Molecular Cloning and Sequence of the thdF Gene, Which Is Involved in Thiophene and Furan Oxidation by Escherichia coli

KISWAR Y. ALAM† AND DAVID P. CLARK*
Department of Microbiology, Southern Illinois University, Carbondale, Illinois 62901-6508

Received 19 November 1990/Accepted 24 July 1991

Our previous work resulted in the isolation of mutant strains of Escherichia coli K-12 which were able to oxidize furans and thiophenes as a result of mutations in several novel genes. Some of the genes involved in thiophene oxidation were cloned into the multicopy vector pUC19. The plasmid pKA10 carries a 3.8-kb chromosomal fragment which encodes a previously undiscovered gene involved in thiophene oxidation. Three proteins with approximate molecular sizes of 48, 30, and 26 kDa were overproduced by cells carrying pKA10. Maxicell experiments and DNA sequence analysis indicated that the 48- and 26-kDa proteins are encoded by pKA10, whereas the 30-kDa protein is apparently chromosomally derived. A cassette specifying kanamycin resistance was inserted into various sites on pKA10. An insertion which abolished the 48-kDa protein also abolished thiophene oxidation. Chromosomal integration of pKA10::Kan allowed us to locate the chromosomal insert of pKA10 at 84 min on the E. coli genetic map by transduction. Since no previously identified genes involved in thiophene metabolism are located in this region, we designated the gene for the 48-kDa protein as thdF. Sequencing of the 3.8-kb insert revealed an overlap of several hundred bases with the regulatory and structural regions of the tnaA gene, which is also located at 84 min. The 26-kDa protein is probably truncated tnaA protein. An open reading frame corresponding to the 48-kDa thdF protein was located next to the tnaA gene, which encodes tryptophanase, but was transcribed in the opposite sense.

The degradation of aromatic hydrocarbons by bacteria has been extensively investigated, and plasmids carrying the toluene or naphthalene degradation systems have been analyzed genetically (14, 36). Much has been made of the ecological importance of such degradative systems and their possible applicability in biodegradation and environmental cleanup. However, many xenobiotics and pollutants are not merely hydrocarbons but contain heteroatoms. Many dyes and detergents contain sulfonic acid groups, whereas most coal and oil contains significant amounts of reduced sulfur moieties, including those in which the sulfur is an integral part of the ring system (18). The simplest aromatic sulfur heterocycles are those based on the thiophene ring system. Although compounds containing thiophene rings form a major fraction of the organic sulfur component of fossil fuels such as crude oil and coal, they are rarely found in living organisms (15, 26). One recently discovered exception is caldariellaquinone, which contains a benzothiophene ring system and is found in certain sulfur-metabolizing archaeabacteria (37).

Since sulfur dioxide emission from burning high-sulfur coals is a major contributor to acid rain, it is important to develop bacteria which are capable of efficiently removing the sulfur from coal before combustion (11, 18, 26). Inorganic sulfur can be removed from coal by certain strains of Thioacellulosis or Sulfolobus (15, 17, 18); however, the organic sulfur remains intransigent. Since high-sulfur Illinois coals typically contain 60 to 70% of their sulfur in the form of the heterocyclic thiophene ring (3), we have started to investigate the biodegradation of derivatives of thiophene and the corresponding oxygen heterocycle, furan (1, 9).

Our previous work resulted in the isolation of a triple mutant, NAR30, capable of oxidizing a range of furan and thiophene derivatives (1). Mutations at three novel loci, thdA, thdC, and thdD, were required together with constitutive mutations in the fadR (= "thdB") (28) and aotC (29) genes. However, NAR30 does not completely degrade thiophenes or furans, and its oxidation of these compounds is slow and inefficient. We decided to clone the thd genes both to increase the efficiency of degradation and to investigate the nature of the reactions involved.

We cloned two different chromosomal fragments from NAR30, both of which confer the ability to oxidize thiophenes and furans. One of these most likely carries the thdA gene itself (2). The other, described in this report, carries a novel chromosomal locus which when present in high copy mediates the oxidation of certain heterocyclic substrates. This new gene has been designated thdF.

MATERIALS AND METHODS

Bacterial strains and media. Bacterial strains are all derivatives of Escherichia coli K-12 and are described in Table 1. Rich broth contains (per liter) 10 g of tryptone, 5 g of NaCl, and 1 g of yeast extract. The minimal medium used was M9 (25) for growth tests or medium E (34) for genetic manipulations. Sugars, succinate, glycerol, etc., were used at concentrations of 0.4% as carbon sources, whereas aromatic and heterocyclic substrates were used at 0.1% (wt/vol), as most of these are moderately toxic at higher concentrations. Amino acids (50 mg/liter) and vitamins (5 mg/liter) were used when appropriate. Solid media contained 1.5% (wt/vol) Bacto Agar (Difco). The tetrazolium indicator plates used were modified from those of Bochner and Savageau (7) by using M9 salts as the buffer and using aromatic substrates at 0.1% final concentrations. Antibiotics were used at the following concentrations: ampicillin, 100 mg/liter; tetracycline, 10 mg/liter; kanamycin, 25 mg/liter; and chloramphenicol, 30 mg/liter.

* Corresponding author.
† Present address: Department of Medicine, Medical College of Ohio, Toledo, Ohio 43614.
TABLE 1. Bacterial strains and plasmids

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Relevant characteristics</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DC625</td>
<td>gyRA fadR atoc adhC</td>
<td></td>
</tr>
<tr>
<td>NAR11</td>
<td>thdA of DC625</td>
<td>1</td>
</tr>
<tr>
<td>NAR30</td>
<td>thdA thdC thdD of DC625</td>
<td>1</td>
</tr>
<tr>
<td>NAR420</td>
<td>zba-300::Tn10 thdA* of</td>
<td>P1 (SG20523) x NAR30</td>
</tr>
<tr>
<td></td>
<td>NAR30</td>
<td></td>
</tr>
<tr>
<td>NAR710</td>
<td>srl::Tn10 recA of DC625</td>
<td>P1 (NK6659) x DC625</td>
</tr>
<tr>
<td>NAR820</td>
<td>recA::Cam of NAR420</td>
<td>P1 (GM5180) x NAR420</td>
</tr>
<tr>
<td>NK6659</td>
<td>srl::Tn10 recA</td>
<td>N. Kleckner</td>
</tr>
<tr>
<td>GM5180</td>
<td>recA::Cam</td>
<td>G. Walker</td>
</tr>
<tr>
<td>SG20523</td>
<td>zba-300::Tn10</td>
<td>S. Gottesman</td>
</tr>
<tr>
<td>NAR955</td>
<td>zic-501::Tn10</td>
<td>P1 (JW355) x W1485</td>
</tr>
<tr>
<td>NAR967</td>
<td>thdE::Kan of DC679</td>
<td>Integration of pKA20</td>
</tr>
<tr>
<td>NAR968</td>
<td>metE::Tn10</td>
<td>P1 (A313) x W1485</td>
</tr>
<tr>
<td>NAR977</td>
<td>thdE::Kan of HfrC</td>
<td>P1 (NAR967) x HfrC</td>
</tr>
<tr>
<td>JW355</td>
<td>asB relA thyA thi-1 deo</td>
<td>J. Wechsler</td>
</tr>
<tr>
<td></td>
<td>zic-501::Tn10</td>
<td></td>
</tr>
<tr>
<td>A313</td>
<td>metE::Tn10 trxA::Kan</td>
<td>M. Russell</td>
</tr>
<tr>
<td>18404</td>
<td>zic-296::Tn10</td>
<td>C. Gross</td>
</tr>
<tr>
<td>JRG997</td>
<td>ilv metB ampA</td>
<td>J. Guest</td>
</tr>
<tr>
<td>BW6159</td>
<td>ilv::Tn10</td>
<td>B. Wanner</td>
</tr>
<tr>
<td>DC679</td>
<td>fadR atoc adhC</td>
<td>Laboratory collection</td>
</tr>
<tr>
<td>DP968</td>
<td>zic::Tn10</td>
<td>D. Fraenkel</td>
</tr>
<tr>
<td>W1485</td>
<td>Wild type</td>
<td>B. Bachmann</td>
</tr>
</tbody>
</table>

Plasmids

- pUC19: Cloning vector, Ap'
- pKA10: 3.8-kb PstI chromosomal fragment in pUC19
- pKA15: 0.9-kb PstI chromosomal fragment in pUC19
- pKA20: Kanamycin cassette in BgII site of pKA10
- pKA30: Kanamycin cassette in HpaI site of pKA10
- pKA35: 1.6-kb ClaI fragment deleted from pKA10
- pKA37: 1.7-kb BamHI fragment of pKA10 subcloned into pUC19
- pKA39: 1.3-kb BamHI fragment of pKA10 subcloned into pUC19

Plasmids and recombinant DNA procedures. Chromosomal DNA was isolated and purified as described by Sato and Mura (32). Plasmids were isolated by alkaline lysis and then ethidium bromide-CsCl density centrifugation as described by Maniatis et al. (22). The rapid plasmid isolation procedure of Birnboim and Doly (6) was used for screening clones by restriction analysis. Ligations with T4 DNA ligase and restriction enzyme digests were performed under conditions recommended by the manufacturer (Bethesda Research Laboratories). DNA fragments were separated by electrophoresis on 0.7% agarose gels in 89 mM Tris-borate-89 mM borate acid–2 mM EDTA. Transformation procedures were as described by Hanahan (13). Plasmids constructed are described in Table 1.

DNA sequencing. DNA sequencing was performed as in our previous analysis of the adhE gene (12). In brief, DNA of plasmid pKA10 was isolated and then digested with appropriate restriction endonucleases. Fragments were ligated into the correspondingly digested single-strand sequencing vectors M13mp18 and M13mp19 (24, 27). Both strands of DNA were completely sequenced by the technique of Sanger et al. (31).

Gel electrophoresis of proteins. Sodium dodecyl sulfate-(SDS)-polyacrylamide gel electrophoresis was done by the method of Laemmli (21). Whole-cell proteins were prepared by boiling in sample buffer for 5 min, and samples were electrophoresed on SDS-12.5% polyacrylamide gels. Molecular weight markers (97,400 to 14,400) were obtained from Bio-Rad.

Plasmid-encoded proteins were labeled by the maxicell technique of Sancar et al. (30). A 1-ml sample of culture (approximately 2 × 10^8 cells per ml) was labeled for 90 min with 10 μCi of [35S]methionine. Proteins were separated on polyacrylamide gels as above.

RESULTS

Cloning of thiophene degradation genes. Chromosomal DNA from the thiophene-degrading strain NAR490 was totally digested with the restriction endonuclease PstI. Fragments were ligated into the PstI site in the polylinker of the multicyclop vector pUC19 (35). The mixture of recombinant plasmids was transformed into strain NAR710, a recA derivative of strain DC625, the fadR atoc parent of the furan-thiophene-degrading mutants. Ampicillin-resistant transformants were selected on tetrazolium indicator plates containing furfuryl alcohol. From among approximately 1,000 transformants, we found 6 which gave a positive response (i.e., a pink color) on the indicator plates. After single-colony purification, these were screened for the presence of plasmids by agarose gel electrophoresis of minislates. Several plasmid-containing transformants were chosen for further work. These fell into two categories: those exemplified by pKA10 and those exemplified by pKA15.

The two thiophene degradation-positive plasmids, pKA10 and pKA15, were purified by the standard CsCl procedure (see Materials and Methods) and subjected to restriction analysis. Plasmid pKA10 contained a 3.8-kb PstI insert, whereas pKA15 had a 0.9-kb PstI fragment. The detailed restriction map of pKA10 is shown in Fig. 1. Of the enzymes used, only BglII cut inside the 0.9-kb insert of pKA15. The restriction maps were not superimposable, it was clear that the smaller chromosomal PstI fragment of pKA15 was not included within the larger PstI fragment of pKA10.

Physiological properties of pKA10 and pKA15. Both pKA10 and pKA15 were transformed into several recA host strains; the transformants were selected by their resistance to ampicillin. Both plasmids conferred the ability to oxidize furfuryl alcohol, thiophene-2-carboxylic acid, and several other heterocyclic substrates on the fadR atoC strain NAR710. These plasmids were also tested in strain NAR820, a derivative of NAR30 which has the wild-type (i.e., degradation-negative) version of the thdA gene but still carries the thdC and thdD mutations. Whereas pKA10 had relatively little effect, pKA15 restored the degradative ability of NAR820 to that of NAR30. The responses of these various constructs to a variety of heterocyclic substrates appear in Table 2. As indicated previously (2), it seems likely that pKA15 carries the thdA gene. Here we are concerned with the insert of pKA10, which apparently carries a novel gene involved in thiophene metabolism.

Insertion analysis of pKA10. The 3.8-kb insert of pKA10 is about three times as long as an average gene. We therefore wished to locate the thd gene(s) more accurately in this segment of DNA. Our approach was to cut open the plasmid at single restriction sites and insert a DNA cassette specify-
ing resistance to kanamycin. This procedure was repeated for a selection of restriction sites around the target plasmid. The insertion derivatives were transformed back into the recA strain, NAR710, which is wild type for thiophene metabolism, and transformants were selected by resistance to ampicillin (on the original plasmid) plus kanamycin (on the insert). The insertion derivatives were then examined for the presence of an active thd gene(s).

We opened up pKA10 using the enzymes BamHI, BglII, and HpaI. Digested pKA10 was run on an agarose gel. The band corresponding to linearized pKA10 was cut out, and the opened-up plasmid was removed from the gel by the Gene-Clean (Bio 101) procedure. The kanamycin resistance cassette was cut out of plasmid pUC4::Kixx (5, 23) by digestion with BamHI or Smal. The cassette was then ligated into the linearized pKA10. After transformation and selection for ampicillin and kanamycin resistance, we isolated several derivatives of pKA10 carrying the kanamycin cassette (Kan). Insertions into pKA10 opened by the restriction enzyme BamHI resulted in derivatives carrying Kan markers in the BamHI site between KpnI and PstI at coordinate 3550 (Fig. 1). These derivatives retained full activity of the thd genes. In addition, derivatives were found which had lost one or more of the small Bam fragments (coordinates 1650 to 2250) and which were defective in thd activity. We also inserted the kanamycin cassette into the unique BglII site (at coordinate 1700) and isolated a plasmid, pKA20, with no thd activity. This plasmid was subjected to restriction analysis to confirm its structure.

The kanamycin cassette of pKA20 could not be removed by either BamHI or BglII digestion since its construction had produced hybrid Bam-Bgl sites. Digestion with PstI gave a characteristic 930-bp fragment from the interior of the Kan (Kixx) element together with the other expected fragments, i.e., a 2,700-bp vector fragment and two pieces of 2,750 and 1,750 bp, each containing part of the chromosomal fragment and an end portion of the Kan (Kixx) fragment. Since pKA20 contains a single insertion which inactivates the thd gene, this must lie around the BglII site, approximately in the middle of the chromosomal portion of pKA10. Insertions of Kan into the unique HpaI site (at coordinate 3000) to create pKA30 did not inactivate the thd gene(s) of pKA10. Table 2 summarizes the responses of host cells with various plasmid derivatives toward furans and thiophene substrates.

### Proteins of plasmid pKA10.

We ran SDS-polyacrylamide gels to analyze the proteins produced by cells carrying plasmid pKA10 and its derivatives pKA20::Kan and pKA30::Kan. Several gels were run, and we found several consistent differences between the parental strain and plasmid carrying derivatives. The gel shown (Fig. 2) compares NAR710 (a recA derivative of the parental strain) with the same strain carrying pUC19 (the vector), pKA10, and several derivatives of pKA10. Cells carrying plasmid pKA10 expressed three prominent protein bands (a, b, and d in Fig. 2) which were largely absent in NAR710 or NAR710(pUC19). The a band is approximately 48 kDa. It was present in pKA10 and pKA30 (both Thd') but absent in pKA20 (Thd-), implying that it is a product of a gene required for thiophene degradation. The 48-kDa protein was produced only in late-exponential and early-stationary-phase cells (i.e., as in Fig. 2) but was not observed in early-exponential-phase cells (data not shown). The 26-kDa protein, band d, was highly expressed by cells carrying pKA10 and pKA37. Plasmid pKA37 carries the BamHI fragment from coordinates 1 to 1675 subcloned into pUC19 and expresses protein d but not a, b, or c. As shown below, this region of our DNA sequence overlaps the previously sequenced mnaA gene encoding typtophanase (10). It seems likely that protein d is a truncated typtophanase gene product. Protein c (29 kDa) was observed only when those plasmids carrying the Kan element were present and is presumably the neomycin-kanamycin phosphotransferase protein, the product of the npt gene. The b protein, of approximately 30 kDa, proved somewhat puzzling. This protein was especially prominent in cells carrying pKA35 and pKA39, which both carry the region between coordinates 2250 and 3550, as well as those with the complete pKA10 insert. However, as discussed below, it does not appear to be plasmid encoded.

To establish which proteins were actually encoded by the

### TABLE 2. Responses of plasmid-bearing strains

<table>
<thead>
<tr>
<th>Host strain*</th>
<th>Geno-type</th>
<th>Plasmid</th>
<th>Substrate*</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIOH</td>
<td>TmOH</td>
<td>TCA</td>
<td>TMSO</td>
</tr>
</tbody>
</table>

| NAR710 Wild type | None | - | - | - | - |
| pKA10 | + | + | + | + | + |
| pKA20 | - | + | - | - | - |
| pKA30 | + + | + + | + + | + + | + |
| pKA15 | + + + | + + + | + + + | + + + | + + + |

| NAR11 thdA | None | + + + | + + + | + + + | + + + |
| NAR30 thdACD | None | + + + + | + + + + | + + + + | + + + + |
| NAR820 thdCD | None | + + | + + | + + | + + |
| pKA10 | + + | + + | + + | + + |
| pKA15 | + + + + | + + + + | + + + + | + + + + |

* Strain NAR710 is a recA derivative of DC625, and NAR820 is a recA derivative of NAR420. All strains are derived from DC625 and contain the fadR, atoC, and adhC mutations.

* FIOH, furfuryl alcohol; TmOH, thiophene methanol; TCA, thiophene-2-carboxylic acid; TMSO, tetramethylenesulfoxide; TMSO2, tetramethylenesulfone. Responses: - no reaction; + + , ++ , +++ , increasing color on tetrazolium indicator medium, with the specified substrate.
plasmid, we performed maxicell experiments. As shown (Fig. 3), only the prominent 48-kDa band (a) and a much fainter 26-kDa band (d), together with the plasmid-derived β-lactamase (31 kDa), were observed from cells carrying pKA10. Plasmid pKA20 showed β-lactamase, npt protein (c), and enhanced production of protein d (26 kDa). The 30-kDa protein (b) observed in extracts of unirradiated cells was not observed in maxicell experiments and is presumably chromosomally encoded. The sequence data support these observations (see below).

**Location of thdF on chromosome.** To decide whether pKA10 actually carried thdA or another thd gene controlled by thdA, it was necessary to find where on the *E. coli* chromosome the DNA inserted in pKA10 had come from.

The plasmid pKA20 contains Kan inserted into the thd gene of plasmid pKA10, causing loss of the thiophene-positive phenotype and the 48-kDa protein band. We therefore recombed pKA20 with the chromosome to insert the Kan marker at the original location of the gene(s) carried on pKA10-pKA20. Plasmid pKA20 was transformed into the recA+ gyr+ strain DC679, and transformants were selected for resistance to both ampicillin and kanamycin. The pKA20-bearing derivative was then grown for many generations in rich broth in the presence of kanamycin but without ampicillin. At appropriate time intervals, aliquots were diluted and plated on rich broth agar. A hundred colonies were picked and tested versus kanamycin and ampicillin. After approximately 100 generations, 9% of the colonies had lost the plasmid, as evidenced by sensitivity to ampicillin but retained kanamycin resistance. After 200 generations, 93% were Ampr and Kanr. The Kan-carrying fragment had presumably inserted into the host chromosome by reciprocal recombination between the chromosomal sequences surrounding Kan on pKA20 and their chromosomal homologs (19). Several such Kan+ derivatives of strain DC679 were purified and crossed with P1 grown on SG20253, which carries the zba::Tn10 insertion close to thdA. No cotransduction was observed between zba::Tn10 and Kan. We also tried Tn10 insertions highly linked to thdC and thdD and again found no cotransduction.

It would appear that pKA10 carries a gene which gives a thiophene-positive phenotype, at least when present in multiple copies, but which is distinct from thdA, -C, or -D. To map this novel gene, designated thdf, we grew P1 on NAR967, a thdf::Kan derivative of DC679, and transduced the thdf::Kan insertion into HfrC. The thdf::Kan derivative of HfrC was conjugated with JC1552, a multiply marked, streptomycin-resistant, F− strain. Recombinants were selected for Strr and Kanr and tested for retention of the auxotrophic markers of JC1552. The results showed that thdf lay closest to metB at 89 min on the chromosome (data not shown). We then used P1 to transduce Tn10 insertions in this neighborhood of the chromosome into NAR967 thdf::Kan. The thdf::Kan insertion showed 27% cotransduction with zic::Tn10 of strain 18404, 16% cotransduction with ilv::Tn10 of strain BW6159, and 30% cotransduction with the ilv mutation of strain JRG997. No cotransduction was observed between thdf::Kan and zic::Tn10 (strain NAR955), metE::Tn10 (strain NAR968), or zyg::Tn10 (strain DF968). The results indicated a location close to the ilvG cluster at 84.8 min on the chromosome (4). It is clear from this that the thdf gene on pKA10 is quite distinct from the thdA gene which maps at 11 min (1, 9).

**Sequence of thdf region.** DNA of plasmid pKA10 was isolated and digested with suitable restriction endonucleases, and the fragments were ligated into correspondingly digested M13mp18 or M13mp19 (24). Both strands were sequenced by the method of Sanger et al. (31). The 3.8-kb chromosomal insert of pKA10 proved to be 3.774 bp long, and part of this sequence is shown in Fig. 4. When this sequence was compared with known sequences present in GenBank, we found an overlap of approximately 400 bp with the tryptophanase (tmaA) gene of *E. coli*. The PstI site at coordinate 0 of our insert corresponds to the PstI site at

---

**FIG. 2.** Electrophoresis of proteins. Cultures of *E. coli* NAR710 carrying pUC19, pKA10, and derivatives were grown overnight in LB containing ampicillin. Cells were lysed in sample buffer and electrophoresed on an SDS-12.5% polyacrylamide gel. Protein bands were visualized by Coomassie blue staining. Lane 5, molecular size markers: phosphorylase B, 97.4 kDa; bovine serum albumin, 66.2 kDa; ovalbumin, 42.7 kDa; carbonic anhydrase, 31.0 kDa; soybean trypsin inhibitor, 21.5 kDa; lysozyme, 14.0 kDa. Lanes: 1, pUC19; 2, pKA10; 3, pKA30; 4, pKA20; 6, pKA39; 7, pKA37; 8, pKA35; 9, pKA15; 10, NAR710 with no plasmid. Band a, 48-kDa thdf protein; band b, 30-kDa protein; band c, 29-kDa npt protein (neomycin-kanamycin phosphotransferase); band d, 26-kDa protein.

**FIG. 3.** Protein analysis by maxicell technique. Plasmid-encoded proteins were labeled with [35S]methionine as described in the Materials and Methods. Lanes: 1, NAR710 (pKA10) labeled for 24 h; 2, NAR710 (pKA10) labeled for 90 min; 3, NAR710 (pUC19) labeled for 90 min; 4, NAR710 (pKA20) labeled for 90 min. Band a, 48-kDa thdf protein; band b, β-lactamase (31 kDa); band c, npt protein (29 kDa); band d, 26-kDa protein.
coordinate 1850 (approximately) on the restriction map of Deeley and Yanofsky (10) and lies about one-third of the way through the tnaA gene. Since tnaA lies close to 84 min on the genetic map (4), this confirms our cotransductional mapping and also serves to locate our 3.8 kb of DNA on the physical map of Kohara et al. (20).

We found a reading frame, of the correct size to encode a protein of approximately 48 kDa, centered about the BgII site (approximately coordinate 1700 bp) used to generate the thdF::Kan insertion of pKA20. The presumed thdF reading frame runs from coordinate 1031 to 2346 and is the gene next to tnaA but is transcribed in the opposite direction (Fig. 4). No homology to any sequences in GenBank was found with the thdF region or the other portion of the chromosomal insert of pKA10, except for the overlap with tnaA already mentioned. No open reading frame of sufficient size to encode a 30-kDa protein (i.e., band b) was found in the 2250- to 3550-bp region of the insert. The sequence of this latter region has been omitted from Fig. 4.

FIG. 4. Sequence of thdF and intragenic region. The sequence of the whole chromosomal insert of pKA10 is shown from the start codon of the tnaA gene (10) through the intragenic region and the thdF open reading frame (coordinates 1031 to 2346). The putative stem-and-loop structure (coordinates 840 to 889) is underlined.
conveyed the phenotype of a thdA mutation upon a recA derivative of the wild-type strain DC625 (2). However, although there is only one thdA gene, the two clones carried distinct DNA segments.

In this study, we demonstrated that the chromosomal insert of pKA10 derives from the 84-min region of the E. coli chromosome and that it encodes a novel gene designated thdF. Plasmids carrying an intact thdF gene confer a thiophene oxidation-positive phenotype, whereas disruption of thdF by inserting a kanamycin cassette at a unique BglII site abolishes the thiophene phenotype. A protein of approximately 48 kDa was observed both in standard cell extracts and by the maxicell technique. This protein is abolished by disruption of the thdF open reading frame and is of the appropriate size. A 30-kDa protein was also observed in cells carrying pKA10 but was not seen in maxicells nor was there an appropriate open reading frame found upon sequencing the insert of pKA10. We presume this protein is somehow involved in the thiophene response but is chromosomally encoded. A third protein (26 kDa) was encoded by pKA10 and is presumably a truncated version of the tnaA gene product.

Sequence analysis showed an overlap of approximately 400 bp with the tnaA gene (10), serving to locate our segment on the E. coli physical map (20). The thdF open reading frame (coordinates 1031 to 2346) is next to tnaA but transcribed in the opposite direction. Tryptophanase converts tryptophan to indole, releasing pyruvate and ammonia (33). Conceivably, thdF is an evolutionary relic of a disused pathway for the further degradation of indole. Perhaps the thdF gene has been recruited for the breakdown of sulfur rather than nitrogen heterocycles in our thd mutants. Note
that their thd mutants do not grow on indole nor do they show significant oxidation of indole itself or derivatives such as indole-2-carboxylate or indole-3-acetate when assessed in tetrazolium indicator medium (data not shown). However, the thiophene-positive strains were more sensitive to indole toxicity than wild-type strains, suggesting possible partial metabolism of indole to some toxic intermediates.

The intercistronic region between tnaA and thdF contains the regulatory elements for tnaA (10). A putative Shine-Dalgarno sequence, TAAAGGG (consensus sequence TAG GAGG), for thdF can be seen at coordinates 1016 to 1022. It also contains a possible stem-and-loop structure with a 20-bp stem, located at coordinates 840 to 889, approximately halfway between tnaA and thdF. We found that both the thdF protein and the 30-kDa protein, as observed on polyacrylamide gels, are much more prominent in cultures grown into the early stationary phase. Moreover, we have recently shown that oxidation activity toward thiophene derivatives in our E. coli strains is also greatest in the early stationary phase (16). Exponential cultures, or those left in the stationary phase for more than 24 h, show little activity (16). It has recently been demonstrated that the nah operon of some naphthalene-degrading Pseudomonas strains shows a similar regulatory pattern (8). The regulatory significance of the putative stem-and-loop structure and the possible evolutionary relationship of thdF to other aromatic degradative systems requires further investigation.

ACKNOWLEDGMENTS

This work was supported by contracts DE-FG22-87PC79912 and DE-FC01-83FE06339 from the U.S. Department of Energy.

REFERENCES