAGCCGGTACTGATCTTTAGTCTGGAGATGCCCGGCGAACAGATCATGATGCGTATGCTGGCCTCGCTG
        G     A         T     CTCG              ATCA                    TCT     G
                                                  Ser                     Ser

276   SerArgValAspGlnThrArgIleArgThrGlyGlnLeuAspAspGluAspTrpAlaArgIleSerGlyThrMet
826   TCCCGCGTCGATCAGACACGTATTCGTACCGGTCAACTTGATGATGAAGACTGGGCGCGAATCTCCGGCACGATG
        G     T  C     TAAA  C         G  G  C     C                C  T        C
                       Lys
```

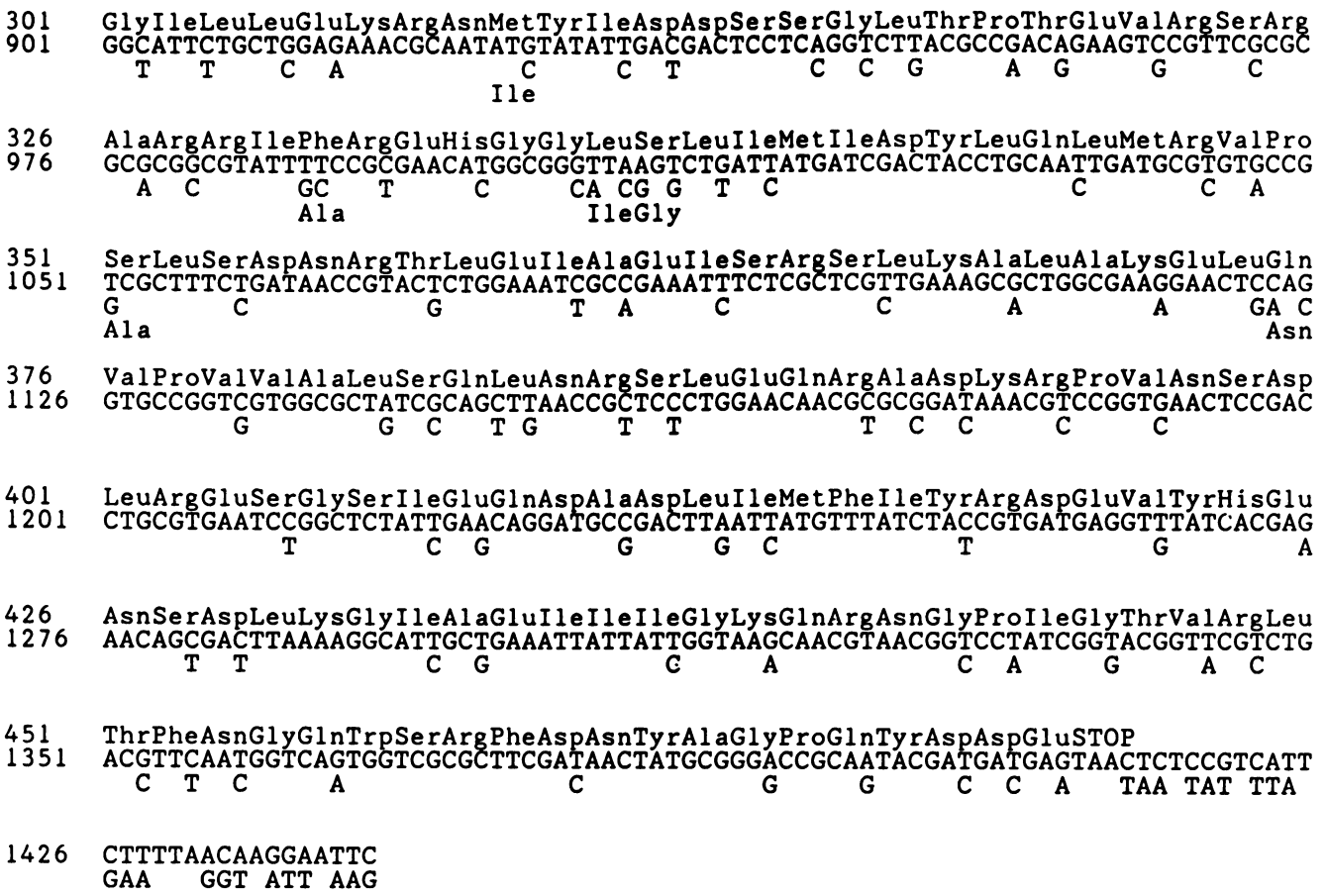

FIG. 2. DNA and inferred amino acid sequence of *S. typhimurium dnaB*, with the differences found in *E. coli*. The DNA sequence shown was determined on both strands. The underlined sequences specify a potential ribosome-binding site and initiation codon for the oppositely oriented open reading frame discussed in the text. The *dnaB34*(Am) mutation in codon 149 is also shown.

of these sequences is the *dnaB* promoter will have to be determined by mapping the 5′ end of the *dnaB* message. It is also possible that *dnaB*, like several other DNA replication genes, can be transcribed from multiple promoters (24, 34, 57).

For 22 nucleotides preceding the initiator ATG, the sequences in *E. coli* and *S. typhimurium* are identical. As noted previously (44), this region contains no extensive complementarity to the 3′ end of 16S rRNA, a factor that may limit translation of *dnaB* and thereby contribute to the low level of DnaB maintained in cells.

Another potential influence on *dnaB* expression is suggested by the presence of an open reading frame, oriented oppositely to *dnaB*, beginning with an ATG (CAT in the strand shown in Fig. 2) at −68. This reading frame is preceded at a distance of four nucleotides by a potential ribosome-binding site (GGAGG; CCTCC as shown) (23). This reading frame continues uninterrupted to the *Hin*dIII end of the *dnaB* fragment, including both the sequence shown and an additional 254 bp sequenced on only one strand. A data bank search did not identify this potential gene product but did reveal significant homology to several alcohol dehydrogenases. If this potential gene product is made in vivo (which cannot be evaluated with the existing plasmids because the reading frame is fused to *lacZ* out of frame), the position of its reading frame implies that either its promoter overlaps the *dnaB* promoter, or its message is complementary to the *dnaB* message. Either possibility could have regulatory significance.

Comparison between *E. coli* and *S. typhimurium* DnaB. The DnaB amino acid sequence is highly conserved between *E. coli* and *S. typhimurium*. Of the 470 residues, only 35 differ between the two species (93% identity). Among the 35 amino acid residue changes, many are conservative, including seven exchanges between glutamate and aspartate. There is no difference in the total number of basic residues (lysine plus arginine) or in the total number of acidic residues (glutamate plus aspartate). It is noteworthy, however, that the few amino acid changes between *E. coli* and *S. typhimurium* DnaB proteins are not distributed evenly across the protein. For example, the carboxy-terminal 95 amino acids are perfectly conserved; in contrast are short stretches exhibiting considerable variation (e.g., five changes out of seven in residues 9 through 15). The distribution of changes seems to be related to proteolytically defined domains of DnaB protein from *E. coli*. The two large trypsin-resistant domains (residues 15 through 125 or 127 and 172 through 470; 45) are 96.4 and 94.0% conserved, respectively, whereas the remainder of DnaB is only 76% conserved. These results indicate that the proteolytically defined domains are important functional units. Nonetheless, specific mutational data also support a functional role for the remainder of DnaB. In strains having synthetic cysteine- or phenylalanine-inserting amber suppressor tRNA molecules (46),

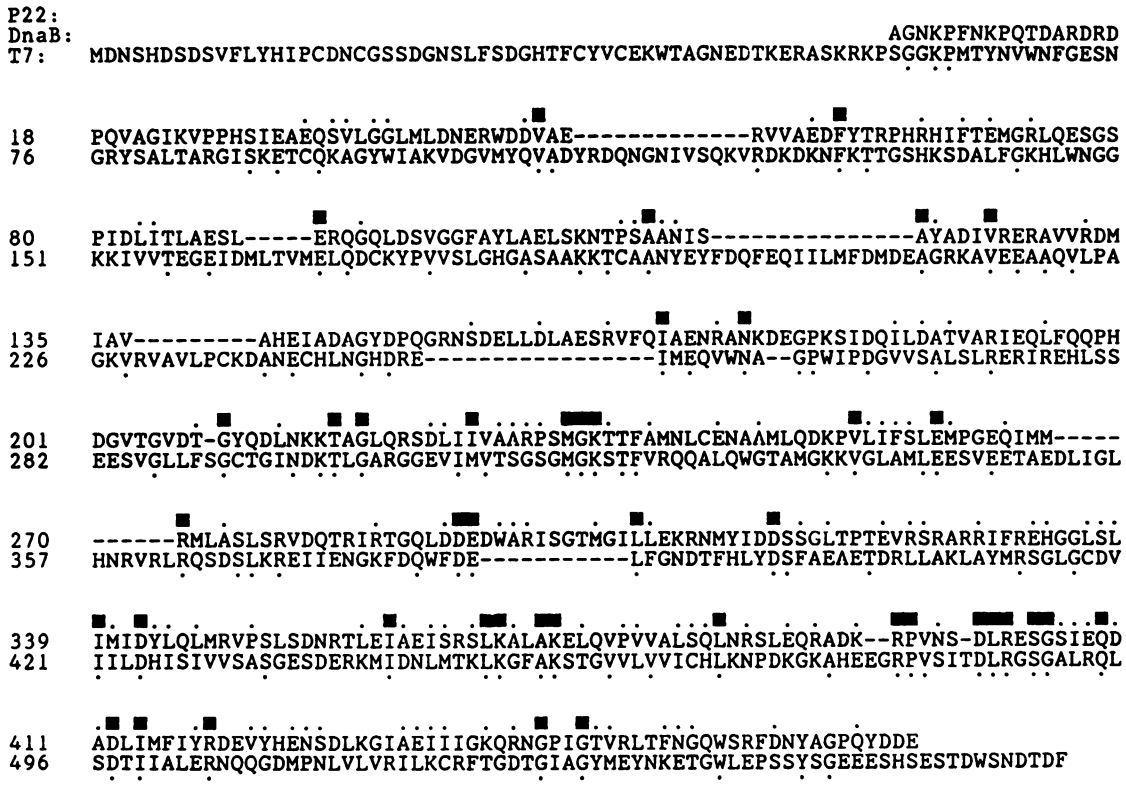

FIG. 3. Alignment of amino acid sequences of *S. typhimurium* DnaB and the gene 4 helicase-primase of phage T7 (17). The longer of the two gene 4 polypeptides is used (16). The alignment in the figure was obtained using the PRTALN program (67) with parameters set at K-tuple = 1, window = 40, and gap penalty = 2. Amino acid identities are marked by dots below the alignment. Marks above the alignment denote identities between *S. typhimurium* DnaB and P22 gene 12 by using the alignment reported by Backhaus and Petri (6). Bars denote identities in all three proteins. Dashes indicate insertions needed to maintain the alignment.

dnaB34(Am) is unable to complement a *dnaB*(Ts) mutation, a result suggesting that substitution of cysteine or phenylalanine for glutamine at position 149 produces nonfunctional DnaB (data not shown). Moreover, *dnaB* mutations exhibiting specific defects in function affect residues Ile-135, Ile-141, and Leu-156 (all conserved residues in the trypsin-sensitive connector), as well as other sites in DnaB (R. Maurer and A. Wong, manuscript submitted).

Codon usage in *S. typhimurium dnaB* scarcely differs from that in *E. coli dnaB*. Codon usage is not thought to play a major role in limiting expression of DnaB (44).

Comparison with gene 4 helicase-primase of phage T7. Figure 3 shows an alignment of the *S. typhimurium* DnaB amino acid sequence with that of the helicase-primase encoded by coliphage T7 (29, 55). Identities are marked by dots below the alignment. Particularly notable are two extended stretches of similarity: 26 of 99 identical residues corresponding to amino acids 167 through 265 in DnaB, and 28 of 107 identical residues corresponding to amino acids 313 through 419 in DnaB. In each stretch, only one or two short (1- to 2-amino-acid) insertions are required to maintain the alignment. Thus, as is also true of the alignment of DnaB with the analogous P22 gene 12 protein (6), the greatest similarity is within the C-terminal domain of DnaB.

No homology of DnaB with UvrD or Rho. Helicase II, the *uvrD* product, has its major in vivo function in connection with DNA repair, but under some circumstances it may also participate in DNA replication (for reviews, see references 32 and 41). According to early reports, helicase II, like DnaB, migrates 5′ to 3′ along the single strand to which it is bound (31); a more recent study inferred 3′-to-5′ migration (36). An attempt to align the amino acid sequence of helicase II from *E. coli* (19 [corrected in reference 69]) with DnaB from either *E. coli* or *S. typhimurium* failed because of insufficient similarity.

Transcription termination factor Rho from *E. coli*, the product of the *rho* gene, is an RNA-DNA helicase that attaches to single-stranded RNA and migrates in a 5′-to-3′ direction to invade the RNA-DNA hybrid region near a transcribing RNA polymerase (11). Like DnaB, Rho monomers join into functional hexamers, and also like DnaB, Rho monomers consist of two protease-resistant domains (9, 18, 20; for a review, see reference 68). Despite these similarities in functional aspects of DnaB and Rho, their amino acid sequences (50) could not be aligned.

DISCUSSION

The DNA sequence data presented in this paper, together with functional data presented previously (39), firmly establish *S. typhimurium* DnaB as a member of the family of proteins whose prototype is the *E. coli* DnaB protein. Other proteins can be considered members of this family if they substitute for DnaB in host chromosome replication, or if they show significant primary amino acid sequence similarity to DnaB and share some distinctive aspect of DnaB function. According to these criteria, other members of the DnaB family include the Ban protein of coliphage P1 (12), one or more DnaB analog proteins encoded by conjugative R plasmids (62), the helicase-primase of coliphage T7 (29, 55; see

Fig. 3), and the gene 12 product of *Salmonella* phage P22 (6, 64). DNA helicase II and transcription termination factor Rho, despite provocative partial overlap with DnaB at a functional level, do not exhibit primary amino acid sequence similarity with DnaB and cannot be considered members of the DnaB family at this time.

To date, *S. typhimurium dnaB* is the only functionally interchangeable *dnaB* analog whose nucleotide sequence has been determined. Comparison of the DnaB sequences from *E. coli* and *S. typhimurium* provides a general indication that most amino acid residues in DnaB are important to its correct functioning. The codon position distribution of nucleotide sequence changes offers additional evidence that the close similarity of the DnaB proteins is not a trivial result of a lack of opportunity for evolutionary divergence. If we assume, for simplicity, that all changes at position 3 are silent at the protein level, whereas changes in position 1 or 2 of codons are not silent, then the excess of nucleotide changes found at position 3 reflects the average likelihood that an amino acid change anywhere in the protein would be selected against during evolution. In fact, there is a 10-fold excess of changes at position 3 over changes at position 1 or 2 (195 versus 24 and 14, respectively). This fact implies that the vast majority of changes at positions 1 and 2 occurring during evolution must have been deleterious to DnaB function. From this information, we can estimate that 9 out of 10 amino acid residues in DnaB cannot be changed without significantly damaging function.

A more specific appreciation of the role of individual amino acid residues might be obtained by comparing DnaB with other proteins that share only a subset of DnaB activities and exhibit a lesser degree of amino acid primary sequence identity. An instructive comparison is possible among DnaB, the gene 12 protein of *Salmonella* phage P22, and the gene 4 protein of coliphage T7. These three proteins are DNA-dependent (DNA-stimulated) ATPases (2, 29, 52, 64, 66). DnaB and T7 gene 4 are helicases (29, 33, 55) and a similar activity appears likely for P22 gene 12 (64). These proteins also have their unique features; only DnaB interacts with DnaC (28, 54), only P22 gene 12 interacts with P22 gene 18 (26), and only T7 gene 4 possesses primase activity (55). Therefore, regions that are conserved across the three proteins may represent core elements of an ATPase-helicase. When the separately determined alignments of *S. typhimurium* (or *E. coli*) DnaB with T7 gene 4 (dots below the alignment in Fig. 3) and with P22 gene 12 (6; dots and bars above the alignment in Fig. 3) are compared, two regions of significant common identity emerge. These regions are evident from inspection of Fig. 3, where all common identities are designated by a bar above the alignment.

One of these regions contains the conserved triad Met-Gly-Lys, beginning with residue 234 in DnaB. This sequence is immediately followed by threonine in DnaB and gene 12 and by serine in gene 4. This sequence resembles the ATP-binding site consensus sequence, GXXXGK(T/S) (21, 61), differing only by substitution of alanine or threonine for the initial glycine. Moreover, the predicted secondary structure surrounding this region in DnaB (44) conforms to the consensus topography predicted for ATP-binding domains by Bradley et al. (10). Finally, this region of DnaB is included in the proteolytic fragment that exhibits ATP binding and hydrolysis (45). For these reasons, it is likely that this conserved sequence represents a portion of the ATP-binding domain. This sequence in DnaB was also noted as a potential nucleotide-binding motif by Finch and Emmerson (19).

The second highly conserved region (27% amino acid identity in the three proteins) corresponds to 60 amino acids (350 through 419) in DnaB. This region, or some portion of it, could represent a helicase motif. When we compared this 60-residue subsequence with UvrD or Rho, we found limited similarity, 9 identical residues with UvrD residues 354 through 413 and 11 identities with Rho residues 104 through 163. The significance of these similarities remains to be determined.

In view of the multiple functions of DnaB (helicase, ATPase, DNA binding, oligomerization, and interactions with several bacterial and phage replication proteins), much remains to be learned about the specific contribution of particular amino acid residues. Sequence comparisons, although useful, can only suggest possibilities. With the wild-type sequence now available, and given suitable genetic and biochemical assays, it should be possible to examine the deficiencies in specific *dnaB* mutants.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant AI-19942 from the National Institutes of Health. L.K. was supported, in part, by the Summer Undergraduate Research Program at Case Western Reserve University.

LITERATURE CITED

1. **Arai, K., and A. Kornberg.** 1979. A general priming system employing only DnaB protein and primase for DNA replication. Proc. Natl. Acad. Sci. USA **76**:4308–4312.
2. **Arai, K., and A. Kornberg.** 1981. Mechanism of DnaB protein. II. ATP hydrolysis by DnaB protein dependent on single- or double-stranded DNA. J. Biol. Chem. **256**:5253–5259.
3. **Arai, K., R. Low, J. Kobori, J. Shlomai, and A. Kornberg.** 1981. Mechanism of DnaB protein action. V. Association of DnaB protein, n′ and other prepriming proteins in the primosome of DNA replication. J. Biol. Chem. **256**:5260–5266.
4. **Arai, K., R. McMacken, S. Yasuda, and A. Kornberg.** 1981. Purification and properties of *Escherichia coli* protein i, a prepriming protein in ϕX174 DNA replication. J. Biol. Chem. **256**:5281–5286.
5. **Arai, K., S. Yasuda, and A. Kornberg.** 1981. Mechanism of DnaB protein action. I. Crystallization and properties of DnaB protein, an essential replication protein in *Escherichia coli*. J. Biol. Chem. **256**:5247–5252.
6. **Backhaus, H., and J. B. Petri.** 1984. Sequence analysis of a region from the early right operon in phage P22 including the replication genes *18* and *12*. Gene **32**:289–303.
7. **Baker, T. A., B. E. Funnell, and A. Kornberg.** 1987. Helicase action of dnaB protein during replication from the *Escherichia coli* chromosomal origin *in vitro*. J. Biol. Chem. **262**:6877–6885.
8. **Baker, T. A., K. Sekimizu, B. E. Funnell, and A. Kornberg.** 1986. Extensive unwinding of the plasmid template during staged enzymatic initiation of DNA replication from the origin of the *Escherichia coli* chromosome. Cell **45**:53–64.
9. **Bear, D. G., C. L. Andrews, J. D. Singer, W. D. Morgan, R. A. Grant, P. H. von Hippel, and T. Platt.** 1985. *Escherichia coli* transcription termination factor rho has a two-domain structure in its activated form. Proc. Natl. Acad. Sci. USA **82**:1911–1915.
10. **Bradley, M. K., T. F. Smith, R. H. Lathrop, D. M. Livingston, and T. A. Webster.** 1987. Consensus topography in the ATP binding site of the simian virus 40 and polyomavirus large tumor antigens. Proc. Natl. Acad. Sci. USA **84**:4026–4030.
11. **Brennan, C. A., A. J. Dombroski, and T. Platt.** 1987. Transcription termination factor Rho is an RNA-DNA helicase. Cell **48**:945–952.
12. **D'Ari, R., A. Jaffé-Brachet, D. Touati-Schwartz, and M. Yarmolinsky.** 1975. A *dnaB* analog specified by bacteriophage P1. J. Mol. Biol. **94**:341–366.

13. **Davis, R. W., D. Botstein, and J. R. Roth.** 1980. Advanced bacterial genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
14. **Dodson, M., H. Echols, S. Wickner, C. Alfano, K. Mensa-Wilmot, B. Gomes, J. LeBowitz, J. D. Roberts, and R. McMacken.** 1986. Specialized nucleoprotein structures at the origin of replication of bacteriophage λ: localized unwinding of duplex DNA by a six-protein reaction. Proc. Natl. Acad. Sci. USA **83**:7638–7642.
15. **Dodson, M., J. Roberts, R. McMacken, and H. Echols.** 1985. Specialized nucleoprotein structures at the origin of replication of bacteriophage λ: complexes with λ O, λ P, and *Escherichia coli* DnaB proteins. Proc. Natl. Acad. Sci. USA **82**:4678–4682.
16. **Dunn, J. J., and W. F. Studier.** 1981. Nucleotide sequence from the genetic left end of bacteriophage T7 DNA to the beginning of gene 4. J. Mol. Biol. **148**:303–330.
17. **Dunn, J. J., and F. W. Studier.** 1983. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. J. Mol. Biol. **166**:477–535.
18. **Engel, D., and J. P. Richardson.** 1984. Conformational alterations of transcription termination protein rho induced by ATP and by RNA. Nucleic Acids Res. **12**:7389–7400.
19. **Finch, P. W., and P. T. Emmerson.** 1984. The nucleotide sequence of the *uvrD* gene of *E. coli*. Nucleic Acids Res. **12**: 5789–5799.
20. **Finger, L. R., and J. P. Richardson.** 1982. Stabilization of the hexameric form of *Escherichia coli* protein rho under ATP hydrolysis conditions. J. Mol. Biol. **156**:203–219.
21. **Fry, D. C., S. A. Kuby, and A. S. Mildvan.** 1986. ATP-binding site of adenylate kinase: mechanistic implications of its homology with *ras*-encoded p21, F_1-ATPase, and other nucleotide-binding proteins. Proc. Natl. Acad. Sci. USA **83**:907–911.
22. **Fuller, R. S., B. E. Funnell, and A. Kornberg.** 1984. The DnaA protein complex with the *E. coli* chromosomal replication origin (*oriC*) and other DNA sites. Cell **38**:889–900.
23. **Gold, L., and G. Stormo.** 1987. Translational initiation, p. 1302–1307. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
24. **Hansen, E. B., F. G. Hansen, and K. von Meyenburg.** 1982. The nucleotide sequence of the *dnaA* gene and the first part of the *dnaN* gene of *Escherichia coli* K12. Nucleic Acids Res. **10**: 7373–7385.
25. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids. Res. **11**:2237–2255.
26. **Hilliker, S., and D. Botstein.** 1976. Specificity of genetic elements controlling regulation of early functions in temperate bacteriophages. J. Mol. Biol. **106**:537–566.
27. **Hoopes, B. C., and W. R. McClure.** 1987. Strategies in regulation of transcription initiation, p. 1231–1240. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
28. **Kobori, J. A., and A. Kornberg.** 1982. The *Escherichia coli* *dnaC* gene product. III. Properties of the DnaB-DnaC protein complex. J. Biol. Chem. **257**:13770–13775.
29. **Kolodner, R., and C. C. Richardson.** 1977. Replication of duplex DNA by bacteriophage T7 DNA polymerase and gene *4* protein is accompanied by hydrolysis of nucleoside 5′-triphosphates. Proc. Natl. Acad. Sci. USA **74**:1525–1529.
30. **Kornberg, A.** 1980. DNA replication. W.H. Freeman and Co., San Francisco.
31. **Kuhn, B., M. Abdel-Monem, H. Krell, and H. Hoffman-Berling.** 1979. Evidence for two mechanisms for DNA unwinding catalyzed by DNA helicases. J. Biol. Chem. **254**:11343–11350.
32. **Kushner, S. R.** 1987. DNA repair, p. 1044–1053. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
33. **LeBowitz, J. H., and R. McMacken.** 1986. The *Escherichia coli* DnaB replication protein is a DNA helicase. J. Biol. Chem. **261**:4738–4748.
34. **Lupski, J. R., A. A. Ruiz, and G. N. Godson.** 1983. Regulation of the *rpsU-dnaG-rpoD* macromolecular synthesis operon of *E. coli* K-12. Mol. Gen. Genet. **195**:391–401.
35. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
36. **Matson, S. W.** 1986. *Escherichia coli* helicase II (*uvrD* gene product) translocates unidirectionally in a 3′ to 5′ direction. J. Biol. Chem. **261**:10169–10175.
37. **Maurer, R., B. J. Meyer, and M. Ptashne.** 1980. Gene regulation at the right operator (O_R) of bacteriophage λ. I. O_R3 and autogenous negative control by repressor. J. Mol. Biol. **139**: 147–161.
38. **Maurer, R., B. C. Osmond, and D. Botstein.** 1984. Genetic analysis of DNA replication in bacteria: *dnaB* mutations that suppress *dnaC* mutations and *dnaQ* mutations that suppress *dnaE* mutations in *Salmonella typhimurium*. Genetics **108**:25–38.
39. **Maurer, R., B. C. Osmond, E. Shektman, A. Wong, and D. Botstein.** 1984. Functional interchangeability of DNA replication genes in *Salmonella typhimurium* and *Escherichia coli* demonstrated by a general complementation procedure. Genetics **108**:1–23.
40. **McMacken, R., C. Alfano, B. Gomes, J. H. LeBowitz, K. Mensa-Wilmot, J. D. Roberts, and M. Wold.** 1987. Biochemical mechanisms in the initiation of bacteriophage λ DNA replication, p. 227–245. *In* T. J. Kelly and R. McMacken (ed.), DNA replication and recombination. Alan R. Liss, Inc., New York.
41. **McMacken, R., L. Silver, and C. Georgopoulos.** 1987. DNA replication, p. 564–612. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 1. American Society for Microbiology, Washington, D.C.
42. **McMilin, K. D., and V. E. A. Russo.** 1972. Maturation and recombination of bacteriophage lambda DNA molecules in the absence of DNA duplication. J. Mol. Biol. **68**:49–55.
43. **Messing, J.** 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9**:309–321.
44. **Nakayama, N., N. Arai, M. W. Bonds, Y. Kaziro, and K. Arai.** 1984. Nucleotide sequence of *dnaB* and the primary structure of the *dnaB* protein from *Escherichia coli*. J. Biol. Chem. **259**:97–101.
45. **Nakayama, N., N. Arai, Y. Kaziro, and K. Arai.** 1984. Structural and functional studies of the *dnaB* protein using limited proteolysis. Characterization of domains for DNA-dependent ATP hydrolysis and for protein association in the primosome. J. Biol. Chem. **259**:88–96.
46. **Normanly, J., J.-M. Masson, L. G. Kleina, J. Abelson, and J. H. Miller.** 1986. Construction of two *Escherichia coli* amber suppressor genes: phenylalanine transfer RNA suppressor and cysteine transfer RNA suppressor. Proc. Natl. Acad. Sci. USA **83**:6548–6552.
47. **Norrander, J., T. Kempe, and J. Messing.** 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene **26**:101–106.
48. **Nossal, N. G.** 1983. Prokaryotic DNA replication systems. Annu. Rev. Biochem. **53**:581–615.
49. **Nossal, N. G., and B. M. Alberts.** 1983. Mechanism of DNA replication catalyzed by purified T4 replication proteins, p. 71–81. *In* C. K. Mathews, E. M. Kutter, G. Mosig, and P. B. Berget (ed.), Bacteriophage T4. American Society for Microbiology, Washington, D.C.
50. **Pinkham, J. L., and T. Platt.** 1983. The nucleotide sequence of the *rho* gene of *E. coli* K-12. Nucleic Acids Res. **11**:3531–3545.
51. **Reha-Krantz, L. J., and J. Hurwitz.** 1978. The *dnaB* gene product of *Escherichia coli*. I. Purification, homogeneity, and physical properties. J. Biol. Chem. **253**: 4043–4050.
52. **Reha-Krantz, L. J., and J. Hurwitz.** 1978. The *dnaB* gene

product of *Escherichia coli*. II. Single-stranded DNA-dependent ribonucleoside triphosphatase activity. J. Biol. Chem. **253:** 4051–4057.

53. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
54. **Schanda-Mulfinger, U. E. M., and H. Schmieger.** 1980. Growth of *Salmonella* bacteriophage P22 in *Escherichia coli dna*(Ts) mutants. J. Bacteriol. **143:**1042–1045.
55. **Scherzinger, E., E. Lanka, G. Morelli, D. Seiffert, and A. Yuki.** 1977. Bacteriophage T7-induced DNA-priming protein. Eur. J. Biochem. **72:**543–558.
56. **Shortle, D.** 1983. A genetic system for analysis of staphylococcal nuclease. Gene **22:**181–189.
57. **Tomasiewicz, H. G., and C. S. McHenry.** 1987. Sequence analysis of the *Escherichia coli dnaE* gene. J. Bacteriol. **169:** 5735–5744.
58. **Ueda, K., R. McMacken, and A. Kornberg.** 1978. *dnaB* protein of *Escherichia coli*. J. Biol. Chem. **253:**261–269.
59. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19:**259–268.
60. **Vogelstein, B., and D. Gillespie.** 1979. Preparative and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. USA **76:**615–619.
61. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay.** 1982. Partly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1:**945–951.
62. **Wang, P. Y., and V. N. Iyer.** 1978. Analogs of the *dnaB* gene of *Escherichia coli* K-12 associated with conjugative R plasmids. J. Bacteriol. **134:**765–770.
63. **Wickner, S.** 1979. DNA replication proteins of *Escherichia coli* and phage λ. Cold Spring Harbor Symp. Quant. Biol. **43:**303–310.
64. **Wickner, S.** 1984. DNA-dependent ATPase activity associated with phage P22 gene 12 protein. J. Biol. Chem. **259:**14038–14043.
65. **Wickner, S.** 1984. Oligonucleotide synthesis by *Escherichia coli dnaG* primase in conjunction with phage P22 gene 12 protein. J. Biol. Chem. **259:**14044–14047.
66. **Wickner, S., M. Wright, and J. Hurwitz.** 1974. Association of DNA-dependent and independent ribonucleoside triphosphatase activities with the *dnaB* gene product of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **71:**783–787.
67. **Wilbur, W. J., and D. J. Lipman.** 1983. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA **80:**726–730.
68. **Yager, T. D., and P. H. von Hippel.** 1987. Transcript elongation and termination in *Escherichia coli*, p. 1241–1275. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
69. **Yamamoto, Y., T. Ogawa, H. Shinagawa, T. Nakayama, H. Matsuo, and H. Ogawa.** 1986. Determination of the initiation sites of transcription and translation of the *uvrD* gene of *Escherichia coli*. J. Biochem. **99:**1579–1590.