Proc. Natl. Acad. Sci. USA
Vol. 77, No. 5, pp. 2611–2615, May 1980
Biochemistry

Sequences of the *recA* gene and protein

(DNA nucleotide sequence/tryptic peptides/NH_2 and COOH termini/cysteines/"pseudo insertion sequence")

AZIZ SANCAR*, C. STACHELEK†, W. KONIGSBERG†, AND W. DEAN RUPP*†

Departments of *Therapeutic Radiology and †Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, Connecticut 06510

Communicated by Evelyn M. Witkin, February 22, 1980

ABSTRACT **We have determined the nucleotide sequence of the *recA* gene of *Escherichia coli*; this permits the formulation of the primary structure for the *recA* protein. This structure is consistent with the amino acid composition of the tryptic peptides obtained from the *recA* protein. The coding region of the *recA* gene has 1059 base pairs, which specify 352 amino acids. The *recA* protein has alanine and phenylalanine as its NH_2- and COOH-terminal amino acids, respectively, and has the following amino acid composition: Cys_3 Asp_{20} Asn_{15} Met_9 Thr_{17} Ser_{20} Glu_{30} Gln_{13} Pro_{10} Gly_{35} Ala_{38} Val_{22} Ile_{27} Leu_{31} Tyr_7 Phe_{10} His_2 Lys_{27} Trp_2 Arg_{14}. Of the three cysteine residues, only two can be alkylated under reducing and denaturing conditions. The molecular weight of the *recA* polypeptide is 37,842.**

The *recA* protein of *Escherichia coli*, which is involved in important cellular functions such as cell division, recombination–repair, mutagenesis, and phage induction (1), has been shown recently to catalyze several reactions *in vitro*: (*i*) proteolytic cleavage of λ repressor in the presence of ATP (2); (*ii*) DNA-dependent hydrolysis of ATP (3, 4); (*iii*) hybridization of single-stranded DNA (5); (*iv*) uptake of single-stranded DNA by homologous double-stranded DNA (6, 7); (*v*) and ATP- and oligonucleotide-dependent unwinding of double-stranded DNA (8). It is also required for the cutting of superhelical double-stranded DNA in the presence of homologous crosslinked DNA and ATP (9). Although the last three reactions are interrelated and might reflect different aspects of a single phenomenon, it is clear that *recA* protein must have at least three active sites for its protease, ATPase, and helicase–recombinase functions. To identify these active sites it is essential to know the primary structure of the protein.

In a previous communication (10), we published a restriction endonuclease map of the *recA* gene and localized the *recA* promoter. Here we have determined the DNA sequence of the *recA* gene, which allowed us to deduce the primary structure of the *recA* protein. The postulated sequence was confirmed in part by the amino acid compositions of peptides isolated from a tryptic digest of the *recA* protein and by NH_2- and COOH-terminal analyses of the intact protein. Our results show that the coding region of the *recA* gene is 1059 nucleotides long and that *recA* protein is made up of 352 amino acids. The calculated molecular weight is 37,842.

MATERIALS AND METHODS

Enzymes and Biochemicals. Restriction endonucleases and polynucleotide kinase were obtained from New England BioLabs; deoxynucleoside triphosphates and dideoxynucleoside triphosphates were purchased from P-L Biochemicals. DNA polymerase I and bovine pancreatic DNase I were from Boehringer Mannheim and bacterial alkaline phosphatase was from Worthington. Reagents used in Maxam–Gilbert reactions were obtained from the sources described by these authors (11). [^{35}S]Methionine (1000 Ci/mmol) and [^{32}P]ATP (3000 Ci/mmol) were purchased from New England Nuclear (1 Ci = 3.7 × 10^{10} becquerels).

Purification of *recA* Gene and Protein. The *recA* plasmids used in this study as a source of restriction fragments for DNA sequence determination were pDR1453, pDR1461, and pDR1464. The construction of these plasmids and the methods used for purification of plasmid DNA and restriction fragments have been described (10). pDR1453 complements *recA* mutations (10), and this plasmid was used as the source of *recA* protein in the studies of Shibata *et al.* (6). [^{35}S]Methionine-labeled *recA* protein was prepared as follows: *E. coli* KM4104/pDR1453 (10) was grown in 500 ml of K medium (12), to an $OD_{600}^{1\,cm}$ of 0.5. Cells were collected by centrifugation, washed, and resuspended in 100 ml of low-sulfate Hershey medium (13); nalidixic acid was added to give a final concentration of 50 μg/ml; and the culture was incubated at 37°C for 30 min. [^{35}S]Methionine was added (1 μCi/ml) to the culture and incubation was continued for another 2 hr. Cells were collected, washed, and resuspended in 5 ml of 0.05 M Tris·HCl, pH 7.6, containing 10% sucrose. The suspension was frozen at −80°C for 10 min. After thawing, lysozyme and NaCl were added to final concentrations of 0.1 mg/ml and 10 mM, respectively. The cell suspension was kept on ice for 1 hr and then centrifuged at 12,000 rpm for 20 min in a Sorvall SS-34 rotor. The supernate was collected and *recA* protein was extracted from it by Polymin-P (Bethesda Research Laboratories, Rockville, MD) precipitation and elution with 1 M NaCl as described by Roberts *et al.* (2). The final *recA* protein preparation was at least 90% pure as judged by sodium dodecyl sulfate/polyacrylamide gel electrophoresis and had a specific activity of about 10^6 cpm/mg of protein.

Unlabeled *recA* protein was prepared by growing cells in a fermentor to an $OD_{600}^{1\,cm}$ of 6. This was followed by nalidixic acid induction, freeze-thaw lysis, Polymin-P precipitation, and extraction with 1 M NaCl as described above. The enzyme was further purified by hydroxyapatite chromatography (6). The yield for *recA* protein prepared this way was about 100 mg of protein per liter of culture.

Determination of DNA Sequence. This procedure was done by the Maxam–Gilbert chemical method (11), employing the A+C, A+G, G, T+C, and C reactions. The nucleotide sequence of about half of the gene was also determined by the dideoxy method (14, 15). Polyacrylamide gel electrophoresis was carried out according to Sanger and Coulson (16).

Determination of NH_2- and COOH-Terminal Residue of *recA* Protein. The NH_2-terminal sequence was determined by the microdansyl/Edman technique (17). Carboxypeptidases A and B were used to liberate amino acids from the COOH terminus of the protein, using the conditions of Ambler (18).

Analysis of Tryptic Fragments of the *recA* Protein. [^{35}S]-Methionine-labeled *recA* protein was oxidized with performic acid by the method of Hirs (19). The oxidized protein was dialyzed against 0.05 M NH_4HCO_3, pH 8.4. A 5 mg/ml solution of trypsin (Worthington, treated with tosylphenylalanine

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

Abbreviation: IS, insertion sequence.

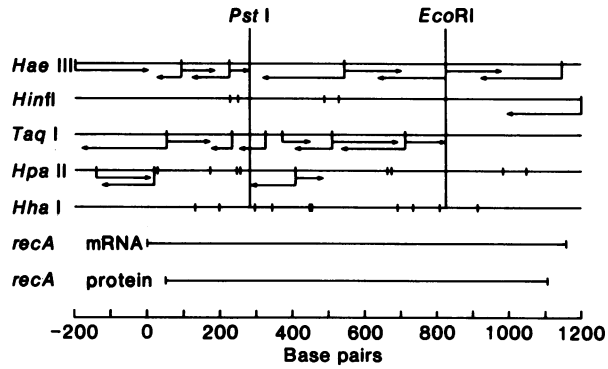

FIG. 1. Restriction map and sequencing strategy for *recA* gene. The arrows indicate the direction of sequencing and the length of sequences determined in individual experiments.

chloromethyl ketone) was prepared in 0.01 M HCl and was added to the *recA* solution to give a trypsin/*recA* protein ratio of 1/100 (wt/wt). The tryptic peptides were isolated by elution from Aminex AG-50W-X4 (Bio-Rad) with a gradient of pyridinium acetate. Radioactive pools were chromatographed on Dowex AG-1X2 (Bio-Rad), while nonradioactive peptide pools were purified by descending chromatography on Whatman 3MM paper in a 1-butanol/pyridine/acetic acid/H_2O (15:10:3:12, vol/vol) solvent.

The amino acid compositions of the *recA* protein and its tryptic fragments were determined, after acid hydrolysis, on a Beckman 121 M amino acid analyzer.

Alkylation of Sulfhydryl Groups. *recA* protein was alkylated with iodo[^{14}C]acetamide in 8.5 M guanidine·HCl/0.1 M Tris·HCl, pH 8.0, under reducing and nonreducing conditions. Samples were taken from the reaction mixture at various times and the reaction was terminated by the addition of 5 μl of 2-mercaptoethanol. The samples were dialyzed against potassium phosphate buffer (50 mM, pH 6.8). An equal volume of 12 M HCl was added to each and hydrolysis was carried out for 22 hr at 110°C. Portions of each hydrolysate were analyzed on a

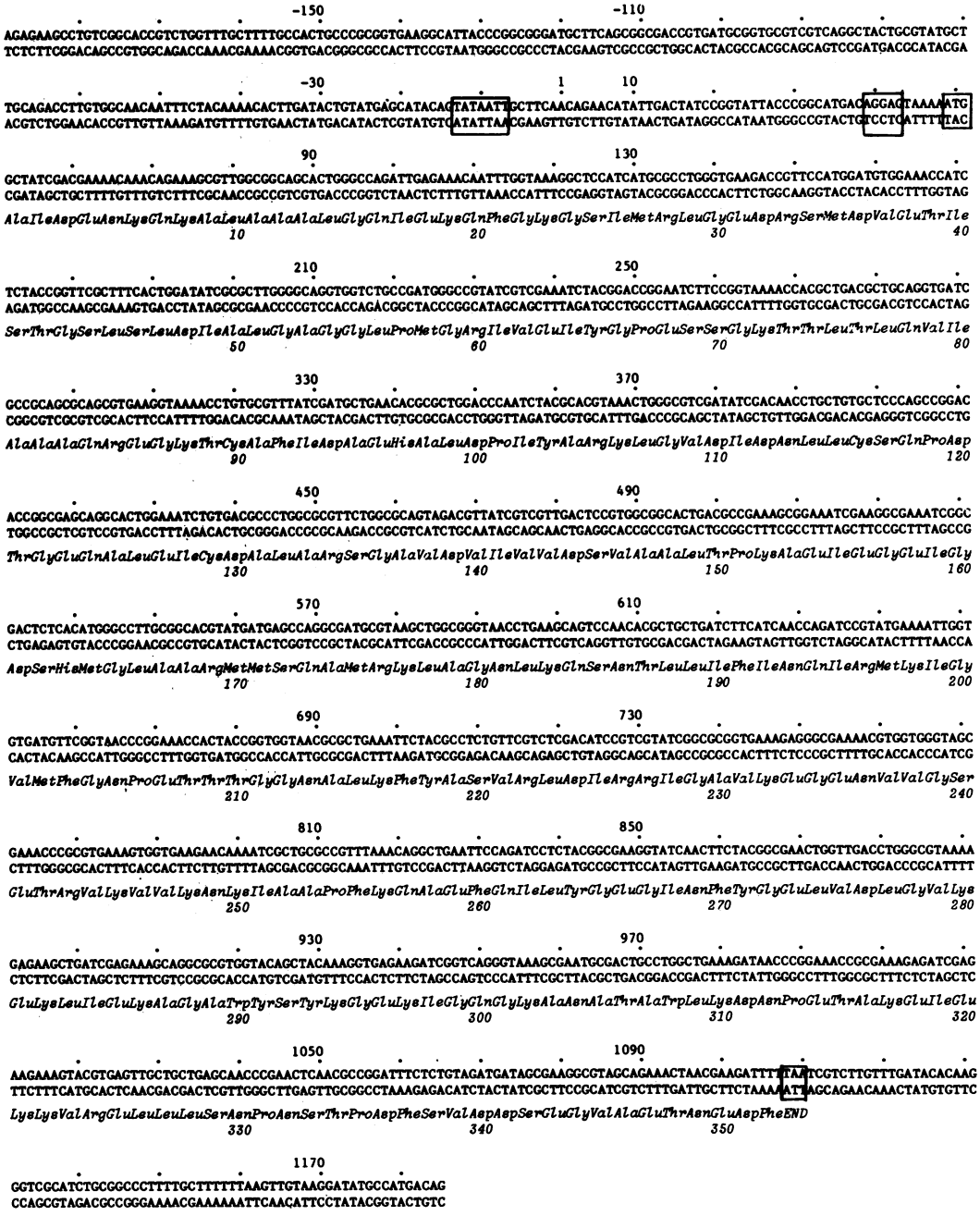

FIG. 2. Sequences of the *recA* gene and protein. The base pair designated +1 corresponds to the 5′ terminus of the *recA* transcript. The Pribnow and Shine-Dalgarno sequences as well as the codons used for initiation and termination of translation are boxed.

Table 1. Amino acid compositions of *recA* protein and isolated tryptic peptides

Amino acid	*recA* protein		Tryptic peptides, residue numbers									
	Analysis*	DNA sequence†	9–19‡	24–28	34–60	61–72	89–105	107–134	106–134	153–169	170–176	178–183
$Cys(O_3H)$	3.6(4)	3					0.9(1)	1.9(2)	1.8(2)			
Asx	35.1(35)	35			2.0(2)		2.1(2)	5.0(5)	4.9(5)	1.0(1)		1.2(1)
$Met(O_2)$	9.5(10)	9		1.0(1)	2.0(2)					1.1(1)	2.6(3)	
Thr	16.5(16)	17			1.7(2)		1.1(1)	1.0(1)	1.2(1)			
Ser	18.7(19)	20		1.0(1)	3.7(4)	1.8(2)		1.1(1)	1.0(1)	0.9(1)	1.1(1)	
Glx	44.2(44)	43	2.4(2)		1.1(1)	2.4(2)	1.2(1)	4.3(4)	4.3(4)	3.1(3)	1.2(1)	
Pro	13.5(14)	10			0.9(1)	1.0(1)	0.9(1)	1.0(1)	1.1(1)			
Gly	34.1(34)	35	1.2(1)	0.8(1)	5.1(5)	2.2(2)		2.1(2)	2.2(2)	3.2(3)		1.0(1)
Ala	37.3(37)	38	3.1(3)		2.0(2)		3.9(4)	3.0(3)	3.0(3)	2.7(3)	1.1(1)	1.1(1)
Val	22.6(23)	22			0.9(1)	0.8(1)		1.3(1)	1.3(1)			
Ile	24.3(24)	27	1.1(1)	1.2(1)	1.9(2)	1.7(2)	1.9(2)	1.8(2)	1.9(2)	2.0(2)		
Leu	32.4(32)	31	2.0(2)		4.0(4)		1.1(1)	4.5(5)	4.5(5)	1.0(1)		1.7(2)
Tyr	7.3(7)	7				0.8(1)	0.9(1)					
Phe	9.4(9)	10					0.9(1)					
His	3.3(3)	2					1.0(1)			1.0(1)		
Lys	24.0(24)	27	0.6(1)			1.1(1)			0.8(1)			1.0(1)
Trp	ND	2										
Arg	14.8(15)	14		1.0(1)	1.0(1)		0.9(1)	0.8(1)	1.0(1)	1.0(1)	1.0(1)	
Total	350	352	10	5	26	12	17	28	29	17	7	6

	Tryptic peptides, residue numbers												
	197–198	199–216	217–222	233–243	246–248	249–250	257–280	295–297	298–302	311–317	323–324	325–352§	Total¶
$Cys(O_3H)$													3
Asx		2.0(2)		1.4(1)		1.0(1)	2.3(2)			1.9(2)		6.7(7)	26
$Met(O_2)$	0.9(1)	0.7(1)											9
Thr		2.6(3)		1.0(1)						1.1(1)		1.9(2)	11
Ser			1.1(1)	1.0(1)								3.1(3)	15
Glx		1.4(1)		3.0(3)			5.4(5)	1.1(1)	1.1(1)	1.2(1)		3.9(4)	30
Pro		0.9(1)								0.8(1)		2.2(2)	8
Gly		3.2(3)		2.1(2)			4.3(4)	0.9(1)	1.9(2)			1.6(2)	29
Ala		1.4(1)	1.2(1)				1.1(1)			1.2(1)		1.6(2)	23
Val		1.0(1)	1.2(1)	1.6(2)	1.8(2)		1.9(2)				1.0(1)	2.3(2)	14
Ile		0.9(1)					2.0(2)		1.0(1)				16
Leu		1.2(1)					2.4(2)					3.3(3)	21
Tyr			0.6(1)				1.7(2)						5
Phe		0.6(1)	0.8(1)				1.8(2)					1.2(1)	6
His													2
Lys	1.0(1)	1.0(1)			1.2(1)	1.0(1)	0.9(1)	1.0(1)	1.0(1)	0.9(1)			12
Trp													
Arg			1.1(1)	0.8(1)							1.0(1)		9
Total	2	17	6	11	3	2	24	3	5	7	2	28	240

Data are numbers of amino acids, with the nearest whole numbers in parentheses. ND, not determined.
* Composition determined by amino acid analysis.
† Composition determined from DNA sequence.
‡ Peptide has been sprayed with ninhydrin for detection.
§ Digestion with carboxypeptidase A revealed Phe as the COOH-terminal residue.
¶ Residues 107–134 are not included in the total because they are a subset of residues 106–134.

Beckman 121 M amino acid analyzer. The peaks corresponding to carboxymethylcysteine were collected and their radioactivities were determined.

RESULTS

Nucleotide Sequence of the *recA* Gene. Restriction fragments of *recA* labeled at 5′ ends were either cut with a second restriction enzyme or denatured to obtain singly end-labeled fragments whose sequences were determined either by chemical methods (11) or by both chemical and enzymatic methods (15). Fig. 1 shows the restriction fragments that were used and gives the strategy followed to obtain the nucleotide sequence of the *recA* gene. Over 70% of the sequence was determined from both strands of the DNA.

The nucleotide sequence of the *recA* gene is shown in Fig. 2. The bases are numbered starting with the first base of *recA* mRNA which was determined from analysis of the *recA* transcript made *in vitro*.‡ Transcription starts six base pairs "downstream" from a putative Pribnow box (20), T-A-T-A-A-T-T, which at the 5′ side overlaps with a 22-base-pair dyad symmetry. At +1155 to +1162 the sequence T-T-T-T-T-T-A-A is found; it is preceded by a 28-base-pair hyphenated molecular palindrome. Because this is a "typical" termination signal (21) we infer that *recA* mRNA probably terminates in this region. Translation starts at base pair +49 (see below), five nucleotides downstream from a Shine–Dalgarno ribosome binding sequence (22), A-G-G-A-G. The structural part of the gene is 1059

‡ A. Sancar and W. D. Rupp, unpublished.

nucleotides long and translation is terminated with a Phe codon at base pair +1107, which is followed by the ochre codon TAA.

Amino Acid Analysis of *recA* Protein. To check the correspondence between the amino acid sequence of *recA* protein predicted from the nucleotide sequence and the actual protein product, we determined the NH_2- and COOH-terminal amino acids as well as the amino acid composition of the intact *recA* protein and many of its tryptic fragments.

The NH_2-terminal analysis gave NH_2-Ala-Ile-, indicating that the Met residue specified by the initiation codon is cleaved off *in vivo*. Carboxypeptidase A and B digestion of intact *recA* protein released one residue of Phe and a fractional residue of Asp, a result consistent with the sequence -Asp-Phe-COOH as predicted from the DNA sequence.

The amino acid compositions of the *recA* protein and its tryptic fragments gave the results summarized in Table 1. It is apparent from a comparison of the data in Table 1 with the results shown in Fig. 2 that the composition of the tryptic peptides predicted from the DNA sequence are in good agreement with the analytical data. The isolated tryptic peptides account for about 70% of the protein. The other tryptic fragments were found as mixtures but in most cases their compositions could be matched with mixtures of peptides predicted from the *recA* gene sequence. Further confirmation of the nucleotide sequence was provided by the amino acid composition of the intact *recA* protein which agrees well with that predicted from the DNA sequence.

Reactivity of Sulfhydryl Groups. The *recA* protein contains three cysteine residues, at positions 90, 116, and 129. To find out if their SH groups were free or present as disulfides, the protein was alkylated with iodo[^{14}C]acetamide under reducing and nonreducing conditions. In both instances only two of the three cysteines were alkylated (Table 2). Although these results eliminated the possibility of intramolecular disulfide bridges, they did not explain the reason for the unreactive cysteine residue nor rule out the chance of intermolecular disulfide links between *recA* protein subunits. We therefore electrophoresed the protein in sodium dodecyl sulfate/acrylamide gels under reducing and nonreducing conditions. The *recA* protein had identical mobilities under both conditions (data not shown), indicating that there was no intersubunit crosslinking. The possibility exists, however, that a prosthetic group is attached to one of the cysteine residues.

Table 2. Reactivity of sulfhydryl groups in the *recA* protein*

Time, hr	mol of *S*-[^{14}C]carboxymethylcysteine per mol of *recA* protein after alkylation		
	Native protein	Denatured protein†	
		Nonreduced	Reduced‡
0.25	0.09		
0.50	0.15		
1.0	0.19	1.0	1.5
1.5	0.25		
2.0		1.7	
3.0		1.2	
4.0		1.9	
5.0		2.2	
6.0		2.0	
24.0	0.40	2.0	1.6

* Reaction conditions: 0.23 mM *recA* protein, 7.1 mM iodo[^{14}C]-acetamide (except for the reduced sample, for which 60 mM was used), 1 mM EDTA, 50 mM Tris·HCl, pH 7.5.
† Protein was denatured in 8.5 M guanidine hydrochloride.
‡ Protein was reduced for 4 hr at 50°C with a 100-fold molar excess of 2-mercaptoethanol over SH groups.

DISCUSSION

The sequence of the *recA* gene was determined as a first step towards studying its regulation and function at a molecular level. The knowledge of both the DNA sequence and the primary structure of the protein has allowed us to make several observations that will be useful in future experiments.

Regulation. In noninduced cells *recA* protein is made only in small quantities, whereas in induced cells the gene is as actively transcribed as a ribosomal RNA gene (cf. refs. 23–26). Our empirical analysis of the *recA* promoter indicates that it is among the "strongest" *E. coli* promoters.‡ Therefore, the low level of *recA* expression under noninducing conditions must be due to a repressor effect. Mount and coworkers (27) have accumulated extensive data indicating that the *lexA* gene product is the repressor. Because all known repressor–operator interactions involve DNA dyad symmetries, we searched for such sequences in the *recA* gene. These are shown in Table 3. In addition to the 22-base-pair symmetrical sequence in the promoter region (−32 to −11) there are a variety of molecular palindromes in the structural part of the gene. Although the probability of chance occurrence of these individual sequences in a 1000-nucleotide-long DNA sequence is high (28), their clustering in the first third of the gene is striking. It is conceivable that these sequences are involved in transcriptional and translational regulation of *recA* expression.

In addition to these dyad symmetries, *recA* contains three translational symmetries (direct repeats) in the DNA that are worth noting. The 16-nucleotide sequence at +162 to +169 and +173 to +180 also occurs at +669 to +684. The sequence at the latter site is flanked by an 8-nucleotide direct repeat, reminiscent of direct repeats surrounding insertion sequences (ISs) (see

Table 3. Dyad symmetries in the *recA* gene

Location	Sequence
−130 to −122	CCCGGCGGG
	GGGCCGCCC
−73 to −66	TGCA TGCA
	ACGT ACGT
−32 to −11*	ATACTGTATGA GCATACAGTAT
	TATGACATACT CGTATGTCATA
+21 to +29	GGTATTACC
	CCATAATGG
+91 to +99	CTGGGCCAG
	GACCCGGTC
+151 to +158	TCCA TGGA
	AGG ACCT
+245 to +264†	ACCGGAAT CTTCCGGT
	TGGCCTTA GAAGGCCA
+351 to +374	CCCAATCTACGC ACGTAAACTGGG
	GGGTTAGATGCG TGCATTTGACCC
+376 to +387	GTCGAT ATCGAC
	CAGCTA TAGCTG
+813 to +820	GTTT AAAC
	CAAA TTTG
+1031 to	TGCTGAGCA
+1039	ACGACTCGT
+1126 to	CACAAGGGTCGCAT CTGCGGCCCTTTTG
+1153‡	GTGTTCCCAGCGTA GACGCCGGGAAAAC

Underlining indicates dyad symmetry.
* Overlaps with Pribnow box.
† Overlaps with a "pseudo-IS."
‡ Terminator.

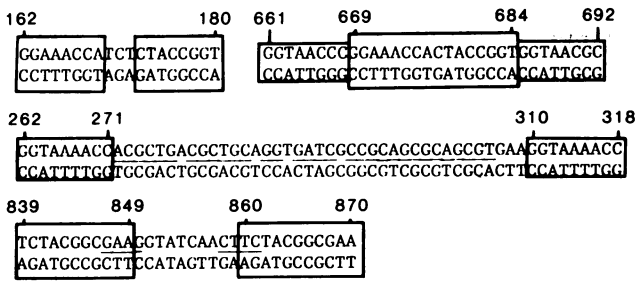

FIG. 3. Pseudo-ISs in the *recA* gene. Direct repeats are boxed and bases in dyad symmetry are underlined.

ref. 24 for a review). Even more striking are the sequences at +262 to +318 and +839 to +870 (Fig. 3). These sequences have inverted repeats at the ends, which are flanked by direct repeats, features possessed by known ISs (29). Because they are too small to code for proteins that promote their transposition, we propose the name "pseudo-ISs" for these and similar sequences. Whether these pseudo-ISs are transposed or whether they play a role in the regulation of *recA* gene expression awaits experimental testing. It is tempting to speculate, however, that these sequences have played a role in the evolution of the *recA* gene and perhaps even of the *E. coli* chromosome itself. The validity of this idea can be put to test when more sequence results become available, allowing detailed comparison of various genes and intergenic regions.

Mechanism of Action. Among the most intensively studied DNA-binding proteins, *lac* repressor (30) and λ repressor (31) have well-defined domains—i.e., an NH_2-terminal region that binds to DNA and a COOH-terminal region that is involved in protein–protein interaction. Similar domains with separate functions have been proposed for the *O* protein of λ phage (32) and the gene 32 protein specified by the T4 bacteriophage (33). Because there are a number of *recA* mutants having various phenotypes (1), it is likely that *recA* has structural domains also. Although the protein sequence itself gives no clear indication of such domains, we suggest that the NH_2-terminal half is involved in protein–protein interactions and the COOH-terminal region participates in DNA binding, on the basis of the following considerations: Chou–Fasman analysis (34) of the protein sequence suggests the presence of an α-helical region at the NH_2 terminus separated from another α-helical segment near the COOH terminus by a middle region containing predominantly a β sheet. The COOH-terminal half of the protein contains about 75% of the basic and 80% of the aromatic amino acids, which are likely to be involved in ionic and stacking interactions with the DNA. The NH_2-terminal half contains all three cysteines and both histidines, amino acids that are part of the active sites of most proteases (35). In addition the sequences Asp-Ile-Ala-Leu, Ala-Glu-His-Ala, and Gly-Asp-Ser present in this part of the *recA* protein are homologous to active site sequences of several serine proteases (35). Clearly, the unambiguous identification of various functional regions should follow from sequence analysis of mutant *recA* genes that give rise to different phenotypic properties of the *recA* protein.

After this work was completed we learned that Emmerson *et al.* (36) had obtained *recA* protein from the plasmid pDR1453 and had determined the NH_2-terminal sequence, which is in agreement with the sequence reported here. Horii *et al.* (37) have independently determined the sequence of the *recA* gene. Their sequence is in agreement with ours with the exception of the base pair at position +1134, which is missing in their sequence.

We are grateful to A. J. Clark for his comments on the manuscript. A.S. is thankful to R. P. Wharton for his careful recording of the DNA sequence. This work was supported by U.S. Public Health Service Grants GM12607 and CA06519.

1. Witkin, E. M. (1976) *Bacteriol. Rev.* **40**, 869–907.
2. Roberts, J. W., Roberts, C. W. & Craig, N. L. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4714–4718.
3. Ogawa, T., Wabiko, H., Tsurimoto, T., Horii, T., Masukata, H. & Ogawa, H. (1979) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 909–915.
4. Roberts, J. W., Roberts, C. W., Craig, N. L. & Phizicky, E. (1979) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 917–920.
5. Weinstock, G. M., McEntee, K. & Lehman, I. R. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 126–130.
6. Shibata, T., DasGupta, C., Cunningham, R. P. & Radding, C. M. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1638–1642.
7. McEntee, K., Weinstock, G. M. & Lehman, I. R. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 2615–2619.
8. Cunningham, R. P., Shibata, T., DasGupta, C. & Radding, C. M. (1979) *Nature (London)* **281**, 191–195.
9. Cassuto, E., Mursalim, J. & Howard-Flanders, P. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 620–624.
10. Sancar, A. & Rupp, W. D. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3144–3148.
11. Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
12. Rupp, W. D., Wilde, C. E., III, Reno, D. L. & Howard-Flanders, P. (1971) *J. Mol. Biol.* **61**, 25–44.
13. Worcel, A. & Burgi, E. (1974) *J. Mol. Biol.* **82**, 91–105.
14. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
15. Maat, J. & Smith, J. H. (1978) *Nucleic Acids Res.* **5**, 4537–4545.
16. Sanger, F. & Coulson, A. R. (1978) *FEBS Lett.* **87**, 107–110.
17. Weiner, A. M., Platt, T. & Weber, K. (1972) *J. Biol. Chem.* **247**, 3242–3251.
18. Ambler, R. P. (1972) *Methods Enzymol.* **25**, 262–271.
19. Hirs, C. H. W. (1956) *J. Biol. Chem.* **219**, 611–621.
20. Pribnow, D. (1975) *J. Mol. Bol.* **99**, 419–443.
21. Gilbert, W. (1976) in *RNA Polymerase*, eds. Losick, R. & Chamberlin, M. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 221–243.
22. Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 1342–1346.
23. Pettijohn, D. E., Clarkson, K., Kossman, C. R. & Stonington, O. G. (1970) *J. Mol. Biol.* **52**, 281–300.
24. Kiss, A., Sain, B. & Venetianer, P. (1977) *FEBS Lett.* **79**, 77–79.
25. Gudas, L. J. & Pardee, A. B. (1976) *J. Mol. Biol.* **101**, 459–477.
26. McPartland, A., Green, L. & Echols, H. (1978) in *DNA Repair Mechanisms*, eds. Hanawalt, P. C., Friedberg, E. C. & Fox, C. F. (Academic, New York), pp. 383–386.
27. Pacelli, L. Z., Edmiston, S. H. & Mount, D. W. (1979) *J. Bacteriol.* **137**, 568–573.
28. Dykes, G., Bambara, R., Marians, K. & Wu, R. (1975) *Nucleic Acids Res.* **2**, 327–345.
29. Reed, R. R., Young, R. R., Steitz, J. A., Grindley, N. D. F. & Guyer, M. S. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4882–4886.
30. Ogata, R. T. & Gilbert, W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5851–5854.
31. Pabo, C. D., Sauer, R. T., Sturtevant, J. M. & Ptashne, M. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1608–1612.
32. Scherer, G. (1978) *Nucleic Acids Res.* **5**, 3141–3156.
33. Williams, K. R. & Konigsberg, W. (1978) *J. Biol. Chem.* **253**, 2463–2470.
34. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 222–244.
35. Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, DC), Vol. 5, pp. 53–66.
36. Emmerson, P. T., Northrop, F. D., Walker, J. E. & West, S. C. (1979) *FEBS Lett.* **106**, 349–351.
37. Horii, H., Ogawa, T. & Ogawa, H. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 313–317.