The Amino-Acid Sequence of lac Repressor
(E. coli/regulation/protein-DNA interaction)

KONRAD BEGREUTHER, KLAUS ADLER, NORBERT GEISLER, AND ALEX KLEMM*

Institut für Genetik der Universität zu Köln, 5 Köln 41, Germany (FRG)

Communicated by Jacques Monod, July 20, 1973

ABSTRACT The amino-acid sequence of lac repressor from Escherichia coli has been determined. The sequence contains 347 residues in the subunit single peptide chain. It shows no similarities with the sequences of histones or the known part of β-galactosidase.

If we want to understand repression (1), we have to elucidate the three-dimensional structure of a repressor-operator complex. The lac system of E. coli offers the possibility of such an analysis. Lac repressor has been isolated (2). It binds in vitro to operator DNA (3, 4) and represses in vitro transcription (5) and translation (6) of the lac operon. It can be isolated in large amounts (7, 8). Genetic analysis of mutants of the lac repressor-producing i gene has shown that the amino-terminus of repressor is involved in operator binding (9). The lac operator has also been analyzed genetically (10) and chemically: Gilbert succeeded recently in determining its structure (11). We report here the sequence of lac repressor. The sequence of its amino-terminus has been presented earlier (9, 12-14).

RESULTS

The amino-acid composition of lac repressor is given in Table 1. The amino-acid sequence is shown in Fig. 1. It was determined by analysis of fragments derived by treating the protein with cyanogen bromide; for the ordering of the fragments, tryptic peptides of repressor were used.

Lac repressor was purified from t° and t+ strains (7, 13) as described earlier (8). Gel electrophoresis in the presence of sodium dodecyl sulfate and endgroup determinations of repressor had indicated the presence of a single polypeptide chain with a molecular weight of 38,000-40,000. On the basis of the amino-acid composition, a molecular weight of 38,000 was calculated (8). The amino-acid sequence gives a molecular weight of 37,200 for the subunit and 148,800 for the functional tetramer.

Cyanogen Bromide Fragments. 10 Fragments were obtained after lac repressor was cleaved with cyanogen bromide (Table 2). The cyanogen bromide digest was chromatographed on a Sephadex G-75 column equilibrated with 1 M acetic acid (Fig. 2). The first peak, in the break-through volume, contained partially cleaved repressor and was not further analyzed. Peak II consists of an "overlap peptide" which proved to be fragment V joined to fragment III. Peak III contained the largest fragment with the two tryptophan residues of the chain. Several recylcings were necessary to separate it from peak IV. Fragments V and VI could be easily separated from the fraction of the small peptides. They correspond to sequences 2-42 (FVI) and 43-98 (FV) (9). Cyanogen bromide fragments VII-XII (including free homoserine) could be separated on DEAE-cellulose with pyridine-acetate buffers at pH 6.5. Gel electrophoresis of the isolated fragments II-VI in Na dodecyl sulfate and endgroup analysis established their molecular weight and their homogeneity. The molecular weights of fragments VII-XII could be determined with the method of Offord (16). The sum of the molecular weights of the ten fragments and the sum of the compositions of the ten fragments agreed with the molecular weight and the amino-acid composition of intact lac repressor (Tables 1 and 2).

Ordering the Cyanogen Bromide Fragments. Tryptic peptides were used to order the cyanogen bromide peptides. Thirty tryptic peptides were isolated after chromatography on high-resolution ion-exchange resins, gel chromatography, thin-layer electrophoresis, and thin-layer chromatography of a tryptic digest of carboxymethylated lac repressor. Five of these peptides contained methionine. The sequence of these peptides (Table 3) provided the overlaps that allowed the cyanogen bromide fragments to be ordered. The order of the cyanogen bromide fragments is: FXII-FVI-FV-FIX-FVII-XII-XI-FXI-FXII. The sequence of the other 25 tryptic peptides has also been determined.

Sequence Analysis of the Cyanogen Bromide Fragments. The structure of cyanogen bromide fragments VII-XII was determined without further enzymatic digestions. Cyanogen bromide fragments II-VI were digested with various proteolytic enzymes (Fig. 3A-D). The peptide mixtures thus obtained were fractionated by the methods described for separation of

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>No. of residues</th>
<th>Amino acid</th>
<th>No. of residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lysine</td>
<td>11</td>
<td>Glycine</td>
<td>22</td>
</tr>
<tr>
<td>Histidine</td>
<td>7</td>
<td>Alanine</td>
<td>44</td>
</tr>
<tr>
<td>Arginine</td>
<td>19</td>
<td>Cysteine</td>
<td>3</td>
</tr>
<tr>
<td>Aspartic Acid</td>
<td>15</td>
<td>Valine</td>
<td>33</td>
</tr>
<tr>
<td>Asparagine</td>
<td>11</td>
<td>Methionine</td>
<td>9</td>
</tr>
<tr>
<td>Threonine</td>
<td>18</td>
<td>Isoleucine</td>
<td>17</td>
</tr>
<tr>
<td>Serine</td>
<td>30</td>
<td>Leucine</td>
<td>40</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>13</td>
<td>Tyrosine</td>
<td>8</td>
</tr>
<tr>
<td>Glutamine</td>
<td>28</td>
<td>Phenylalanine</td>
<td>4</td>
</tr>
<tr>
<td>Proline</td>
<td>13</td>
<td>Tryptophan</td>
<td>2</td>
</tr>
</tbody>
</table>

Total no. of residues 347

Abbreviation: Hsr, homoserine.
the tryptic peptides. The sequences of the purified peptides were determined by the most appropriate current techniques:

- Met-Lys-Pro-Val-Thr-Leu-Tyr-Asp-Val-Ala-Glu-Tyr-Ala-Gly-Val-10
- Ser-Tyr-Gln-Thr-Val-Ser-Arg-Val-Ala-Asn-Gln-Ala-Ser-His-Val-20
- Ser-Ala-Lys-Thr-Arg-Glu-Lys-Glu-Val-Ala-Met-Ala-Glu-Leu-40
- Asn-Tyr-Ile-Pro-Asn-Arg-Val-Ala-Gln-Leu-Ala-Gly-Lys-Gln-50
- Ser-Leu-Leu-Ile-Gly-Val-Ala-Thr-Ser-Leu-Leu-Ala-Leu-His-Ala-70
- Pro-Ser-Gln-Ile-Ala-Ile-Lys-Ser-Arg-Asp-Ala-Gln-Leu-80
- Gly-Ala-Ser-Val-Val-Ser-Met-Glu-Arg-Ser-Gly-Val-Glu-100
- Ala-Cys-Lys-Ala-Ala-Val-His-Asn-Leu-Leu-Asn-Arg-Val-Ser-110
- Gly-Leu-Ile-Asn-Tyr-Pro-Leu-Asp-Arg-Asn-Ala-Ile-Ala-120
- Val-Glu-Ala-Ala-Cys-Thr-Asn-Val-Pro-Asp-Leu-Phe-Ile-Ile-140
- Ser-His-Asn-Gly-Ile-Arg-Glu-Leu-Gly-Glu-Ala-Val-Ala-150
- Leu-Gly-His-Ile-Gln-Ala-Leu-Gly-Leu-Ala-Val-Ala-160
- Ser-Leu-Ala-Arg-Leu-Ala-Gly-Leu-Leu-Thr-Arg-Val-Ala-170
- Glu-Leu-Ala-Ile-Ile-Ile-Asn-Arg-Ile-Val-Arg-Ile-180
- Val-Ser-Ala-Arg-Leu-Ala-Gly-Leu-Leu-Thr-Arg-Val-Ala-190
- Arg-Asn-Ile-Val-Pro-Ile-Arg-Asn-Arg-Asp-Trp-Leu-200
- Ala-Met-Asp-Gly-Ile-Gln-Thr-Leu-Asn-Glu-Glu-Val-210
- Pro-Thr-Ala-Met-Leu-Val-Ala-Val-Leu-Glu-Ala-220
- Met-Arg-Ile-Thr-Glu-Ser-Glu-Leu-Ala-Asp-Ile-230
- Ser-Leu-Gly-Tyr-Asp-Thr-Glu-Ser-Ser-Arg-Val-Leu-Gly-240
- Pro-Pro-Thr-Ile-Leu-Gly-Ala-Phe-Arg-Leu-Leu-Gly-250
- Thr-Ser-Val-Asp-Arg-Leu-Glu-Leu-Ser-Leu-Gly-Gln-Ala-260
- Lys-Gly-Asn-Leu-Leu-Leu-Val-Ser-Leu-Ala-270
- Thr-Leu-Pro-Asn-Thr-Thr-Thr-Val-Thr-Pro-Arg-Ala-Leu-280
- Gly-Gln-347

**Fig. 1.** The amino acid sequence of lac repressor from *E. coli.*

**Table 2.** Characterization of cyanogen bromide fragments of lac repressor

<table>
<thead>
<tr>
<th>Cyanogen bromide fragment</th>
<th>Molecular weight (SDS gel)</th>
<th>Electrophoretic mobility (± 5%)</th>
<th>No. of amino acids</th>
<th>NH&lt;sub&gt;T&lt;/sub&gt;-Terminus</th>
<th>COOH-Terminus</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>FI</td>
<td>30,000-35,000</td>
<td>—</td>
<td>—</td>
<td>Lys, Ala, Val</td>
<td>Hsr, Glu</td>
<td>—</td>
</tr>
<tr>
<td>FII</td>
<td>20,000</td>
<td>—</td>
<td>170</td>
<td>Ala</td>
<td>Hsr</td>
<td>43-212</td>
</tr>
<tr>
<td>FIIII</td>
<td>14,000</td>
<td>—</td>
<td>114</td>
<td>Val</td>
<td>Hsr</td>
<td>99-212</td>
</tr>
<tr>
<td>FIV</td>
<td>11,000</td>
<td>—</td>
<td>93</td>
<td>Arg</td>
<td>Hsr</td>
<td>242-334</td>
</tr>
<tr>
<td>FV</td>
<td>5,800</td>
<td>—</td>
<td>56</td>
<td>Ala</td>
<td>Hsr</td>
<td>43-98</td>
</tr>
<tr>
<td>FVI</td>
<td>4,500</td>
<td>1,500</td>
<td>13</td>
<td>Gln</td>
<td>Gln</td>
<td>335-347</td>
</tr>
<tr>
<td>FVII</td>
<td>—</td>
<td>1,100</td>
<td>10</td>
<td>Leu</td>
<td>Hsr</td>
<td>220-229</td>
</tr>
<tr>
<td>FIX</td>
<td>—</td>
<td>750</td>
<td>7</td>
<td>Leu</td>
<td>Hsr</td>
<td>230-236</td>
</tr>
<tr>
<td>FX</td>
<td>—</td>
<td>750</td>
<td>7</td>
<td>Ser</td>
<td>Hsr</td>
<td>213-219</td>
</tr>
<tr>
<td>FXI</td>
<td>—</td>
<td>450</td>
<td>5</td>
<td>Ala</td>
<td>Hsr</td>
<td>237-241</td>
</tr>
<tr>
<td>FXII</td>
<td>—</td>
<td>—</td>
<td>1</td>
<td>Hsr</td>
<td>—</td>
<td>1</td>
</tr>
<tr>
<td>Lac repressor</td>
<td>38,000-40,000</td>
<td>—</td>
<td>347</td>
<td>Met</td>
<td>Gln</td>
<td>1-347</td>
</tr>
</tbody>
</table>

SDS, sodium dodecyl sulfate; Hsr, homoserine.

**Fig. 2.** Separation of the cyanogen bromide fragments of lac repressor on a 2 × 200-cm column of Sephadex G-75 equilibrated with 1.0 M acetic acid. The digest of 7.5 μmol (300 mg) was dissolved in 10 ml of 2.0 M acetic acid; 2-ml portions were applied to the column at once and eluted with 1.0 M acetic acid. The column was developed at 22° at a rate of 30 ml/hr. The effluent was monitored both at 280 nm (broken line) and by ninhydrin analysis after alkaline hydrolysis of 10-μl aliquots (solid line).
The amino-acid sequence of cyanogen bromide fragments III (A), IV (B), V (C), and VI (D) and alignment of peptides after hydrolysis with trypsin (T), chymotrypsin (C), and thermolysin (TL).

FIG. 3.

DISCUSSION

There is no obvious feature in the sequence of lac repressor that could explain its tight and specific binding to lac operator. However, genetic analysis of the i-gene mutants has shown that the amino-terminus is involved in operator-DNA bind-

Sequence of Lac Repressor

1 Ala-Glu-Leu-Asn-Tyr-Ile-Pro-Arg-Val-Ala-Gln-Glu-Leu-Ala-Gly-Lys-Gln-Ser-Leu-Leu-Ile-Gly-Val-Ala-

25 TSC ➔ T6 ➔

< C5a ➔ C5b ➔ C5,6 ➔ C6,7 ➔ C7a

< TL5a ➔ TL5b ➔ TL6a ➔ TL6,7 ➔ TL7a ➔

26 Thr-Ser-Ser-Leu-Ala-Leu-His-Ala-Pro-Ser-Gln-Ile-Val-Ala-Ile-Lys-Arg-Ala-Asp-Glu-Leu-Gly-Ala-

40 T8 ➔ T9N

< TL7b ➔ TL7c ➔ TL7d ➔ TL7,8,9N ➔ TL9a

51 Ser-Val-Val-Ser-Met

56 C

1 Lys-Pro-Val-Thr-Leu-Tyr-Asp-Ala-Glu-Tyr-Ala-Gly-Val-Ser-Tyr-Gln-Thr-Ala-Glu-Ala-Lys-Thr-Arg-Glu-Lys-

25 Val-Ala-Asp-Gln-Leu-Gly-Ala-Ser-Ala-Ala-Asp-Arg-Gln-

< C1a ➔ C1b ➔ C1c ➔ C1,2 ➔

< TL1a ➔ TL1b ➔ TL1c ➔ TL1d ➔ TL1,2 ➔

26 Ala-Ser-His-Ala-Pro-Leu-His-Thr-Arg-Glu-Lys-Glu-Ala-Ala-

41 Met-

< C2a ➔ C2,3,4,5N ➔

< TL2a ➔ TL2,3,4,5N ➔ D

Fig. 3. (continued)

The sequence does not resemble the sequence of histones (20). It may be pointed out that 50% of the tyrosine residues are found between sequences 1 and 50, the region which according to genetic analysis is involved in operator binding. Furthermore the two tryptophans occupy positions (190 and 209) in a region which is probably involved in inducer binding (21).

We found considerable deamidation of Asn25, Gln54, Gln55, Gln131, and Gln153, but of no other glutamines or asparagines. Asp154 could possibly have been an asparagine since extensive deamidation is known to occur at asparagine residues followed by a glycine residue (22, 23). Asn25 has been invoked in operator binding (9), and a similar involvement of Gln54 and Gln55 seems possible. Pure repressor has the annoying property of being rather inactive with regard to operator binding (8). Deamidation of asparagines and glutamines that are involved in operator-DNA recognition could explain the loss of operator-binding activity.

Table 3. Sequence of tryptic peptides containing methionine

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Overlap for cyanogen bromide fragments</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>T 1</td>
<td>FXII-FVI</td>
<td>Met-Lys-Pro-Val-Thr-Leu-Tyr-Asp-Val-Ala-Glu-Tyr-Ala-Gly-Val-Ser-Tyr-Gln-Thr-Val-Ser-Arg</td>
</tr>
<tr>
<td>T 5</td>
<td>FVI-FV</td>
<td>Val-Glu-Ala-Ala-Met-Ala-Glu-Leu-Asn-Tyr-Ile-Pro-Asn-Arg</td>
</tr>
<tr>
<td>T 9</td>
<td>FV-FIII</td>
<td>Ala-Asp-Gln-Leu-Gly-Ala-Ser-Val-Val-Ser-Met-Val-Glu-Arg</td>
</tr>
<tr>
<td>T 18</td>
<td>FIII-FX-</td>
<td>Glu-Gly-Asp-Trp-Ser-Ala-Met-Ser-Gly-Phe-Gln-Gln-Thr-Met-Asp</td>
</tr>
<tr>
<td>T 28</td>
<td>FIV-FIX-</td>
<td>Asn-Glu-Gly-Ile-Val-Pro-Thr-Asp-Met-Val-Val-Ala-Asn-Asp-Gln</td>
</tr>
<tr>
<td>T 28</td>
<td>FIV-FIX-</td>
<td>Met-Ala-Leu-Gly-Ala-Met-Arg</td>
</tr>
</tbody>
</table>
Platt et al. (13) have shown that trypsin and chymotrypsin readily destroy operator binding of native lac repressor, leaving inducer binding intact. The trypsin-resistant core lacks 79 residues (14). 59 Residues are derived from the aminoterminus according to Platt et al. The present sequence shows that the remaining 20 residues originate from the carboxyterminus. In the same context, we would like to point out that our previous determination of the carboxy-terminal residue, was probably produced from the internal Lys-Arg-Lys (312–314) sequence by trypsin contaminating our chymotrypsin B.

Did β-galactosidase and lac repressor evolve from a common ancestor? We have shown that antibody against β-galactosidase does not react with lac repressor in an Ouchterlony test. Similarly there is no crossreaction between antibody against repressor and β-galactosidase. Since about 30% of the sequence of β-galactosidase is known (24), a comparison of the sequences cannot yet give the final answer. Visual inspection of the published 350 sequences of β-galactosidase did not show any homology with the sequence of lac repressor.

We thank Ellen Fanning for her help, Carolyn Murray, and Santosh Pinto for excellent assistance, B. Müller-Hill for giving us the opportunity to conduct this work and for his continual interest, and G. Braunitzer, B. Gutte, and B. S. Hartley for discussions. Part of the equipment was provided by Stiftung Volkswagenwerk. The work was supported by grants from Deutsche Forschungsgemeinschaft through the SFB 74 to K.B. and B.M.H.