AMINO TERMINAL SEQUENCE OF THE *recA* PROTEIN OF *ESCHERICHIA COLI*

P. T. EMMERSON, F. D. NORTHROP+, J. E. WALKER+ and S. C. WEST*

Department of Biochemistry, University of Newcastle upon Tyne, Newcastle upon Tyne, +MRC Laboratory of Molecular Biology, Hills Road, Cambridge, England and **Department of Therapeutic Radiobiology, Yale University, New Haven, CN, USA*

Received 2 August 1979
Revised version received 28 August 1979

1. Introduction

The *recA* gene of *E. coli* is involved in genetic recombination [1], DNA repair [1,2], induction of prophage λ [3,4], mutagenicity [5] and coordination of cell division with DNA synthesis [6,7]. The product of this gene is a protein of ~40 000 mol. wt [8] which is induced copiously when DNA is damaged or when DNA synthesis is inhibited [9–11]. The *recA* gene is thought to be regulated negatively by a repressor coded by *lexA* and induction of the gene appears to require positive action by the *recA* product itself [9–11].

The *recA* protein is known to inactivate the phage repressor by cleaving it [12]. However, it is also known to bind to single-stranded DNA [13] and to catalyse ATP-dependent renaturation of DNA [14]. In addition, it is involved in endonucleolytic cleavage of intact superhelical DNA triggered by damage in homologous duplex DNA (cutting in trans) [15], and homologous pairing of single-stranded DNA to superhelical DNA [16].

A better understanding of the complex multifunctional roles of this crucial protein in recombination and repair may be helped by sequence analysis of the wild-type gene and various mutant genes. With this in mind, we have determined the amino-terminal amino acid sequence of the wild-type as an aid to DNA base sequence analysis of the gene.

2. Materials and methods

2.1. *Purification of the* recA *protein*

The *recA* protein was prepared by a modification of the method in [14]. *E. coli* K12 strain KM4104 carrying the *recA* Δ7 deletion on the chromosome and the *recA*$^+$ gene on the plasmid pDR1453 [17] was used for preparation of the protein.

Cells (8 l) were grown at 37°C in Luria broth to A_{650} = 0.5, treated with 40 μg/ml nalidixic acid for 90 min and then harvested. The cells were lysed and acidic proteins precipitated by the addition of polymin P as in [14]. The *recA* protein was extracted from the pellet and recovered by ammonium sulphate precipitation. The precipitate was washed with phosphate buffer (20 mM sodium phosphate (pH 6.8), 10% glycerol, 10 mM β-mercaptoethanol) containing 0.28 g/ml ammonium sulphate and the final pellet resuspended in 4 ml phosphate buffer.

After overnight dialysis against the same buffer the proteins were loaded onto an equilibrated 10 × 1.6 cm hydroxyapatite column (Biorad Biogel-HPT) and eluted with a 400 ml gradient of 20–400 mM phosphate (pH 6.8). *recA* protein eluted at ~200 mM phosphate. Fractions were pooled and precipitated by the addition of 0.3 g/ml $(NH_4)_2SO_4$. Following centrifugation, the pellet was resuspended in 2 ml TEGD buffer (20 mM Tris–HCl (pH 7.5), 10% glycerol, 0.1 mM EDTA, 50 mM NaCl, 1 mM dithiothreitol) and applied to a 5.0 × 1.0 cm column of equilibrated DEAE-cellulose (Whatman DE52). Proteins were eluted with a 50–400 mM NaCl gradient in TEGD buffer and the *recA* protein fractions were pooled and precipitated with $(NH_4)_2SO_4$. The pellet was dialysed extensively against 20 mM Tris–HCl (pH 7.0), 20% glycerol, 0.1 mM EDTA, 50 mM NaCl, 2 mM dithiothreitol, and the *recA* protein was stored at −70°C at 18 mg/ml. From 8 l culture ~60 mg

Table 1
N-terminal sequence of *recA* protein

Residue number	Identification method			Quantity (nM)	Identity
	GC	HPLC	BH		
1	+	+	+	15.81	Ala
2	+	+	+	17.54	Ile
3	–	+	+	13.77	Asp
4	–	+	+	13.69	Glu
5	–	+	+	17.42	Asn
6	–	+	+	7.31	Lys
7	–	+	+	10.77	Gln
8	–	+	+	5.06	Lys
9	+	+	+	7.58	Ala
10	+	+	+	7.28	Leu
11	+	+	+	4.97	Ala
12	+	+	+	7.46	Ala
13	+	+	+	8.00	Ala
14		+	+	7.34	Leu
15		+	+	5.51	Gly
16		+	+	3.42	Gln
17		+	+	6.98	Ile
18		+	+	2.33	Glu
19		–	+	1.82	Lys
20		+	+	1.16	Gln
21		+	+	2.27	Phe
22		+	+	2.45	Gly
23		+	+	0.84	Lys
24		+	+	1.95	Gly
25		–	–	not identified	
26		+	+	4.11	Ile
27		–	+	1.00	Met
28		–	+	1.08	Arg
29		+	+	1.53	Leu
30		+	+	1.30	Gly
31		+	+	0.53	Glu

The degradation was performed once with 1 mg (25 nM) of protein. Released phenylthiohydantoin amino acids were identified by gas chromatography (GC), high-pressure liquid chromatography (HPLC) and amino acid analysis following back hydrolysis (BH). The yields of sequence amino acids are determined by amino acid analysis and are corrected for hydrolytic losses [19]

recA protein was obtained. The protein was >95% pure as judged by SDS–polyacrylamide gel electrophoresis.

2.2. *Sequence analysis*

The N-terminal sequence of the protein was determined with the aid of an updated Beckman 890B spinning cup sequencer. The protein (1 mg, 25 nmol) was degraded in the presence of polybrene using a dilute Quadrol single cleavage programme [18]. Phenylthiohydantoin amino acids were identified by high-pressure liquid chromatography or by amino acid analysis. These methods are detailed in [19].

3. Results

The sequence of 31 amino acids from the N-terminus of the *recA* protein is shown in table 1. No amino acid could be identified with certainty at position 25. The yield of alanine released during the first cycle is consistent with the amount of protein used in the experiment, assuming protein mol. wt 40 000. This confirms the purity of the protein sample.

The amino acid composition of the *recA* protein is shown in table 2.

4. Discussion

A restriction map of the *recA* gene has been determined and sequence analysis is in progress [17]. The N-terminal sequence (table 1) should permit easy iden-

Table 2
Amino acid composition of *recA* protein

Amino acid	% Composition[a]
Aspartic acid	7.4
Threonine	4.6
Serine	10.5
Glutamic acid	13.6
Proline	3.7
Glycine	14.8
Alanine	10.2
Valine	5.5
Methionine	1.8
Isoleucine	5.2
Leucine	7.4
Tyrosine	1.8
Phenylalanine	2.5
Histidine	1.8
Lysine	5.5
Arginine	3.4

[a] The compositions of cysteine and tryptophan were not determined

Details of the determination of amino acid composition are given in [19]

tification or confirmation of the initiation codon and the reading frame of the *recA* mRNA. Knowledge of the approximate molecular weight (40 000) should also permit identification of the termination codon.

The amino acid composition (table 2) could be useful in fingerprint analysis of peptide fragments of *recA* protein isolated from wild-type and mutant cells.

Acknowledgements

We are very grateful to Drs A. Sancar and W. D. Rupp for generously providing the plasmid pDR1453. This work was supported by the Medical Research Council.

References

[1] Clark, A. J. and Margulies, A. D. (1965) Proc. Natl. Acad. Sci. USA 53, 451–459.
[2] Howard-Flanders, P. and Theriot, L. (1966) Genetics 53, 1137–1150.
[3] Hertman, I. and Luria, S. E. (1967) J. Mol. Biol. 23, 117–123.
[4] Devoret, R. and Blanco, M. (1970) Mol. Gen. Genet. 107, 272–280.
[5] Witkin, E. M. (1969) Mut. Res. 8, 8–14.
[6] Green, M. H. L., Greenberg, J. and Donch, J. (1969) Genet. Res. Camb. 14, 150–162.
[7] Inouye, M. (1971) J. Bacteriol. 106, 539–542.
[8] McEntee, K., Hesse, J. E. and Epstein, W. (1976) Proc. Natl. Acad. Sci. USA 73, 3979–3983.
[9] Emmerson, P. T. and West, S. W. (1977) Mol. Gen. Genet. 155, 77–85.
[10] McEntee, K. (1977) Proc. Natl. Acad. Sci. USA 74, 5275–5279.
[11] Gudas, L. J. and Mount, D. W. (1977) Proc. Natl. Acad. Sci. USA 74, 5280–5284.
[12] Roberts, J. W., Roberts, C. W. and Craig, N. L. (1978) Proc. Natl. Acad. Sci. USA 75, 4714–4718.
[13] Gudas, L. J. and Pardee, A. B. (1975) Proc. Natl. Acad. Sci. USA 72, 2330–2334.
[14] Weinstock, G. M., McEntee, K. and Lehman, I. R. (1979) Proc. Natl. Acad. Sci. USA 76, 126–130.
[15] Cassuto, E., Mursalim, J. and Howard-Flanders, P. (1978) Proc. Natl. Acad. Sci. USA 75, 620–624.
[16] Shibata, T., Das Gupta, C., Cunningham, R. P. and Radding, C. M. (1979) Proc. Natl. Acad. Sci. USA 76, 1638–1642.
[17] Sancar, A. and Rupp, W. D. (1979) Proc. Nat. Acad. Sci. USA in press.
[18] Brauer, A. W., Margolies, M. N. and Haber, E. (1975) Biochemistry 14, 3029.
[19] Walker, J. E., Carne, A. F., Runswick, M. J., Bridgen, J. and Harris, J. I. (1979) Eur. J. Biochem. in press.
[20] Walker, J. E., Shaw, D. C., Northrop, F. D. and Horsnell, T. (1977) in: Solid Phase Methods in Protein Sequence Analysis (Previero, A. and Previero, M. A. eds) p. 277, Elsevier/North-Holland, Amsterdam, New York.