In the preceding paper[9] Calos presents the sequence of the promoter for the *lacI* gene. I have extended the sequence through the *lacI* gene to the *lac* promotor/operator control region, which has been sequenced by Gilbert and Maxam (ref. 10 and personal communication) and Dickson *et al.*[11]. The repressor protein sequence of Beyreuther *et al.*[12,13], a sequence for the initial portion of the mRNA reported by Steege[14], and the DNA sequence of the terminal 223 base pairs of *lacI* constructed by A. Maxam with G. Copenhaver and W. Herr (personal communication) were known. The known protein sequence was used to position DNA sequences within the gene; the DNA sequence agrees with the protein sequence except for four emendations.

Experimental design

The plasmid[9] which provided the DNA for this analysis consists of two DNA fragments covering the *lacI* region carried on the pMB9 vector[15]. The *I* gene DNA was isolated by digesting the plasmid with *Hin*dII + III (ref. 16) and separating the products on a 5% polyacrylamide (1:29 bisacrylamide) gel. The fragments containing *I* gene information are 935 and 789 base pairs long and correspond to the N and C-terminal portions of the protein respectively. Fragments resulting from cleavage with a second restriction endonuclease (either *Hae*III, *Hpa*II, *Alu*I or *Mbo*III (ref. 16)) were then labelled: treatment with bacterial alkaline phosphatase removes the 5′ terminal phosphate, and T4 polynucleotide kinase catalyses the transfer of the radioactive phosphate of [γ-^{32}P]ATP to the 5′ termini of DNA to generate fragments with two 5′ labels[1]. The Maxim–Gibert technique requires singly-labelled DNA molecules; cleavage with a second restriction endonuclease or strand separation yielded two such molecules.

The sequencing technique[1] involves the modification of the nucleotide bases of these end-labelled fragments in a limited and random fashion using a purine-specific (dimethylsulphate, DMS) or a pyrimidine-specific (hydrazine) reagent. The

Revised sequence for the *lac* repressor

THE elucidation of the DNA sequence of the *lacI* gene[1] has indicated the existence of an additional 13 amino acids not reported in the original protein sequence[2]. Recently a direct protein chemical verification of the suggested additional 11 amino acids in the repressor after position 147 and of 2 amino acids after residue 230 (formerly residue 219[2]) has been made. The strategy of the re-examination of the protein sequence included the isolation of [^{14}C]carboxyamidomethylated cysteine-containing peptides which contain the sought-after amino acids and the automated Edman degradation[3] of these peptides. Thus, a quantitative detection and sequence determination was possible of all fragments including insoluble peptides since there was the suspicion that solubility problems and semi-quantitative manual sequencing methods[2] gave rise to the encountered discrepancies.

Repressor was cleaved at the two tryptophanyl residues 201 and 220 by BNPS-skatole[4] and one of the fragments thus obtained was further digested with trypsin and subsequently with peptidyl-L-glutamate hydrolase (an endopeptidase cleaving specifically peptide bonds at glutamic acid residues)[5]. The results summarised in Fig. 1 are in complete agreement with the amino acid sequence predicted from the DNA sequence of the *lacI* gene[1].

KONRAD BEYREUTHER

Institut für Genetik der Universität zu Köln, Köln, FRG

Received 27 February; accepted 26 June 1978

1. Farabaugh, P. J. *Nature* **274**, 765–769 (1978).
2. Beyreuther, K., Adler, K., Geisler, N. & Klemm, A. *Proc. natn. Acad. Sci. U.S.A.* **70**, 3576–3580 (1973); Beyreuther, K., Adler, K., Fanning, E., Murray, C., Klemm, A. & Geisler, N. *Eur. J. Biochem.* **59**, 491–509 (1975).
3. Edman, P. & Begg, G. *Eur. J. Biochem.* **1**, 80–91 (1967).
4. Omenn, G. S., Fontana, A. & Anfinsen, C. B. *J. biol. Chem.* **245**, 1895–1902 (1970).
5. Houmard, J. & Drapeau, G. R. *Proc. natn. Acad. Sci. U.S.A.* **69**, 3506–3509 (1972).
6. Capra, J. D. in *Solid Phase Methods in Protein Sequence Analysis* (eds Previero, A. & Coletti-Previero, M.-A.) (North-Holland, Amsterdam, in the press).

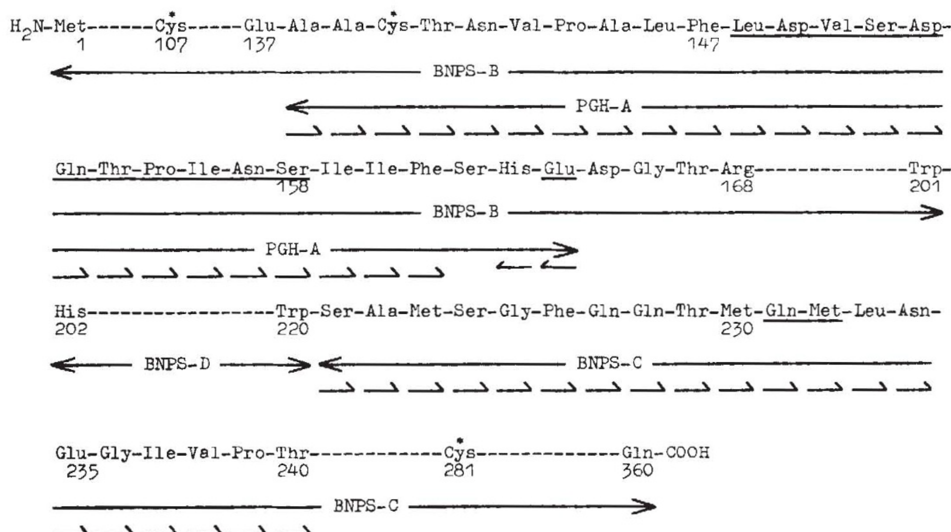

Fig. 1 Revised amino acid sequence of *lac* repressor of *E. coli*. Peptides are designated as BNPS when derived from cleavage with BNPS-skatole[4] or as PGH when derived from digestion with peptidyl-L-glutamate hydrolase[5]. The BNPS-fragments were separated on BioGel P150 in 6 M guanidine hydrochloride. BNPS-fragment B was treated with trypsin, the largest tryptic peptide (residues 119–168) was isolated by chromatography on Sephadex G-50 in 0.05 M NH_4HCO_3 and cleaved with peptidyl-L-glutamate hydrolase. The PGH-peptides were isolated from the precipitate of the digest by chromatography on Sephadex G-25 in 2 M acetic acid. A second radioactive PGH-peptide (PGH-B, residues 138–168) gave the same sequence 148–158 as shown for PGH-A. Residue positions determined by automated Edman degradation[3] (→) or by hydrolysis with carboxypeptidase C[2] (←) are indicated by arrows. The asterisk indicates [^{14}C]carboxyamidomethylated cysteine. Automated sequence analyses were performed by using a Beckman updated model 890B Sequencer and a Quadrol (0.1 M) program[2]. Polybrene (Aldrich Chemical Co.) (0.7 mg per run) was used as carrier[6]. Identification of PTH-amino acids was as described[2]. The residues not reported in the original sequence[2] are underlined. Sequences 1–147, 159–230 (formerly residues 148–219) and 233–360 (formerly residues 220–347) with the exception of residue 164 are as reported previously[2]. The 360 residues of *lac* repressor have a calculated molecular weight of 38,590.

© 1978 Nature Publishing Group