Novel structure of the recA locus of Mycobacterium tuberculosis implies processing of the gene product.

E O Davis, S G Sedgwick and M J Colston

Novel Structure of the recA Locus of Mycobacterium tuberculosis Implies Processing of the Gene Product

ELAINE O. DAVIS, STEVEN G. SEDGWICK, and M. JOSEPH COLSTON

Laboratory of Leprosy and Mycobacterial Research, and Genetics Division, National Institute for Medical Research, Mill Hill, London NW7 1AA, Great Britain

Received 29 March 1991/Accepted 28 June 1991

A fragment of Mycobacterium tuberculosis DNA containing recA-like sequences was identified by hybridization with the Escherichia coli recA gene and cloned. Although no expression was detected from its own promoter in E. coli, expression from a vector promoter partially complemented E. coli recA mutants for recombination, DNA repair, and mutagenesis, but not for induction of phage λ. This clone produced a protein which cross-reacts with antisera raised against the E. coli RecA protein and was approximately the same size. However, the nucleotide sequence of the cloned fragment revealed the presence of an open reading frame for a protein about twice the size of other RecA proteins and the cloned product detected by Western blotting (immunoblotting). The predicted M. tuberculosis RecA protein sequence was homologous with RecA sequences from other bacteria, but this homology was not dispersed; rather it was localized to the first 254 and the last 96 amino acids, with the intervening 440 amino acids being unrelated. Furthermore, the junctions of homology were in register with the uninterrupted sequence of the E. coli RecA protein. Identical restriction fragments were found in the genomic DNAs of M. tuberculosis H37Rv and H37Ra and of M. bovis BCG. It is concluded that the ancestral recA gene of these species diversified via an insertional mutation of at least 1,320 bp of DNA. Possible processing mechanisms for synthesizing a normal-size RecA protein from this elongated sequence are discussed.

The gram-positive bacterium, Mycobacterium tuberculosis, is the causative agent of tuberculosis. This disease continues to be responsible for widespread illness, particularly in developing countries, with 3 million fatalities occurring annually (20). Vaccination with M. bovis BCG has proved effective in some areas but has provided little or no protection against the disease in others (19). The identification and cloning of particular antigens involved in the immune response to this intracellular parasite (for a review, see reference 61) has been pursued with a view to developing an improved subunit vaccine. This approach might be optimized by using recombination-deficient mutants in which introduced DNA fragments might be more stably maintained. In Escherichia coli, homologous recombination relies on recA activity. Hence, we have cloned the recA-like gene of M. tuberculosis as a preliminary step in the construction of a recA mutant strain.

M. tuberculosis bacilli reside in macrophages where they would be exposed to hydrogen peroxide as part of the macrophage defense mechanism (36). In E. coli, resistance to hydrogen peroxide correlates with the recA genotype of the cells, rather than with levels of the protecting enzymes catalase, peroxidase, or superoxide dismutase (9), although this appears not to be the case with Neisseria gonorrhoeae (22). SOS induction, requiring RecA activity, enhances the virulence of Erwinia carotovora by increasing the synthesis of pectin lyase and carotovorin (39). Therefore, it is possible that the recA gene of M. tuberculosis has a similar function which might affect virulence.

The RecA protein of E. coli has been shown to have an important role in general recombination, DNA repair, mutagenesis, and the expression of SOS functions (58). Analogous recA genes have been isolated from more than 20 gram-negative bacteria by complementation of E. coli recA mutants (for a review, see references 42 and 47), and an analogous gene has been identified in the gram-positive bacterium Bacillus subtilis (37). A striking degree of similarity between these proteins has been shown, although it appears that the mode of regulation may vary.

In this study, the recA+ gene of M. tuberculosis has been cloned and sequenced and its protein product has been identified. As many mycobacterial genes are expressed poorly in E. coli (27, 57), DNA hybridization was used to search for a recA-like gene from M. tuberculosis rather than screening by functional complementation. The ability of this gene, expressed from a plasmid lac promoter, to complement the different functions of RecA in E. coli strains with various recA alleles has been investigated. An unexpected discordance between the coding capacity of the M. tuberculosis recA gene and the size of its product was revealed.

MATERIALS AND METHODS

Bacterial strains, plasmids, and growth conditions. The bacterial strains and plasmids used in this study are described in Table 1. E. coli strains were grown at 37°C in L medium (48) containing the following antibiotics where appropriate at the following concentrations: ampicillin, 100 μg ml⁻¹; tetracycline, 10 μg ml⁻¹; and kanamycin, 50 μg ml⁻¹. MH1521 is a srl::Tn10 recA1 P1 transductant of the sfa::ModAp lac) fusion strain, N3331 (33). MH1522 is a similar srl::Tn10 recA+ derivative. Mycobacteria were grown in modified Dubos medium (Difco) at 37°C. UV mutagenesis and survival. Quantitative assays of UV survival and induced mutagenesis used mid-log-phase cells. UV survival was assayed by colony formation on L-agar plates containing the appropriate antibiotic. UV-induced resistance to rifampin used L agar in the triple overlay technique (51). UV light came from a mercury vapor germicidal lamp calibrated with a UV Products UV meter.

Phage λ lysogenic induction. Phage λ papa lysogens were
tested for UV-induced lysogenic induction by infective-center formation on GY4015 Amp<sup>+</sup> indicator bacteria on GT agar plates containing 15 μg of ampicillin per ml (43).

**Bacterial conjugation.** Matings were performed between *E. coli* HfrH and AB2463 recA13 carrying plasmids with cloned recA sequences (41). Plating on Davis and Mingioli minimal agar (14) lacking threonine selected for growth of females which had acquired the *thr*<sup>+</sup> marker from the male. Streptomycin at 100 μg ml<sup>-1</sup> counterselected for growth of the donor Hfr bacteria. Frequencies of recombinants were normalized to the number of viable female cells, as determined by their colony forming ability on L-agar plates.

**SOS-induced gene expression.** The ability of cloned *M. tuberculosis* sequences to derepress SOS genes was tested in MH1521, a recA13 sfa::Mud (Ap lac) fusion strain. For SOS induction, cultures were treated with mitomycin C for 4 h and assayed for β-galactosidase synthesis as described elsewhere (41).

**Preparation of genomic DNA from mycobacteria.** Genomic DNA was isolated from 250- to 1,000-mL mycobacterial cultures. Two to three generations before harvesting, 0.1 volume of 2 M glycerol was added. Cells were harvested; washed in 15 to 30 mL of 0.3 M sucrose–50 mM Tris-Cl (pH 8.0)–10 mM EDTA; resuspended in 2 mL (per gram of wet weight of cells) of the same buffer containing 2 mg of both lysozyme and lipase per mL; and incubated at 37°C for 30 to 60 min. Four volumes of 6 M guanidinium chloride–1% Sarkosyl–20 mM EDTA was added, and the incubation was continued for 1.5 to 2 h. The mixture was extracted with chloroform and precipitated with ethanol overnight at −20°C. The precipitate was dissolved in 400 to 800 μL of TE (10 mM Tris Cl [pH 8.0], 1 mM EDTA) containing 0.5% sodium dodecyl sulfate (SDS), 0.5 mg of proteinase K per mL, and 4,000 U of RNase T1 by incubating at 37°C for 2 to 3 h; phenol–chloroform extracted; and ethanol precipitated. The DNA was dissolved in an appropriate volume of TE, typically 100 μL per 250 mL of starting culture.

**Recombinant DNA techniques.** Standard molecular and recombinant DNA techniques were used (48). Minipreparation of plasmid DNA has been described elsewhere (24), and large-scale preparations were by alkali lysis followed by pZ523 column purification (Northumbria Biologicals Ltd.). DNA was digested with restriction enzymes (Bethesda Research Laboratories and Northumbria) according to the manufacturers’ instructions. DNA fragments were purified from agarose gels by electrophoresis into Whatman 3MM paper backed with dialysis membrane presoaked in running buffer. Recovery of the DNA-containing solution by centrifugation was followed by precipitation with ethanol. Preparation of competent cells and transformation of *E. coli* strains were performed by the calcium chloride method (12). Nitrocellulose was used for Southern hybridizations (Biotrace NT; Gelman Sciences). Labeling of probes with digoxigenin, hybridization at 65°C, and detection of hybridizing bands was as described by Boehringer Mannheim Ltd. Posthybridization washes were carried out at 65°C with either 2× SSC–0.1% SDS or 0.2× SSC–0.1% SDS as indicated (20× SSC is 3 M NaCl plus 0.3 M sodium citrate). DNA sequencing was performed on random clones in M13 generated by sonication (3) by the dideoxy chain termination

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Relevant genotype</th>
<th>Reference(s) or source&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>M. tuberculosis</em> H37Rv</td>
<td>thr&lt;sup&gt;-&lt;/sup&gt;1 ara-14 leuB6 Δggt-proA62 supE44 tss-33 hisG4 rpsL31 argE3 thi&lt;sup&gt;-&lt;/sup&gt;1</td>
<td>25</td>
</tr>
<tr>
<td><em>M. tuberculosis</em> H37Ra</td>
<td>Same as AB1157 plus recA13</td>
<td>25</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td></td>
<td>Laboratory stock</td>
</tr>
<tr>
<td>HfrH</td>
<td>rasL&lt;sup&gt;+&lt;/sup&gt;</td>
<td>48</td>
</tr>
<tr>
<td>DH5</td>
<td>recA13</td>
<td>7</td>
</tr>
<tr>
<td>HB101</td>
<td>recA56</td>
<td>31</td>
</tr>
<tr>
<td>CL83</td>
<td>ΔrecA</td>
<td>38, 49</td>
</tr>
<tr>
<td>KM4104</td>
<td>sfa::Mud(Ap lac)</td>
<td>33</td>
</tr>
<tr>
<td>MH1521</td>
<td>sfa::Mud(Ap lac) src::Tn10 recA1</td>
<td>43</td>
</tr>
<tr>
<td>MH1522</td>
<td>Same as MH1521 but recA&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>GY4015</td>
<td>Amp&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>NM522</td>
<td>supE thi Δ(lac-proA) Δhisd-5 F&lt;sup&gt;+&lt;/sup&gt;[proAB&lt;sup&gt;+&lt;/sup&gt; lacF]lacZΔM15]</td>
<td>21</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pBR322</td>
<td>Same as pUC18 (58), but with T7 promoter and fl origin</td>
<td>40</td>
</tr>
<tr>
<td>pTZ18R</td>
<td>recA E. coli</td>
<td>49</td>
</tr>
<tr>
<td>pDR145</td>
<td>recA M. tuberculosis 14-kb EcoRI fragment in pBR322</td>
<td></td>
</tr>
<tr>
<td>pEJ126</td>
<td>Same as pEJ126 with insert in opposite orientation</td>
<td></td>
</tr>
<tr>
<td>pEJ127</td>
<td>recA M. tuberculosis 2.8-kb PvuII fragment in pTZ18R recA in opposite orientation to lacZ</td>
<td></td>
</tr>
<tr>
<td>pEJ134</td>
<td>Same as pEJ135, but recA in same orientation as lacZ</td>
<td></td>
</tr>
<tr>
<td>pEJ135</td>
<td>Same as pTZ18R but Amp&lt;sup&gt;+&lt;/sup&gt; Kan&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pEJ141</td>
<td>Same as pEJ134 but Amp&lt;sup&gt;+&lt;/sup&gt; Kan&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pEJ143</td>
<td>Same as pEJ135 but Amp&lt;sup&gt;+&lt;/sup&gt; Kan&lt;sup&gt;+&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>pEJ145</td>
<td>Same as pEJ143</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Strains or plasmids with no reference or source were constructed as part of this study.
method (50), and the data were assembled by using the DB system of Staden (56). Further analysis of the DNA and protein sequences were carried out with The University of Wisconsin Genetics Computer Group (GCG) package of programs (15).

Detection of the RecA protein. Cell pellets were lysed and electrophoresed through 10% SDS–polyacrylamide gels and subjected to Western blotting (immunoblotting) with polyclonal rabbit antibodies to E. coli RecA protein by using standard methods (48). Primary antibody binding was visualized with a secondary antibody–alkaline phosphatase conjugate.

Nucleotide sequence accession number. The sequence has been deposited with the EMBL data bank and has been assigned accession no. X58485.

RESULTS

Identification of the M. tuberculosis recA+ gene. The presence of a recA-like gene in M. tuberculosis genomic DNA was investigated by hybridization with a 0.6-kb PstI-EcoRI probe internal to the recA+ gene of E. coli, and a partial restriction map of this region of the chromosome of M. tuberculosis was constructed (Fig. 1a).

The smallest fragment hybridizing was a 2-kb SsrI-HindIII fragment. Given the size of recA genes isolated from other organisms, it could be predicted that the putative recA-like sequence detected here would be contained within a 14-kb EcoRI fragment of M. tuberculosis DNA (Fig. 1a). Therefore, size-fractionated EcoRI fragments of M. tuberculosis DNA approximately 14 kb long were cloned into pBR322. Plasmids from the 227 transformants were screened by DNA hybridization in pools of 10 plasmids using the E. coli recA+ probe. One clone was obtained from each of two pools. These two plasmids, pEJ126 and pEJ127, have the insert in opposite orientations.

The restriction maps of these clones were consistent with the partial map of the recA region of the M. tuberculosis chromosome previously constructed (Fig. 1). To verify further the authenticity of these clones, the 14-kb EcoRI insert was purified and used as a probe against DNAs of M. tuberculosis H37Rv and H37Ra and M. bovis BCG. These DNAs and the clones themselves were each digested with EcoRI plus SsrI. EcoRI plus HindIII digests were also done with H37Rv total DNA and the clones. Identical bands of hybridization were obtained for each of the DNAs (Fig. 2), chromosomal or cloned, confirming that the cloned DNA originated from M. tuberculosis and had not undergone major rearrangement during cloning.

Further Southern hybridization of pEJ126 with the E. coli recA probe detected a 2.8-kb PvuII fragment which spanned the SsrI site of the SsrI-HindIII fragment previously identified, and both SsrI-PvuII fragments resulting from a double digest gave signals. This suggested that the recA gene crossed the SsrI site.

Functional analyses. Complementing activity of the putative recA clones in pEJ126 and pEJ127 was tested in E. coli recA1, recA13, recA56, and ΔrecA mutants. However, no complementation of the UV sensitivity of any of these mutants by either plasmid was detected (data not shown). A possible explanation for the lack of complementation could be that the M. tuberculosis sequences were not being expressed in E. coli. Hence, the 2.8-kb PvuII fragment identified above by hybridization with E. coli recA was subcloned into the actively transcribed lacZ gene of pTZ18R, giving plasmids pEJ134 and pEJ135. The putative recA gene of M. tuberculosis was in the same transcriptional orientation as lacZ in pEJ135 and in the opposite orientation to lacZ in pEJ134. Complementation assays for recA activity in recombination, repair, and SOS regulation were done with these
TABLE 2. Plating efficiency of λ red gam

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>KM4104 (ΔrecA)</th>
<th>HB101 (recA13)</th>
<th>DH5 (recA1)</th>
<th>CL83 (recA56)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pDR1453</td>
<td>1 &lt;10&lt;sup&gt;-5&lt;/sup&gt;</td>
<td>1 &lt;10&lt;sup&gt;-5&lt;/sup&gt;</td>
<td>1 &lt;10&lt;sup&gt;-5&lt;/sup&gt;</td>
<td>1 &lt;10&lt;sup&gt;-5&lt;/sup&gt;</td>
</tr>
<tr>
<td>pEJ135</td>
<td>0.9 &lt;10&lt;sup&gt;-5&lt;/sup&gt;</td>
<td>0.9 &lt;10&lt;sup&gt;-5&lt;/sup&gt;</td>
<td>0.4 &lt;10&lt;sup&gt;-5&lt;/sup&gt;</td>
<td>0.4 &lt;10&lt;sup&gt;-5&lt;/sup&gt;</td>
</tr>
<tr>
<td>pEJ134</td>
<td>&lt;10&lt;sup&gt;-5&lt;/sup&gt;</td>
<td>&lt;10&lt;sup&gt;-5&lt;/sup&gt;</td>
<td>&lt;10&lt;sup&gt;-5&lt;/sup&gt;</td>
<td>&lt;10&lt;sup&gt;-5&lt;/sup&gt;</td>
</tr>
<tr>
<td>pTZ18R</td>
<td>&lt;10&lt;sup&gt;-5&lt;/sup&gt;</td>
<td>&lt;10&lt;sup&gt;-5&lt;/sup&gt;</td>
<td>&lt;10&lt;sup&gt;-5&lt;/sup&gt;</td>
<td>&lt;10&lt;sup&gt;-5&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

* Values are relative to those of pDR1453 for each strain.

clones to test whether the *M. tuberculosis* sequence was functional when expressed from the heterologous *lacZ* promoter of pTZ18R.

Complementation of the *recA* defect in homologous recombination was assayed by the plating efficiency of λ red gam mutants and by recombination of incoming Hfr DNA in a bacterial mating. λ red gam phages can only grow in recombination-proficient *E. coli*. Thus, the high plating efficiency of these phages in *recA1, recA13, recA56*, and Δ*recA* mutants carrying pEJ135 indicates that the *M. tuberculosis* DNA cloned in this plasmid provides recombinational activity (Table 2). Indeed, the levels of λ red gam growth were comparable to those provided by the control plasmid, pDR1453, which encodes the *E. coli recA* gene. Furthermore, pEJ134, in which the insert DNA is in the opposite orientation to the direction of *lacZ* transcription, did not allow growth of λ red gam and so did not complement. A second, conjugational assay of recombination activity used the multiple auxotroph, AB2463 recA13, as a female in a mating with *E. coli* HfrH. When the female carried the control plasmid, pDR1453, a high frequency of 7 × 10<sup>-2</sup> Thr<sup>+</sup> recombinants per recipient was found. The frequency of recombinants in *recA13* mutants carrying pEJ135 was 10<sup>-2</sup>. Nevertheless, this lower level of complementation by pEJ135 was still 10-fold higher than that found with pEJ134 or with the vector alone. Thus, pEJ135 appears to be recombinationally functional in two different assays, although with different efficiencies.

Complementation of the *recA* defect in repair was assayed by UV survival and was allele dependent. A level of UV resistance intermediate between that provided by the *E. coli* clone pDR1453 and the extreme sensitivity with the vector alone was observed with pEJ135 in *recA13* HB101 and *recA56* KM4104, while a lesser resistance was seen in *recA1* DH5; no complementation at all was seen in the *recA56* mutant, CL83 (Fig. 3). In contrast, *recA13, recA56*, and Δ*recA* mutants carrying pEJ134 remained as sensitive to UV as cells carrying the vector alone and so were not complemented. However, in *recA1* DH5, pEJ134 did confer some UV resistance.

One of the *recA*-dependent processes contributing to UV resistance is a mutagenic DNA repair process which can enhance survival with an inherent probability of inducing a mutational change. UV-induced mutagenesis with rifampin resistance is an indicator of the activity of this repair system. pEJ135 restored mutability to the *recA13* mutant HB101 to levels similar to those of the cloned *E. coli recA* gene of pDR1453 (Fig. 4). No complementation of the mutagenesis defect of HB101 was found with either pEJ134 or the vector pTZ18R. This further demonstrates RecA protein-like activity stemming from expression of the cloned *M. tuberculosis* sequences of pEJ135.

Cloned *M. tuberculosis* DNA was also tested for its ability to complement the SOS regulatory defects of *E. coli recA* mutants in inducing SOS genes and prophages. The ability of the *M. tuberculosis* RecA protein to derepress the λ prophage was assessed by UV induction of λ lysogenic derivatives of *E. coli* HB101, DH5, and CL83. KM4104 has the λ *att* site deleted and could not be used. To allow measurement of the infective centers by counterselection of the lysogenic strains in the presence of ampicillin-resistant indicator bacteria (43), the vector pTZ18R and the clones pEJ134 and pEJ135 were made ampicillin sensitive by cloning a kanamycin resistance gene into the *bla* gene, yielding plasmids pEJ141, pEJ143, and pEJ145, respectively. No infective centers were observed either before or after UV irradiation of lysogenic derivatives of the *recA* mutant strains HB101, DH5, and CL83 carrying pEJ145, although *recA* cells carrying pDR1453 yielded 10% infective centers at 3 J m<sup>-2</sup> of UV irradiation. Thus, the cloned *M. tuberculosis* sequences did not appear to complement for cleavage of the λ repressor.

Complementation of SOS gene expression resulting from cleavage of the LexA repressor was tested with a *recA1* mutant carrying an SOS-inducible *sfiA-lacZ* fusion (33). In these cells, LexA cleavage allows derepression of the *sfiA-lacZ* fusion for the synthesis of β-galactosidase. After 4 h of
TABLE 3. Induction of a sfiA-lacZ fusion by mitomycin C

<table>
<thead>
<tr>
<th>Mitomycin C (μg ml⁻¹)</th>
<th>Mean β-galactosidase activity (U)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MH1521(pEJ141)</td>
</tr>
<tr>
<td>0</td>
<td>78</td>
</tr>
<tr>
<td>0.1</td>
<td>46</td>
</tr>
<tr>
<td>0.3</td>
<td>73</td>
</tr>
<tr>
<td>0.5</td>
<td>62</td>
</tr>
<tr>
<td>1.0</td>
<td>64</td>
</tr>
</tbody>
</table>

* Treatment was for 4 h.

Values are the means from duplicate measurements.

exposure to up to 0.3 μg of mitomycin per ml, there was increased β-galactosidase synthesis in cells carrying pEJ145 expressing M. tuberculosis recA (Table 3). Induction in these cells and in control recA⁺ E. coli MH1522 reached its maximum at this concentration of mitomycin C. The level of induction produced by the cloned M. tuberculosis recA gene was approximately one-third that of normal E. coli induction. There was no increase in β-galactosidase synthesis with vector plasmid pEJ141 or pEJ143 with the unexpressed M. tuberculosis recA gene. These results indicate that the M. tuberculosis RecA protein is able to bring about LexA cleavage and SOS induction.

**Detection of the M. tuberculosis RecA protein.** Many RecA proteins from other organisms interact with antibodies raised to the E. coli RecA protein, suggesting their high degree of structural conservation. Accordingly, an E. coli ΔrecA mutant carrying plasmids with cloned M. tuberculosis recA sequences was tested for production of RecA-like proteins by Western blotting with E. coli RecA antibodies. A band of similar size to that of the E. coli control, which was absent with the vector pTZ18R, was detected in extracts of E. coli containing pEJ135 (Fig. 5). This band was not detected in E. coli carrying pEJ134 with the cloned sequence in the opposite orientation or with the original clone pEJ126, which has 8.5 kb of upstream DNA, confirming that the M. tuberculosis recA gene was not expressed from its own promoter. Although the RecA-like band detected was relatively faint, it is not clear whether this is due to poor expression or weak cross-reactivity with the E. coli RecA antibody.

**DNA sequence analysis.** The 2.8-kb PvuII fragment cloned in pEJ135 was sequenced on both strands (Fig. 6). An open reading frame (ORF) whose translation had homology to E. coli RecA and which spanned the SsrI site thought to be internal to the gene by DNA hybridization analyses was found. However, this ORF was approximately twice the size of other recA genes. The protein encoded by this gene exhibited a high degree of homology with the E. coli RecA sequence over its first 254 and its last 96 amino acids (Fig. 7), but in between there was a large nonhomologous region of 440 amino acids. Strikingly, the point in the E. coli sequence at which homology with the N terminus of the M. tuberculosis sequence stopped was coincident with where homology with the C terminus began. This is suggestive of a sequence having been inserted into an original recA gene and then having been modified to produce a continuous ORF. No homology with the intervening sequence was found by searching the GenBank (version 10.0) or Swiss (version 15.0) protein data bases or the EMBL (version 25.0) DNA data base by using the FASTA algorithm of Lipman and Pearson (32).

The protein encoded by the M. tuberculosis recA gene contains 790 amino acids and would be predicted to have a molecular mass of 85,390 Da, in contrast to the apparent molecular mass of ca. 38,000 Da seen by Western blotting. No LexA binding site upstream of the recA gene was identified, and no good E. coli-like promoter was recognizable, in keeping with the lack of expression of RecA in E. coli from the original clones pEJ126 and pEJ127, which would rely on M. tuberculosis promoter sequences. However, a sequence resembling that of heat shock promoters was found 100 bp upstream of the initiation codon (Fig. 6); it
consisted of the elements TGTCACACTTGAAG and CGGC
TACTG, compared with the E. coli consensus sequences of
TNtCNCCCTTGA and CCCCATYTyta (13). In the same
region, there is a sequence identity to a consensus GAAC
N2-GTTC sequence identified in various DNA damage-
inducible promoters in B. subtilis (11).

FIG. 6. Sequence of the recA gene of M. tuberculosi s. The DNA sequence of the 2.8-kb PvuII fragment is shown with the protein translation of the recA gene. A probable Shine-Dalgarno sequence is underlined, and a possible heat shock promoter-like sequence is overlined.

... of the elements TGTCACACTTGAAG and CGGC
TACTG, compared with the E. coli consensus sequences of
TNtCNCCCTTGA and CCCCATYTyta (13). In the same
region, there is a sequence identity to a consensus GAAC
N2-GTTC sequence identified in various DNA damage-
inducible promoters in B. subtilis (11).

FIG. 6. Sequence of the recA gene of M. tuberculosi s. The DNA sequence of the 2.8-kb PvuII fragment is shown with the protein translation of the recA gene. A probable Shine-Dalgarno sequence is underlined, and a possible heat shock promoter-like sequence is overlined.

... of the elements TGTCACACTTGAAG and CGGC
TACTG, compared with the E. coli consensus sequences of
TNtCNCCCTTGA and CCCCATYTyta (13). In the same
region, there is a sequence identity to a consensus GAAC
N2-GTTC sequence identified in various DNA damage-
inducible promoters in B. subtilis (11).

FIG. 6. Sequence of the recA gene of M. tuberculosi s. The DNA sequence of the 2.8-kb PvuII fragment is shown with the protein translation of the recA gene. A probable Shine-Dalgarno sequence is underlined, and a possible heat shock promoter-like sequence is overlined.

... of the elements TGTCACACTTGAAG and CGGC
TACTG, compared with the E. coli consensus sequences of
TNtCNCCCTTGA and CCCCATYTyta (13). In the same
region, there is a sequence identity to a consensus GAAC
N2-GTTC sequence identified in various DNA damage-
inducible promoters in B. subtilis (11).

FIG. 6. Sequence of the recA gene of M. tuberculosi s. The DNA sequence of the 2.8-kb PvuII fragment is shown with the protein translation of the recA gene. A probable Shine-Dalgarno sequence is underlined, and a possible heat shock promoter-like sequence is overlined.

... of the elements TGTCACACTTGAAG and CGGC
TACTG, compared with the E. coli consensus sequences of
TNtCNCCCTTGA and CCCCATYTyta (13). In the same
region, there is a sequence identity to a consensus GAAC
N2-GTTC sequence identified in various DNA damage-
inducible promoters in B. subtilis (11).

FIG. 6. Sequence of the recA gene of M. tuberculosi s. The DNA sequence of the 2.8-kb PvuII fragment is shown with the protein translation of the recA gene. A probable Shine-Dalgarno sequence is underlined, and a possible heat shock promoter-like sequence is overlined.

... of the elements TGTCACACTTGAAG and CGGC
TACTG, compared with the E. coli consensus sequences of
TNtCNCCCTTGA and CCCCATYTyta (13). In the same
region, there is a sequence identity to a consensus GAAC
N2-GTTC sequence identified in various DNA damage-
inducible promoters in B. subtilis (11).

FIG. 6. Sequence of the recA gene of M. tuberculosi s. The DNA sequence of the 2.8-kb PvuII fragment is shown with the protein translation of the recA gene. A probable Shine-Dalgarno sequence is underlined, and a possible heat shock promoter-like sequence is overlined.

... of the elements TGTCACACTTGAAG and CGGC
TACTG, compared with the E. coli consensus sequences of
TNtCNCCCTTGA and CCCCATYTyta (13). In the same
region, there is a sequence identity to a consensus GAAC
N2-GTTC sequence identified in various DNA damage-
inducible promoters in B. subtilis (11).

FIG. 6. Sequence of the recA gene of M. tuberculosi s. The DNA sequence of the 2.8-kb PvuII fragment is shown with the protein translation of the recA gene. A probable Shine-Dalgarno sequence is underlined, and a possible heat shock promoter-like sequence is overlined.

... of the elements TGTCACACTTGAAG and CGGC
TACTG, compared with the E. coli consensus sequences of
TNtCNCCCTTGA and CCCCATYTyta (13). In the same
region, there is a sequence identity to a consensus GAAC
N2-GTTC sequence identified in various DNA damage-
inducible promoters in B. subtilis (11).

FIG. 6. Sequence of the recA gene of M. tuberculosi s. The DNA sequence of the 2.8-kb PvuII fragment is shown with the protein translation of the recA gene. A probable Shine-Dalgarno sequence is underlined, and a possible heat shock promoter-like sequence is overlined.

... of the elements TGTCACACTTGAAG and CGGC
TACTG, compared with the E. coli consensus sequences of
TNtCNCCCTTGA and CCCCATYTyta (13). In the same
region, there is a sequence identity to a consensus GAAC
N2-GTTC sequence identified in various DNA damage-
inducible promoters in B. subtilis (11).

FIG. 6. Sequence of the recA gene of M. tuberculosi s. The DNA sequence of the 2.8-kb PvuII fragment is shown with the protein translation of the recA gene. A probable Shine-Dalgarno sequence is underlined, and a possible heat shock promoter-like sequence is overlined.

... of the elements TGTCACACTTGAAG and CGGC
TACTG, compared with the E. coli consensus sequences of
TNtCNCCCTTGA and CCCCATYTyta (13). In the same
region, there is a sequence identity to a consensus GAAC
N2-GTTC sequence identified in various DNA damage-
inducible promoters in B. subtilis (11).

FIG. 6. Sequence of the recA gene of M. tuberculosi s. The DNA sequence of the 2.8-kb PvuII fragment is shown with the protein translation of the recA gene. A probable Shine-Dalgarno sequence is underlined, and a possible heat shock promoter-like sequence is overlined.
coding capacity. Furthermore, protein products from the two species are antigenically cross-reactive and have the same electrophoretic mobility. At the functional level, heterologous expression of the *M. tuberculosis* sequences could partially complement the defects of *E. coli recA* mutants in homologous recombination and in the repair and regulatory responses needed for recovery from DNA damage. The degree of complementation by the *M. tuberculosis* DNA depended on which assay was used as an indicator of RecA-like activity and also upon which *recA* allele of *E. coli* was present. Partial complementation can be understood when the requirements for full complementation of the various roles of RecA are considered.

In homologous recombination, *E. coli* RecA protein interacts with single-stranded DNA and ATP as well as single-stranded binding protein to form multimeric spiral filaments which ultimately bring about synthesis of duplex DNA (26). Homologous recombination activity of the *M. tuberculosis* clone was indicated by the high plating efficiency of *λ* red zymogen phages and, to a lesser degree, by the intermediate levels of recombination of Hfr DNA in a bacterial mating. Differences in the degree of complementation between the two assays may indicate different abilities for recombination on different types of substrate or may be based on inadequate levels of gene expression from the heterologous promoter of the cloned sequence. RecA–single-stranded DNA–ATP ternary complexes are also the allosteric effectors of the autocatalytic digestion of LexA and λ repressors, whose cleavage derepresses the SOS response (54, 55), and of UmuD cleavage to produce UmuD' needed for the mutagenic response to DNA damage (8, 44, 53). Recovery after UV irradiation is therefore a composite of the direct action of RecA protein acting in recombinational repair, its indirect action in derepressing other DNA repair genes by LexA repressor autocleavage, and its indirect action in priming UmuD protein autocleavage to UmuD'. The intermediate levels of UV resistance produced by the *M. tuberculosis* clone therefore indicate the presence of RecA activity but not necessarily which aspect. However, the assays of homologous recombination would predict some recombination repair activity. Also, some autocleavage events would appear to be triggered because *recA* mutants expressing the cloned *M. tuberculosis* recA gene showed induction of a *sfiA*-lacZ fusion. The induction of mutants by UV irradiation also requires the expression of the *umuDC* operon by LexA inactivation (2) and the posttranslational conversion of UmuD to UmuD' (8, 44, 53). A further role of RecA protein in UV-induced mutagenesis has also been identified (44), and so presumably this action too is provided by the cloned *M. tuberculosis* sequences. Thus, recombination, repair, and mutagenesis are all complemented at least to some extent.

However, λ lysogenic induction was not detected after UV irradiation of *E. coli recA* mutants carrying the *M. tuberculosis* clone, indicating that the λ repressor is not triggered into autodigestion even though LexA and UmuD proteins appear to be. The lack of even spontaneous λ lysogenic induction is similar to previous observations with *Legionella pneumophila* RecA (16), which inhibited spontaneous induction of a *recA*+ *E. coli* strain as well as reducing background spontaneous induction in a *recA* mutant to zero, although *M. tuberculosis* RecA does not appear to repress λ induction in a *recA*+ strain (data not shown). Knowing the effect on mycobacterial phages would be of interest, as *B. subtilis* RecA stimulates the induction of bacillus prophages (37), although it does not bring about cleavage of the λ repressor (35), whereas the reverse is true of *E. coli* RecA, which does not stimulate the induction of bacillus prophages (34).

The partial ability of cloned heterologous *recA* genes to interact with some *E. coli* proteins and not others is well documented (40). Indeed, similar effects are seen with *recA*430, *recA*432, *recA*1730, and *recA*1734 mutations of *E. coli* itself, where split phenotypes are produced by these particular mutant gene products performing some of the usual RecA activities while being defective in others (17, 18). Thus, it was proposed that the allosteric effector sites for protein cleavage are not identical, although they may be overlapping. In the case of the *M. tuberculosis* RecA protein, it might be surmised that these sites exist for UmuD and LexA proteins but not for the λ repressor. Alternatively, there might be competitive inhibition of λ repressor cleavage by other target molecules, as was demonstrated in vitro between noncleavable LexA and a λ repressors (54).

The allele dependency of these partial complementation events may be caused by dominance effects of the mutated *E. coli* RecA proteins. This has been described for the *recA*1 and *recA*56 alleles and a family of transposition mutants of cloned *E. coli recA* (18, 30, 60). Such effects are thought to derive from the residual abilities of these mutated proteins to form heteromultimers with other RecA proteins (18). The occurrence of such abortive interactions is in itself an
indicator of functional conservation at the protein-protein recognition level (51).

Superimposed on these functional requirements of the protein, full complementation in E. coli would need inducible gene expression to provide additional RecA protein in response to DNA-damaging treatments. However, the M. tuberculosis DNA isolated here was only expressed from a heterologous lacZ promoter rather than its own upstream sequences. Examination of the putative promoter region revealed no motifs like E. coli SOS boxes, and it is not known whether RecA activity is inducible in M. tuberculosis itself or whether a LexA-like protein exists. Nevertheless, a GAAC-N4-7 GTT motif, common to the damage-inducible promoters of the B. subtilis SOS response (11), is present in the M. tuberculosis recA DNA. Thus, there may be inducible expression through the regulation of a LexA-like protein with a recognition sequence different from that of its E. coli counterpart. The Anaebaena variabilis recA gene has no E. coli-like SOS box, but its expression is UV-, mitomycin C-, and MMS-inducible in A. variabilis itself (45). In contrast, in Thiobacillus ferrooxidans, the recA gene has no SOS box and is not inducible (46).

The conservation of RecA-like activities in M. tuberculosis adds to a long list of similar activities from more than 20 other species. However, the structure of the M. tuberculosis recA gene is radically different from any other so far described in that it consists of an ORF approximately twice the usual conserved size. RecA protein homology with the product of this elongated gene is not dispersed but is localized at the beginning and the end. The break points of the RecA-like region are in register with the uninterrupted E. coli RecA sequence. Thus, elongation of the M. tuberculosis recA gene appears to have occurred by the in-frame insertion of 1,320 bp of DNA approximately two-thirds of the way through the prototypical recA sequence. The resulting 2,370-bp gene therefore has homology split to the beginning and end of the ORF. The same size recA locus was found in genomic DNAs of two isolates of M. tuberculosis and one stock of M. bovis BCG, and so it is not an artifact of cloning.

Although the M. tuberculosis recA gene is more than twice the usual length, the protein product appeared to be approximately the same size as E. coli RecA. This result therefore points to some hitherto unknown processing reaction in the production of M. tuberculosis RecA protein. It should be stressed that this reaction is occurring in E. coli and that the nature of the RecA product of M. tuberculosis itself has not been examined. The putative processing, therefore, is either using E. coli components or relying on integral activities of the M. tuberculosis recA transcript or gene product. There are three general possibilities for how processing might occur, namely, RNA splicing, protein truncation by proteolysis, or protein splicing.

The first possibility is that a mechanism of transcript splicing occurs in the synthesis of M. tuberculosis RecA protein. RNA splicing in prokaryotes is not widespread, and examples have been found only in bacteriophages and archeabacteria (11, 29, and references therein). However, the boundaries of introns are marked by termination signals, in contrast to the single, continuous ORF found here (4).

Protein truncation envisages that an elongated RecA protein is made which then undergoes cleavage around residue 352 to yield an N-terminal fragment approximately the same size as conventional RecA proteins. In a multiple alignment of RecA protein sequences (47), diversification appears to have occurred through a gradual process of base change and mutagenesis, particularly in the C-terminal third of the molecules, and it is around here that the radical diversification of M. tuberculosis occurs. Thus, M. tuberculosis RecA would, like all other RecA proteins, retain many conserved N-terminal motifs but would have acquired a new pseudo C-terminus by insertional mutagenesis and proteolytic cleavage to discard the original prototypical RecA C terminus. Preliminary work shows that deletion of 1,118 bp of 3'-terminal coding sequences from the clone abolishes all complementation activity, even though more of the ORF would be retained than would be needed to make a RecA protein by simple proteolytic truncation.

The final possibility for producing normal-size RecA protein from the elongated M. tuberculosis recA gene would be some form of protein splicing to retain the RecA-like N and C termini and discard the central atypical section (6, 10, 28). A similar arrangement of interrupted homology occurs in the Saccharomyces cerevisiae VMA1/TFP1 gene (23, 52) coding for vacuolar H+-adenosine triphosphatase, and evidence of protein splicing has been found previously (28). A comparison of the protein sequences at the proposed splice sites of VMA1/TFP1 with those at the break points of homology in M. tuberculosis RecA shows some conserved residues, particularly a VVHNC motif at the second junction (Fig. 8). The alignment generated by the GCG GAP program would suggest that the junction in M. tuberculosis RecA is offset by two amino acids, but it seems more likely that the splicing would occur at the same point relative to the amino acids conserved with VMA1/TFP1, as this would only remove one amino acid identity, and the break points in homology are more precisely defined in the VMA1/TFP1 protein.

In summary, a recA locus of M. tuberculosis has been isolated and characterized by both physical and biological criteria. This gene has approximately twice the coding capacity of other recA genes, yet it produces a normal size RecA protein. Clearly, the elucidation of this processing event should reveal its significance in bacterial physiology.
reaction in the synthesis of RecA protein has general importance to our understanding of gene expression in mycobacteria.

ACKNOWLEDGMENTS

We are extremely grateful to Stephen West and Carol Parsons of the Imperial Cancer Research Fund for testing for the presence of RecA-like proteins in this work. We thank Nigel Davis and Kevin Hiom for discussions and helpful advice and Pat Brooks for valuable technical assistance.

REFERENCES