Nucleotide sequence of a type II DNA topoisomerase gene. Bacteriophage T4 gene 39

Wai Mun Huang

Department of Cellular, Viral and Molecular Biology, University of Utah Medical Center, Salt Lake City, UT 84132, USA

Received 11 July 1986; Revised and Accepted 3 September 1986

ABSTRACT

T4 DNA topoisomerase is a type II enzyme and is thought to be required for normal T4 DNA replication. T4 gene 39 codes for the largest of the three subunits of T4 DNA topoisomerase. I have determined the nucleotide sequence of a region of 2568 nucleotides of T4 DNA which includes gene 39. The location of the gene was established by the identification of the first fifteen amino acids in the large open reading frame in the DNA sequence as those found at the amino-terminus of the purified 39-protein. The coding region of gene 39 has 1560 bases, and it is followed by two in-frame stop codons. The gene is preceded by a typical Shine-Dalgarno sequence as well as possible promoter sequences for *E. coli* RNA polymerase. T4 39-protein consists of 520 amino acids, and it has a calculated molecular weight of 58,478. By comparing the amino acid sequences, T4 39-protein is found to share homology with the *gyrB* subunit of DNA gyrase. This suggests that these topoisomerase subunits may be equivalent functionally. Some of the characteristics of the 39-protein and its structural features predicted from the DNA sequence data are discussed.

INTRODUCTION

T4 DNA topoisomerase is an ATP-dependent multisubunit enzyme. Mechanistically, it causes transient double-stranded breaks in the DNA backbone through which DNA is passed, resulting in changes in DNA topology. The enzyme consists of three subunits which are the products of T4 genes 39, 52 and 60 (1, 2, 3). These gene products are recognized genetically as proteins required for normal T4 DNA replication and are postulated to be involved in the initiation event of T4 chromosomal DNA replication (1, 4), although the exact roles of these proteins in phage DNA metabolism and development have yet to be clearly defined. Purified T4 DNA topoisomerase exhibits multiple activities *in vitro*; these include (i) The DNA strand passage reaction which results in the relaxation of superhelical DNA and the related reactions of knotting, unknotting, catenation and decatenation of circular DNA; (ii) A DNA-dependent ATPase, (iii) DNA binding, (iv) The induced double-stranded

DNA cleavage reaction; (v) Covalent transfer of one of the protein subunits to the 5'-ends of the cleaved DNA (for a review, see ref. 5). In order to decipher whether individual subunits can carry out any partial reactions, and to investigate if these subunits have functions in addition to being components of the phage DNA topoisomerase we began the molecular cloning of the three genes separately into E. coli cells. T4 gene 39 codes for the largest of the three subunits. It has been shown that it is a positive determinant which regulates the specificity of T even phage DNA replication (6). More recently, we have purified the 39-protein subunit. It is the ATPase portion of the T4 DNA topoisomerase (7).

For an in depth investigation of the interaction of the 39-protein with other subunits of the topoisomerase and possibly with other replication proteins, the availability of detailed information concerning its primary structure is of great interest. DNA sequence information will also greatly enhance our ability to manipulate the gene. Although T4 DNA topoisomerase is a prokaryotic enzyme, its in vitro enzymatic activity differs from the other prokaryotic type II DNA topoisomerase, namely DNA gyrases, in some fundamental way. DNA gyrases are more proficient in introducing supercoils into closed circular DNA than in removing the superhelical twists. The latter reaction is ATP-independent. On the other hand, the T4 enzyme can only remove superhelical twists in an ATP-dependent reaction. In this regard, the T4 enzyme shares more similarities with the eukaryotic DNA topoisomerases than the prokaryotic type II DNA topoisomerase. Thus, it may serve as a prototype for other type II non-gyrase DNA topoisomerases which have been identified in diverse cell types (8). I report here the nucleotide sequence of the coding and flanking regions of T4 gene 39. The reading frame of the coding sequence was confirmed by independently sequencing the N-terminal amino acids of the purified 39-protein. In spite of the apparent difference in enzymatic activities between the prokaryotic type II topoisomerases, the amino acid sequence of the T4 39-protein shares significant homology with the gyrB subunit of B. subtilis DNA gyrase suggesting that they are equivalent functionally and possibly similar in structure. Some characteristics of the protein and the possible regulation of the expression of the gene, deduced from the nucleotide sequence, are discussed.

MATERIALS AND METHODS

Bacterial and Phage Strains. The *Escherichia coli* strain JM101 (9) was used as the recipient for M13 phage DNA transformation and as the host for the propagation of M13 phages. *E. coli* 802 (supE , $\text{r}_k^- \text{m}_k^+$) (10) was used as the recipient for plasmid DNA transformations. Phage M13 derivatives mp8 and mp9 (11) were obtained from New England Biolabs.

Enzymes and Biochemicals. Restriction enzymes were purchased from Bethesda Research Laboratories and New England Biolabs. T4 DNA ligase was prepared from *E. coli* K1100 endo I⁻ lysogenized with the lambda phage NM989 Wam403 Eam110 cIts857 nin5 Sam100 (12, kindly provided by G. Wilson) according to a published procedure (13). *E. coli* DNA polymerase I Klenow fragment was purchased from Bethesda Research Laboratories. Polynucleotide kinase, deoxy- and dideoxynucleotides were purchased from P-L Biochemicals. [α -³²P]-dATP were obtained from New England Nuclear and [α -³⁵S]-thio-dATP was from Amersham.

DNA Sequencing. Both the chemical modification method of Maxam and Gilbert (14) and the chain-termination method of Sanger (15) were used according to published procedures. In the latter method, recombinant M13 phage DNA carrying inserts of T4 DNA were used as single-stranded DNA templates, and a 15-mer primer (New England Biolabs) was used in the chain-termination reactions. When [α -³⁵S]-thio-dATP was used as the label in the DNA sequencing reactions, the modifications described by Biggin et al. (16) were used, and the sequencing gels were dried using a slab gel dryer (Bio-Rad) prior to autoradiography.

Purification of T4 39-Protein. T4 39-protein was purified from *E. coli* cells which harbor a plasmid with the T4 gene 39 under the control of the P_L promoter of coliphage lambda (17) according to the procedure of Nicholson and Huang (7). For amino acid composition determination, 39-protein was further fractionated by preparative slab gel electrophoresis in SDS according to a published procedure (18).

Other Methods. Plasmid DNAs and intracellular RF (replicative form) DNAs of M13 phages were prepared according to a standard cleared lysate procedure followed by banding in CsCl/ethidium bromide gradients. Restriction enzyme digestions were performed according to the recommendations of the commercial suppliers. For repairing the protruding ends of the restriction fragments, the following conditions were used: The reaction mixture (0.04 ml) contained 50 mM Tris-HCl (pH

7.5), 10 mM $MgCl_2$, 0.4 mM 4dNTP, 1 mM dithiothreitol, 5-10 ug DNA fragment and 3 units of DNA polymerase I Klenow fragment. Incubation was for 10 min at 20°C. The reaction was stopped by heating to 65°C for 15 min. The amino-terminal sequence analysis of T4 39-protein was performed using a Beckman 890C sequencer. The amino acid composition of the protein was determined after hydrolysis with hydrochloric acid for 24 or 48 hrs at 110°C using a Beckman amino acid analyzer. The DNA sequence was analyzed using the SEQ system (19) with a DEC-20 computer. The promoter search was done using the program described by Mulligan et al. (20). The protein alignment was analyzed using the Lipman and Wilbur algorithm (21).

RESULTS AND DISCUSSION

Cloning and Sequence Analysis of gene 39. The source of DNA for this study was the EcoRI fragment number 19 (22) of cytosine-containing T4 DNA (23). This fragment, which is approximately 3 kb in length, when

Figure 1. Restriction map of the EcoRI fragment number 19 of T4 DNA and the sequencing strategy of the region. The 3 kb EcoRI fragment is shown inserted into pBR322 and the location of the 39-protein corresponds to the shaded area. The neighboring two HindIII fragments of 0.68 kb and 1.8 kb are expanded to show the subcloning (see text for detail) and sequencing strategies. Open arrows represent sequences that were obtained by method of Maxam and Gilbert and closed arrows cover regions whose sequences were obtained from Sanger dideoxy chain-termination method.

inserted into a plasmid, was found to be capable of marker rescuing T4 phages defective in gene 39 (24). It was concluded that this fragment harbored the entire 39 gene based on the following observations: (i) Plasmids harboring this fragment are capable of complementing T4 amber mutants in gene 39 under non-permissive conditions (6). (ii) When the EcoRI fragment was cloned into a pBR322 derivative plasmid under the transcriptional control of the *E. coli* UV5 lac promoter (25), a protein of approximately 60,000 in SDS-molecular weight was synthesized. It was identical in size and antigenically related to the 39-protein subunit of the complete T4 DNA topoisomerase (3). Figure 1 shows a restriction map of the T4 EcoRI fragment inserted into the unique EcoRI site of pBR322 (26). T4 gene 39 was found to span two neighboring HindIII fragments of 0.68 and 1.8 kb. To facilitate nucleotide sequence analysis, these two HindIII fragments were subcloned into HindIII-cleaved pBR322. DNA fragments were isolated and sequenced by the chemical modification method of Maxam and Gilbert (14). These fragments were also cloned into the phage M13 vector mp8 (11). Restriction enzyme analysis of the 1.8 kb HindIII fragment indicated the presence of five AluI sites, shown in Fig. 1. The 1.8 kb HindIII fragment, after its protruding 5'-ends were repaired to blunt ends with *E. coli* DNA polymerase I Klenow fragment, was digested with AluI, and the resulting six fragments were then cloned into the HincII site of mp9 in a shotgun fashion. The AluI junctions were crossed using subclones containing T4 DNA fragments generated by the restriction enzymes MluI and EcoRV whose recognition sites are present only once within the 1.8 kb HindIII region (Fig. 1). Similarly, the protruding 5'-ends of these fragments were repaired with *E. coli* DNA polymerase I Klenow fragment before they were inserted into the blunt-ended HincII site of the M13 vector mp9. All the T4 DNA inserted into M13 was sequenced by the dideoxy chain-termination method described by Sanger et al. (17). A summary of the sequencing strategy is presented in Figure 1.

Nucleotide Sequence of Gene 39 and the Derived Primary Structure of 39-protein. The nucleotide sequence determined for these 2565 bp, including the T4 39 gene is shown in Figure 2. The reported sequence starts from an EcoRI site and ends at a HindIII site. Using both the chemical modification method and the chain-termination method, about 90% of the final sequence was determined from both strands and the entire region has been covered more than twice. Within this sequence, a large

Figure 2. Nucleotide sequence surrounding T4 gene 39 and the corresponding amino acid sequence of 39-protein. The sequence is numbered from the left end of the EcoRI fragment and the sequence of 2565 nucleotides extends to the right end of the 1.8 kb HindIII fragment (as defined in Fig. 1). The assignment of the reading frame of gene 39 was confirmed by the amino-terminal analysis of the protein.

open reading frame starting with an ATG codon at position 600 and terminating after position 2159 with two consecutive, in-frame stop codons was found. The N-terminal fifteen amino acids of purified 39-protein were determined by amino acid sequencing and were found to be Met-Ile-Lys-Asn-Glu-Ile-Lys-Ile-Leu-Ser-Asp-Ile-Glu-His-Ile. These are in perfect agreement with the first fifteen amino acids deduced from the open reading frame. Based on the DNA sequence, the 39-protein consists

Table I
Amino Acid Composition of 39-Protein

	DNA sequence	Amino acid analysis ^b
Ala	30	33
Arg	26	23
Asx ^a	63	62
Cys	7	— ^d
Glx ^a	52	60
Gly	41	— ^c
His	13	12
Ile	49	44
Leu	34	35
Lys	45	38
Met	18	17
Phe	20	22
Pro	15	15
Ser	25	24
Thr	30	28
Trp	8	— ^d
Tyr	8	10
Val	36	32
total	520	

a. The values for Asx and Glx represent the sums of aspartic acid + asparagine and glutamic acid + glutamine, respectively.

b. The values determined by amino acid analysis were obtained after hydrolysis with 6M HCl for 24 and 48 hrs at 110°C.

c. The actual value obtained for glycine was 75. It was not considered as a reliable value because the protein used for the analysis was extracted from SDS-polyacrylamide gel in which Tris-glycine was used as the running buffer.

d. Not determined.

of 520 amino acids, and it has a calculated molecular mass of 58,478 daltons. Table I summarizes the amino acid composition of 39-protein derived from the DNA sequence. It is in general agreement with the amino acid composition independently determined using purified 39-protein.

The frequency of codon usage in gene 39 is given in Table II. For comparison, the codon usage of the sum of four T4 replication genes, namely genes 44, 62, 45 and a portion of gene 43, reported by Spicer and Konisberg (27), as well as a composite of *E. coli* ribosomal proteins, chosen as typical representatives of abundant *E. coli* proteins, reported by Post and Nomura (28) are also listed. Since all the T4 proteins listed here are synthesized immediately or shortly after phage infection (the pre-replicative proteins), it is reasonable to expect that many of the abundant isoaccepting tRNA species of *E. coli*, such as those recognizing the anticodons ACU (Thr), GUU (Val), AAA (Lys), GAA (Glu), CGU (Arg) and GGU (Gly) would be used preferentially (as is seen for many of the sequenced *E. coli* proteins (28, 29)), and this is in fact what is seen. However, there are other codons that are recognized by

Table II Frequency of Codon Usage in T4 Gene 39

codon	Gene 39 ^a	T4 replication Genes ^b	R-protein Genes ^c	codon	Gene 39 ^a	T4 replication Genes ^b	R-protein Genes ^c
Phe TTT	14	24	9	Tyr TAT	5	21	4
Phe TTC	6	8	17	Tyr TAC	3	2	13
*Leu TTA ^e	7	16	4	— TAA	1	2	7
Leu TTG	8	11	3	— TAG	1	1	1
Leu CTT	10	18	4	His CAT	9	9	4
Leu CTC	2	5	3	His CAC	4	3	9
Leu CTA	5	4	0	*Gln CAA	11	17	8
Leu CTG ^d	2	10	67	*Gln CAG	3	9	26
Ile ATT	39	53	16	Asn AAT	21	33	4
Ile ATC	8	7	37	Asn AAC	10	8	32
Ile ATA	2	2	1	Lys AAA	36	57	77
Met ATG	18	11	24	Lys AAG	9	8	28
Val GTT	19	20	47	Asp GAT	28	45	14
Val GTC	6	3	8	Asp GAC	4	3	31
Val GTA	9	16	42	Glu GAA	31	40	58
Val GTG	2	4	17	Glu GAG	7	5	15
Ser TCT	9	18	24	Cys TGT	4	5	1
Ser TCC	2	5	21	Cys TGC	3	4	6
Ser TCA	6	13	1	— TGA	0	1	1
Ser TCG	1	2	0	Trp TGG	8	8	4
Pro CCT	8	6	4	Arg CGT	15	13	46
Pro CCC	0	1	1	Arg CGC	4	3	24
Pro CCA	5	11	6	Arg CGA	2	4	0
Pro CCG	2	3	32	Arg CGG	1	1	1
Thr ACT	19	21	31	Ser AGT	6	7	4
Thr ACC	2	4	21	Ser AGC	1	2	7
*Thr ACA	8	11	3	*Arg AGA	2	2	1
Thr ACG	1	2	2	*Arg AGG	2	0	0
Ala GCT	13	37	75	Gly GGT	28	27	44
Ala GCC	1	5	13	Gly GGC	10	3	35
Ala GCA	10	22	42	Gly GGA	3	7	1
Ala GCG	6	3	27	Gly GGG	0	2	0

a. Codons for gene 39 include 520 amino acid residues and two stop codons.

b. T4 replication genes include: gene 45 protein, 227 amino acid residues and 2 stop codons; gene 44 protein, 319 amino acid residues and 1 stop codon; and gene 62 protein, 187 amino acid residues and 1 stop codon. The numbers represent a total of 733 amino acids and 4 stop codons. Data are from Spicer and Konigsberg (27).

c. R-protein genes represent a composite of *E. coli* ribosomal genes which have a total of 1108 amino acid residues and 9 stop codons. Data are from Post and Nomura (28).

d. Codons recognized by abundant *E. coli* tRNAs are underlined.

e. *, codons recognized by T4 tRNAs.

abundant host tRNA species which are not preferential used by T4 gene 39; CUG (Leu), CCG (Pro), CAG (Gln) and GGC (Gly) are in this category. Table II also shows that some codons, that are recognized by T4 tRNAs (marked by *), are not preferentially used to a significant extent in the translation of gene 39 mRNA. This is likely due to the fact that 39-protein is one of the proteins synthesized immediately after phage infection (an immediate early protein) and is synthesized before or at about the same time T4 tRNA's are synthesized and processed (30). T4 phage DNA is rich in A and T, and gene 39 also shows an overall composition of 62.9% A + T, and 72.6% of the codons terminate in A or T. In general, the codon usage in gene 39 is consistent with that seen in other sequenced T4 replication genes similarly analyzed (27).

Expression of gene 39. T4 39-protein is synthesized immediately after

Table III
Possible Promoters for Gene 39

	location ^a	-35 region	space	-10 region	weight ^b
(1)	373	TTACTA	ATATTGAAATCACTGGA	TATAAT	57.4%
(2)	358	TTGTCT	TCAAGTGGTTTTAC	TAATAT	51.5%
(3)	358	TTGTCT	TCAAGTGGTTTTACTAATA	TTGAAA	47.9%
<i>E. coli</i> promoter consensus		TTGACA	(17±1 bp)	TATAAT	

a. The numbering of nucleotides is based on that of Fig 2. The designated location is that of the first nucleotide in the -35 region.

b. The value of the overall weight includes the homology scores among the -35 and -10 region hexamers and the spacing between them correlated to their occurrence among 112 known *E. coli* promoters and is based on the the algorithm of Mulligan et. al. (20).

phage infection, and it has been suggested that one of the early T4 transcripts starts near the beginning of gene 39 (31). Since T4 early promoters are anticipated to be recognized by host *E. coli* RNA polymerase, the DNA sequence determined here was examined for the presence of *E. coli* promoter-like sequences using a computer program designed to predict *in vitro* RNA polymerase selectivity (21). Table III summarizes the findings of such a search and lists the three most probable promoters. Starting at position 373 (Fig. 2), a hexamer TTACTA which is 17 bases 5' to a TATAAT (Pribnow box) is found. The region correlates with an overall weight of 57.4% relative to known *E. coli* promoters. For comparison, pBR322 *tet* and *bla* promoters are calculated to have 63.9% and 52.7%, respectively (21). Two other less favorable promoter-like sequences are also found near position 358. They share the same -35 region, but have different spacer lengths and -10 signals. Direct analysis of transcription will be needed to determine if any of these possible promoters are used. Although it is possible that the active gene 39 promoter lies outside of the sequenced area, one line of evidence suggested that it is within the analyzed R1-fragment. A pBR322 derivative was constructed by inserting the 2.8 kb T4 fragment into the EcoRI site with the 39 gene oriented in the same direction as the *tet* gene of the plasmid. Under these conditions, the 39-protein was synthesized both *in vivo* and *in vitro* in a coupled transcription and translation experiment (data not shown). In the *in vitro* experiment, purified *E. coli* RNA polymerase and an S30 fraction of an *E. coli* extract (32) were used. It has been shown that the main plasmid

transcription in this vicinity that is oriented in the direction of the tet gene originates from the P2 promoter which is downstream of the EcoRI site (33). Thus, it is likely that the observed expression of the T4 39 gene is derived from a promoter(s) in the inserted fragment. It should be noted that the observed expression in the uninfected, plasmid-containing cell was relatively low, and it is increased approximately 4-fold upon infection (data not shown). This suggests that although the 39 gene can be transcribed by unmodified E. coli RNA polymerase, its expression can be enhanced by phage coded factors.

Sequence information characteristic of rho-independent termination signals (34) was not found in the 406 bases sequenced 3' to the coding region. It has been proposed that there is a rho-dependent terminator within gene 39 (31). The sequence characteristics of rho-dependent terminators are not clearly defined, so the presence of such a signal remains to be experimentally demonstrated.

An inspection of the sequence also reveals the presence of the sequence AGGA at position 590. It is six bases 5' to the gene 39 initiation codon. This Shine-Dalgarno sequence (36) should adequately serve as the ribosome binding site for efficient translation of the gene according to the rules described by Stormo et al. (37). In fact, if a strong promoter such as the P_{lac}^{UV5} or lambda P_L is placed 5' to the EcoRI fragment harboring gene 39, a protein of the expected size for 39-protein is synthesized at very high levels (5-10% total cellular proteins) (3, 7).

Homology between T4 39-protein and other topoisomerases. When the amino acid sequence of T4 39-protein as derived from its DNA sequence was compared with other topoisomerases, I found that it shares significant homology with the gyrB subunit of B. subtilis DNA gyrase (35). Figure 3 shows an amino acid alignment between the two proteins. Although the B. subtilis protein is considerably larger than the T4 protein, homology starts from the N-terminus and extends for about 544 residues to encompass the 520 amino acids of the T4 protein. There is no homology between the C-terminal 94 amino acids of this gyrase subunit and T4 39-protein. The overall homology is 34.6% of identical amino acids at corresponding positions. Although T4 DNA topoisomerase and bacterial DNA gyrase are both ATP-dependent type II enzymes, there are characteristic differences between their activities. The protein alignment described here further confirms the structural and functional equivalence of the two protein subunits.

suggests that the protein has clusters of hydrophobic residues. Specifically, the most highly hydrophobic region of 39-protein spans amino acid residues 201 to 218. The hydropathy average of these 19 residues was found to be 1.25, which is above the mean hydropathy average for a number of soluble globular proteins whose hydrophobic sequences are known from X-ray crystallographic analysis (41). Thus, it is possible that the unusual behavior of the 39-protein subunit of the T4 DNA topoisomerase may be due to the presence of this and other hydrophobic regions long enough to bind to the bacterial membrane.

Since we have found that the isolated 39-protein subunit contains an ATPase activity (7), the possible presence of an adenine nucleotide binding sequence was therefore investigated. Based on the alignment of homologies in the amino acid sequences of proteins which are involved in ATP utilization and nucleotide binding, a conserved region common to many (but not all) adenine nucleotide binding proteins, called an "adenine nucleotide binding fold" has been described by Walker et al. (43). *E. coli* RecA protein (43) and T4 gene 44 protein (27), both of which are recognized to have ATPase activities, contain amino acid sequences consistent with the consensus for this ATP utilization domain. An inspection of the protein sequence of gene 39 showed that the amino acid sequence from residue 125 to 142, Lys Glu Arg Val Thr Gly Gly Met Asn Gly Val Gly Ser Ser Leu Thr Asn Ile, agrees with this postulated "ATP binding fold" (conserved amino acids are underlined). The significance of this prediction remains to be experimentally tested. It suggests that the ATP binding site of 39-protein may be located in the N-terminal one-third of the molecule.

According to the Chou and Fasman secondary structure prediction, 39-protein has a relatively low helical content. However, the presence of a long potentially helical region in the C-terminal region is noted. It starts at amino acid 365 and extends to amino acid 414. This region is unusually rich in alanine and lysine. An extreme example of this is the presence of 12 alanines and 9 lysines among the 32 amino acids from 370 to 401. These values correspond to an abundance of 37.5% alanine and 28% lysine. The overall abundance of alanine and lysine for the entire 39-protein is 5.8% and 8.7%, respectively. The high concentration of alanine in this region probably increases its propensity to be in an α -helical conformation (46). Furthermore, if the lysines present in this region are projected onto a helical wheel with a

100°C rotation for each of the amino acid side chains as described by Schiffer and Edmundson (46), it is seen that nearly all of these lysine side chains would protrude from one side of a potential α -helix. This configuration, if substantiated, would provide a positively charged surface which might interact with negatively charged DNA. It is known that T4 39-protein alone does bind both single and double stranded DNA (7, 44). It is also noted that the histone-like small DNA-binding proteins of E. coli (type II DNA binding proteins), such as the HU proteins and the H-protein have unusually high levels of alanine and lysine (45). Thus, it is tempting to postulate that this region near the C-terminal portion of the 39-protein is responsible for its DNA-binding property. It is curious that when the T4 39-protein was subjected to a homology search with protein data-bases (Dayhoff and NIH data-bases), it was found that limited but statistically nonrandom local homology exists between T4 39-protein and E. coli DNA polymerase I. Specifically, in the vicinity of the proposed helical region in T4 39-protein, 20 of the 67 residues from amino acid 372 to 438 are homologous to a 65 residue region of E. coli DNA polymerase I from amino acids 559 to 623. This extended stretch of sequence is about 30% homologous between the two proteins, although several single residue gaps were included in the amino acid sequence comparison. This alignment is shown here.

T4 39-protein

372 LARKLAA EKAAETKAAKKASKAKVHKHIKANLCGKDADTTFLFTEGDSAIGYLLIDVRDKELHGGGYPL 438
: : : : : : : : : : : : : : : :
559 LTLRLAELEKKHAHEIAGEEFNLST KQLQTLFEKGQIKPLKKTTPGG APSTSEEVLE ELALDYPL 623
E. coli DNA polymerase I

The corresponding region of the E. coli DNA polymerase I molecule represents a flexible domain whose coordinates were not ordered with certainty in the recently published X-ray structure of the large fragment of the enzyme, the Klenow fragment (47). Nonetheless, it was speculated that this domain constitutes a flexible area in the DNA binding cleft which allows a dynamic interaction between the DNA substrate and the polymerase (46). The homology between the alanine-lysine rich region on the T4 39-protein and the flexible sub-domain of the Klenow fragment of E. coli DNA polymerase I binding area encourages our assignment of the region on 39-protein as a potential non-sequence specific DNA-binding domain.

The above analysis suggests that 39-protein may be organized into

functional domains, an ATP binding fold in the N-terminal region and a DNA binding domain in the C-terminal region. These two functional domains are separated by a hydrophobic region. This observation is supported by the fact that discrete protein fragments are generated when purified 39-protein is subjected to limited proteolysis. Further experiments are in progress to test the functional partitioning of the T4 39-protein predicted by the DNA sequence analysis.

ACKNOWLEDGEMENTS

I thank Drs. S. R. Casjens for critical reading of the manuscript, T. Mattson for making available to us a plasmid which harbors the T4 gene 39, and J. Maizel for helping with some of the computer analysis of the DNA sequence. Ms. J. Cohenour's help in the preparation of the manuscript is greatly appreciated. This work was supported by National Institutes of Health Grant GM-21960.

REFERENCES

1. Stetler, G., King, G. and Huang, W.M. (1979) *Proc. Natl. Acad. Sci. USA*, 76, 3737-3742.
2. Liu, L., Liu, C. and Alberts, B. (1979) *Nature*, 281, 465-461.
3. Huang, W.M., Wei, L. and Casjens, S. (1985) *J. Biol. Chem.*, 260, 8973-8977.
4. McCarthy, D., Minner, C., Bernstein, H. and Bernstein, C. (1976) *J. Mol. Biol.*, 106, 963-981.
5. Kreuzer, K. and Huang, W. (1983) in *Bacteriophage T4*, edited by Mathews, K., Kutter, E. Mosig, G. and Berget, P., A.S.M. Publications, p. 90-96.
6. Huang, W.M. (1979) *Cold Spring Harbor Symp. Quant. Biol.*, 48, 495-499.
7. Nicholson, G. and Huang, W.M. (manuscript in preparation).
8. Wang, J. C. (1985) *Ann. Rev. Biochem.*, 54, 665-697.
9. Messing, J., Crea, R. and Seeburg, P. (1981) *Nucleic Acids Res.*, 9, 309-321.
10. Selzer, G., Bolle, A., Krisch, H. and Epstein, R. (1978) *Molec. gen. Genet.*, 159, 301-309.
11. Messing, J. and Vieira, J. (1982) *Gene*, 19, 269-276.
12. Murray, N.E., Bruce, S.A. and Murray, K. (1979) *J. Mol. Biol.*, 132, 493-504.
13. Weiss, B. and Richardson, C. (1967) *Proc. Natl. Acad. Sci. USA*, 57, 1021-1028.
14. Maxam, A.M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA*, 74, 560-564.
15. Sanger, F., Nicklen, S. and Coulson, A. (1977) *Proc. Natl. Acad. Sci. USA*, 74, 5463-5467.
16. Biggin, M., Gibson, T. and Hong, G. (1983) *Proc. Natl. Acad. Sci. USA*, 80, 3963-3965.
17. Remant, E., Stanssens, P. and Fiers, W. (1981) *Gene*, 15, 81-93.
18. O'Farrell, P., Gold, L. and Huang, W.M. (1973) *J. Biol. Chem.*, 248, 5499-5501.

19. Brutlag, D.L., Clayton, J., Friedland, P. and Kedes, L.H. (1982) *Nucleic Acids Res.*, 10, 279-294.
20. Mulligan, M., Hawley, D., Enriken, R. and McClure, W. (1984) *Nucleic Acids Res.*, 12, 789-800.
21. Wilbur, W. and Lipman, D. (1983) *Proc. Natl. Acad. Sci. USA*, 80, 726-730.
22. Niggemann, F., Green, I., Meyer, H. and Ruger, W. (1981) *Molec. gen. Genet.*, 184, 289-299.
23. Snyder, L., Gold, L. and Kutter, E. (1976) *Proc. Natl. Acad. Sci. USA*, 73, 3098-3102.
24. Mattson, T., Houwe, G., Bolle, A., Selzer, G. and Epstein, R. (1977) *Molec. gen. Genet.*, 154, 319-326.
25. Fuller, F. (1982) *Gene*, 19, 43-54.
26. Sutcliffe, J.G. (1978) *Cold Spring Harbor Symp. Quant. Biol.*, 43, 77-90.
27. Spicer, E.K. and Konigsberg, W.H. (1983) in *Bacteriophage T4*, edited by Mathews, C., Kutter, E. Mosig, G. and Berget, P. A.S.M. Publication, p. 291-301.
28. Post, L.E. and Nomura, M. (1980) *J. Biol. Chem.*, 255, 4660-4666.
29. Ikemura, T. (1981) *J. Mol. Biol.*, 146, 1-21.
30. Fukada, K. and Abelson, J. (1980) *J. Mol. Biol.*, 139, 377-391.
31. Brody, E., Rabussay, D. and Hall, D. (1983) in *Bacteriophage T4*, edited by Mathews, K., Kutter, E., Mosig, M. and Berget, P. A.S.M. Publications, p. 174-183.
32. Zubay, G. (1973) *Ann. Rev. Genet.*, 7, 267-287.
33. Huber, D. and Bujard, H. (1981) *Proc. Natl. Acad. Sci. USA*, 78, 167-171.
34. McClure, W.R. (1985) *Ann. Rev. Biochem.*, 54, 171-204.
35. Moriya, S., Ogasawara, N. and Yoshikawa, H. (1985) *Nucleic Acid Res.*, 13, 2251-2265.
36. Shine, J. and Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA*, 71, 1342-1346.
37. Stormo, G., Schneider, T.D. and Gold, L. (1982) *Nucleic Acids Res.*, 10, 2971-2996.
38. Chou, P.Y. and Fasman, G.D. (1978) *Ann. Rev. Biochem.*, 47, 251-276.
39. Huang, W.M. (1975) *Virology*, 66, 508-521.
40. Huang, W.M. (1979) *J. Virology*, 32, 917-924.
41. Kyte, J. and Doolittle, R.F. (1982) *J. Mol. Biol.*, 157, 105-132.
42. Liu, L., Liu, C. and Alberts, B. (1980) *Cell*, 19, 697-707.
43. Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982) *The EMBO J.*, 1, 945-951.
44. Huang, W.M. and Buchanan, J.M. (1974) *Proc. Natl. Acad. Sci. USA*, 71, 2226-2230.
45. Hubscher, U., Lutz, H. and Kornberg, A. (1980) *Proc. Natl. Acad. Sci. USA*, 77, 5097-5101.
46. Schiffer, M. and Edmundson, A. (1967) *Biophysical J.*, 7, 121-135.
47. Ollis, D., Brick, P., Hamlin, R., Xuong, N. and Steitz, T. (1985) *Nature*, 313, 762-766.