Mol Gen Genet (1989) 218:516–522

MGG

Identification and DNA sequence of *tdcR*, a positive regulatory gene of the *tdc* operon of *Escherichia coli*

Herbert P. Schweizer and Prasanta Datta
Department of Biological Chemistry, University of Michigan, Medical School, Ann Arbor, MI 48109-0606, USA

Summary. Efficient in vivo expression of the biodegradative threonine dehydratase (*tdc*) operon of *Escherichia coli* is dependent on a regulatory gene, *tdcR*. The *tdcR* gene is located 198 base pairs upstream of the *tdc* operon and is transcribed divergently from this operon. The nucleotide sequence of *tdcR* and two unrelated reading frames has been determined. The deduced amino acid sequence of TdcR indicates that it is a polypeptide of M_r 12000 with 99 amino acid residues and contains a potential helix-turn-helix DNA binding motif. Deletion analysis and minicell expression of the *tdcR* gene suggest that TdcR may serve as a *trans*-acting positive activator for the *tdc* operon.

Key words: Gene regulation – *tdc* operon – *Escherichia coli* – *tdc* activator

Introduction

The structural gene (*tdcB*) for the anerobically induced biodegradative threonine dehydratase (EC 4.2.1.16) of *Escherichia coli* has previously been cloned in this laboratory on a 6.3 kilobase pair (kb) *Eco*RI fragment (Goss and Datta 1985). Subcloning and DNA sequence and insertion analysis revealed that *tdcB* is part of a single transcriptional unit, the *tdc* operon, that contains three genes: *tdcA*, *tdcB* and *tdcC* (Goss et al. 1988). In minicells, these genes directed the synthesis of polypeptides with apparent molecular weights of 32, 35 and 45 kilodaltons (kDa), respectively. Whereas the 35 kDa protein was identified as threonine dehydratase (Goss and Datta 1985; Datta et al. 1987), the functions of the 32 kDa and 45 kDa proteins still remain unknown. In an earlier report, Hobert and Datta (1983) showed that anaerobic expression of the *tdcB* gene has an absolute requirement for four amino acids, threonine, serine, valine and isoleucine, plus an electron acceptor, fumarate, and cyclic AMP. They postulated that some metabolite(s) derived from these amino acids during anaerobic metabolism is needed for dehydratase induction. In addition, the enzyme is subject to catabolite repression control by rapidly utilizable carbohydrates (Umbarger 1978; Hobert and Datta 1983). In this report we present evidence that efficient expression of the *tdc* operon is positively regulated by a gene located upstream of the *tdc* promoter.

Offprint requests to: P. Datta

Materials and methods

Bacterial strains. Strain DH5αF′ (F′ *ϕ80d lacZΔM15 Δ(lacZYA-argF)U169 recA1 endA1 hsdR17* ($r_k^- m_k^+$) *supE44* λ^- *thi-1 gyrA relA1* (Bethesda Research Laboratories, Gaithersburg, Md) was used as a host for selection and maintenance of plasmids. Strain X1411 (F$^-$ *glnU42 minA1 minB1* λ^-) (Clark-Curtiss and Curtiss 1983) was used for isolation of minicells. TdcB-LacZ fusions were assayed in strain DL291 (F$^-$ *Δ(argF-lac)U169 araD139 rpsL150 relA1 thiA1 deoC1 ptsF25 flbB5301 rbsR glpR2 Δ(glpT-glpA)593 gyrA recA1)* (Ludtke et al. 1982).

Growth media. Media were prepared according to Miller (1972) and Davis et al. (1980). If needed, antibiotics were added at the following concentrations: 100 μg/ml ampicillin (Ap), and 50 μg/ml kanamycin (Km). Lactose phenotypes were screened on LB plates (Miller 1972) containing 40 μg of 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (XGal) per ml.

Plasmids. Plasmid pSH207 was constructed by deleting a 3.7 kb *Sal*I fragment of pEC61 (Fig. 1; Goss and Datta 1985) and contains 2.6 kb of chromosomal DNA. For the construction of the fusion plasmid pFR6, a 303 bp *Dra*I fragment of pSH207 containing the first 66 codons of the *tdcR* gene was ligated to *Sma*I cleaved pMC1403 DNA (Casadaban et al. 1980). This procedure fused codon 66 of *tdcR* in-frame to codon 8 of *lacZ*. For the construction of the fusion plasmid pFR1, a 368 bp *Hinc*II fragment of plasmid pSH266, harboring the *tdcR* gene on a 600 bp *Dra*I-*Eco*RI fragment from pBal4 in pUC19, was ligated to *Sma*I cleaved pMC1403 DNA. This procedure fused codon 82 of the *tdcR* gene in-frame to codon 8 of *lacZ*. The correct fusion constructs were verified by DNA sequence analysis (data not shown). pBal4, pBal9, pBal11 and pTG122 were from a previous collection of nested deletions (Goss and Datta 1985; T.J. Goss, PhD Thesis, University of Michigan 1986) generated by *Bal*31 treatment of the *Eco*RI terminus of pTG4B. The purified *Eco*RI-*Hin*dIII fragments from pTG4B and these deletion plasmids, carrying decreasing lengths of sequences upstream of the *Hin*dIII site, were ligated separately to the *Hin*dIII site of pSH202 to generate plasmids pSH238, pSH242, pSH241 and pSH240 (Fig. 1). The construction of pSH202 harboring the *tdcB-lacZ* fusion has been described previously (Schweizer and Datta 1988). pSH239 was constructed by

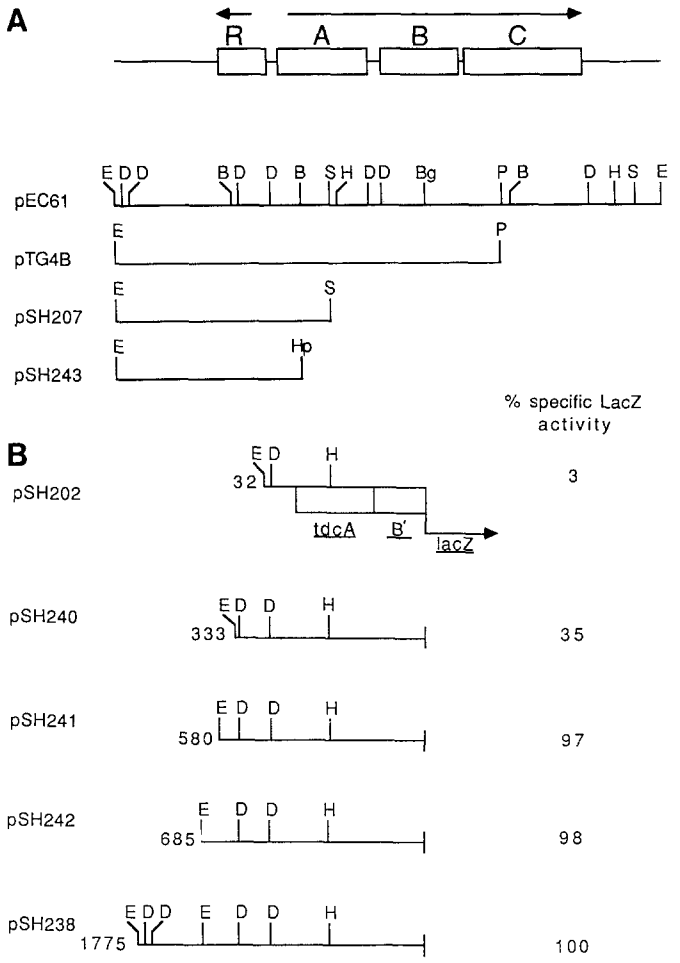

Fig. 1 A and B. Physical structure of the *tdc* region and the effect of upstream sequences on *tdc* expression. **A** Restriction maps of plasmids discussed in the text. Abbreviations: B, *Bcl*I; Bg, *Bgl*II; D, *Dra*I; E, *Eco*RI; H, *Hin*dIII; Hp, *Hpa*I; P, *Pst*I; S, *Sal*I. **B** Effect of upstream sequence on *tdc* expression. A nested set of deletions extending from the *Eco*RI (E) site of pTG4B was obtained by *Bal*31 treatment. The deletion end-points were determined by DNA sequence analysis (see Fig. 3). The numbers next to the *Eco*RI sites of various plasmids represent distances in bp from the first nucleotide of the *Dra*I site upstream of *tdcA*. β-Galactosidase activity from the indicated plasmids was measured in strain DL291 (*Δlac*) grown aerobically in LB medium and β-galactosidase activity expressed from pSH238 was arbitrarily set at 100%

ligating a 3.7 kb *Eco*RI-*Pvu*II fragment from pSH207, containing 2.3 kb of the upstream sequence plus 1.4 kb pBR322 DNA, with pACYC184 (Chang and Cohen 1978) previously digested with *Eco*RI and *Pvu*II.

Plasmid pSH243 (Fig. 1) was constructed by cloning a 2.0 kb *Eco*RI-*Hpa*I fragment of pSH207 into *Eco*RI-*Sma*I cleaved pUC13 (Vieira and Messing 1982) DNA. For the construction of pSH245, pSH243 was first passed through *dam*$^-$ strain GM33 (G. Marinus via Bethesda Research Laboratories), cleaved with *Bcl*I, and then ligated to a 1.4 kb *Bam*HI kanamycin-resistance gene cassette isolated from pUC4K (Vieira and Messing 1982). The transcriptional orientation of the kanamycin-resistance gene in pSH245 is opposite to that of the *tdc* operon.

DNA sequence analysis. DNA fragments were isolated from low-melting agarose gels as described by Wieslander (1979). Cloning of DNA fragments into M13mp18 and M13mp19 and isolation of single stranded template DNA were carried out as described by Messing (1983). Sequence analysis by the dideoxy chain termination procedure was performed utilizing the Sequenase™ kit and the protocol provided by the supplier (United States Biochemicals, Cleveland, Ohio). Labeled DNA fragments were separated in buffer gradient gels as described by Biggin et al. (1983).

Purification and labeling of minicells. Minicells were purified, labeled and analyzed as previously described (Goss et al. 1988). In some experiments, protein patterns were analyzed on tricine-sodium dodecylsulfate-polyacrylamide gels as described by Schägger and von Jagow (1987).

Other methods. Restriction enzymes, Klenow DNA polymerase and T4 DNA ligase were purchased from Bethesda Research Laboratories (Rockville, Md), Boehringer Mannheim Biochemicals (Indianapolis, Ind) or International Biotechnologies (New Haven, Conn) and were utilized according to the supplier's specifications. DNA sequences were analyzed using the Beckman MicroGenie software (Beckman Instruments, Palo Alto, Calif).

Results

Effect of upstream sequences on tdc *expression*

While studying the extent of DNA needed for *tdcB* expression, Goss and Datta (1985) found that 159 base pairs (bp) of nucleotide sequence upstream of the first codon of *tdcA* present in pTG122 was sufficient for dehydratase synthesis. However, pTG122 expressed only approximately 27% of the activity expressed from pTG4B, a plasmid containing ca. 1.9 kb of upstream DNA (Fig. 1). This result indicates that upstream sequences might be required for efficient expression of the *tdc* genes. To examine in detail the effect of upstream sequences on *tdc* expression, a set of nested deletions extending from the *Eco*RI end of pTG4B was fused with *lacZ* as described in Materials and methods, and *tdc* expression was monitored by measuring the *tdcB-lacZ* encoded β-galactosidase activity. The results shown in Fig. 1 clearly demonstrate that DNA sequences located within ca. 700 bp upstream of the *tdcA* coding region are required for optimal *tdc* expression.

A careful comparison of β-galactosidase activity expressed from pSH202, pSH240 and pSH241 revealed an interesting aspect of *tdcB-lacZ* expression: plasmid pSH240, which contained an extra 300 bp of DNA upstream from the end of pSH202, exhibited significantly higher levels of β-galactosidase activity as compared to those produced by pSH202 but failed to express enzyme activity maximally as that seen with pSH241; the latter plasmid contained an additional segment of upstream DNA of about 250 bp (Fig. 1). One possible interpretation of the data is that pSH241, but not pSH240, encodes a specific protein which is necessary for optimal expression of *tdc*, and that pSH240 contains a DNA sequence that is needed in *cis* for transcriptional activation; pSH202 lacks this *cis* sequence, resulting in basal levels of enzyme expression. The intermediate level of *tdc-lacZ* expression in pSH240 would presumably reflect transcriptional activation mediated by the chromosomally-encoded activator protein.

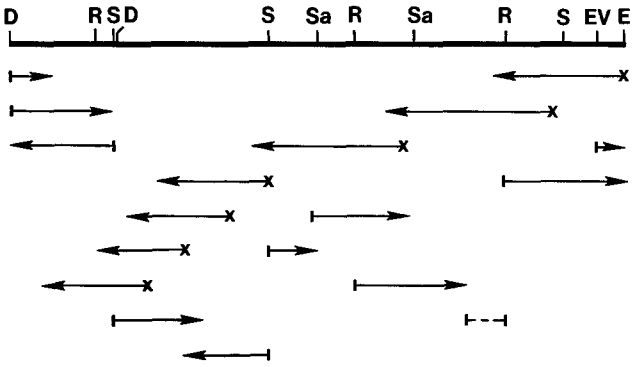

Fig. 2. Sequencing strategy. Sequence information was obtained from clones that were constructed by: (1) cloning fragments derived from a nested set of *Bal*31 deletions into M13mp19 (marked x); and (2) directed cloning of fragments into M13mp18 or M13mp19 (marked I). Both strands were sequenced with the exception of ca. 100 bp (marked – – –) which was sequenced twice on the same strand. Abbreviations: D, *Dra*I; E, *Eco*RI; EV, *Eco*RV; R, *Rsa*I; S, *Ssp*I; Sa, *Sau*3AI

To test this notion, strain DL291 harboring pSH202 or pSH240 was transformed with the compatible plasmid pSH239 containing the entire 1.9 kb upstream sequences. *tdcB-lacZ* expression from pSH240 was stimulated 3- to 5-fold by the presence of pSH239, whereas expression from pSH202 was unaffected (data not shown). These results provide strong evidence that a specific protein might be encoded by the upstream sequences which is acting in *trans* as a transcriptional activator at the *tdc* promoter. The results further suggest that a potential *cis*-acting sequence(s) necessary for transcriptional activation is present on pSH240 but not on pSH202. From the data presented in Fig. 1 it is evident that the structural gene for the putative *tdc* activator protein (designated *tdcR*) must be located between the left boundary of pSH202 and the left boundary of pSH241 (as depicted in Fig. 1), encompassing ca. 600 bp of DNA.

Nucleotide sequence of the tdcR *region*

Figure 2 outlines the strategy used to determine the nucleotide sequence of the entire upstream region (displayed in inverted orientation) between the *Dra*I site located 130 bp upstream of the *tdcA* coding region and the *Eco*RI site representing the left-most boundary of the cloned chromosomal DNA. Both strands were sequenced with the exception of ca. 100 bp which were sequenced twice on the same strand. All restriction sites indicated in Fig. 2 were verified by restriction enzyme digestion except a *Hpa*I site at nucleotide 721 (Fig. 3; not shown in Fig. 2) that could only be verified by nucleotide sequence analysis. The nucleotide sequence of the coding strand, presented in Fig. 3, revealed two complete open reading frames (*orf*) corresponding to nucleotides 106–402 and 661–1197, and a third partial *orf* starting from nucleotide 1222 (numbered from the *Dra*I site at the 5′ end). No other open reading frames were detected in the same or on the opposite strand. Based on the extent of *tdc-lacZ* expression from various plasmids (see Fig. 1), the 297 bp long first *orf* most likely represents the putative *tdcR* gene product: plasmid pSH241, which extends up to nucleotide 580, harbors the complete coding

Fig. 3. Nucleotide sequence of the upstream region. The nucleotide number is assigned arbitrarily, beginning with the first nucleotide of the *Dra*I site located upstream of *tdcA*. Putative promoter consensus sequences (−10 and −35) and ribosome binding sites (SD) of the open reading frames *tdcR*, *orfA* and *orfB* are indicated. An *asterisk* below the line marks a repeated sequence with the consensus AGPyATAAA. The *arrowhead* marks the fusion joints of the TdcR-LacZ fusion proteins. Also indicated is a possible weak transcriptional terminator for *tdcR*

Table 1. Codon usage in *tdcR*[a]

Amino acid	Codon	*tdcR*	S	W
Ala	GCU	0	33	17
	GCC	0	9	34
	GCA	2	23	20
	GCG	1	25	28
Arg	CGU	1	42	19
	CGC	1	19	25
	CGA	0	1	5
	CGG	0	0.2	8
	AGA	1	1	5
	AGG	1	0.2	3
Asn	AAU	9	2	19
	AAC	3	30	19
Asp	GAU	2	22	35
	GAC	1	39	20
Cys	UGU	3	2	6
	UGC	1	4	7
Gln	CAA	3	7	17
	CAG	1	32	32
Glu	GAA	2	63	40
	GAG	2	20	19
Gly	GGU	0	43	24
	GGC	0	33	27
	GGA	0	1	8
	GGG	1	3	13
His	CAU	2	4	18
	CAC	1	14	11
Ile	AUU	5	13	30
	AUC	0	15	23
	AUA	2	0.4	5
Leu	UUA	4	2	14
	UUG	3	3	12
	CUU	0	5	14
	CUC	1	6	13
	CUA	0	1	4
	CUG	1	66	56
Lys	AAA	5	49	31
	AAG	3	20	8
Met	AUG	1	27	25
Phe	UUU	9	7	29
	UUC	3	22	19
Pro	CCU	0	4	6
	CCC	0	0.4	9
	CCA	3	5	9
	CCG	1	31	19
Ser	UCU	2	18	7
	UCC	0	17	9
	UCA	4	1	7
	UCG	1	2	12
	AGU	0	2	11
	AGC	1	9	12
Thr	ACU	0	20	9
	ACC	0	26	23
	ACA	2	3	6
	ACG	1	5	15
Trp	UGG	0	5	13
Tyr	UAU	4	6	18
	UAC	1	19	12
Val	GUU	1	37	21
	GUC	1	8	13
	GUA	0	23	9
	GUG	2	16	24

[a] Codon usage in *tdcR* is compared with that in strongly (S) and weakly (W) expressed *Escherichia coli* genes (tabulated by Grosjean and Fiers 1982). The *tdcR* data are for a total of 99 codons; the Grosjean and Fiers data are for a relative usage per 1000 codons

sequence of the polypeptide and shows maximum expression of β-galactosidase, whereas plasmid pSH202 lacks the coding sequence and shows only basal levels of LacZ activity. The 73 bp untranslated region between the end of pSH202 and the start of the putative *tdcR* polypeptide could easily accommodate a potential *cis*-site for transcription activation. As mentioned above, this sequence present in pSH240 may stimulate transcription to a significant level by interacting with the activator polypeptide encoded by the chromosomal *tdcR* gene.

The results displayed in Figs. 1 and 3 clearly indicate that the *tdcR* open reading frame is oriented in the opposite direction to that of the *tdc* genes. The region preceding the *tdcR* coding sequence contains recognizable −10 and −35 consensus *E. coli* promoter sequences, and a weak Shine-Dalgarno translation initiation sequence precedes the initiating GUG codon (Fig. 3). A stretch of 258 nucleotides separates *tdcR* from *orfA; orfB* is located 24 bp downstream from the 3′ end of *orfA* and extends beyond the *Eco*RI site. From the sequence data, both *orfA* and *orfB* are preceded by strong Shine-Dalgarno sequences, however, only *orfA* is preceded by a sequence typical of *E. coli* promoters. It is conceivable that *orfA* and *orfB*, but not *tdcR*, are part of a single transcriptional unit (see below).

Expression of cloned DNA in minicells

The deduced amino acid sequences of the three reading frames reveal that the putative *tdcR* polypeptide consists of 99 amino acid residues with a calculated M_r of 11816; in addition, *orfA* and *orfB* encode polypeptides of M_r 20579 and 20649, respectively, the latter representing a truncated molecule originating from a larger protein of unknown molecular size. An analysis of the *tdcR* polypeptide indicates that the codon usage of *tdcR* is relatively random, and a comparison with the codon preference statistics of weakly and strongly expressed *E. coli* genes, compiled by Grosjean and Fiers (1982), suggests that *tdcR* would have a low level of expression (Table 1). In accordance with these predictions, minicell analyses of protein patterns expressed by plasmids harboring the upstream region revealed three polypeptides: a polypeptide of approximately 22 kDa, most likely representing *orfA*; a weakly expressed 21 kDa polypeptide presumably originating from the truncated *orfB*; and a very weak, barely visible polypeptide of about 12 kDa approximating the predicted size of the *tdcR* gene product (Fig. 4A, lane 2). Interruption of transcription at the single *Bcl*I site located within *tdcR* (Fig. 3) by insertion of a 1.4 kb kanamycin-resistance gene cassette from

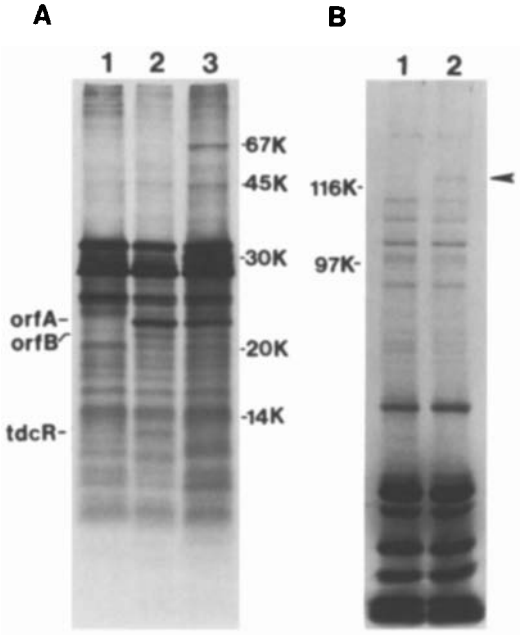

Fig. 4A and B. Expression of proteins in minicells. **A** Minicells harboring the indicated plasmids were labeled with [^{35}S]methionine and proteins were separated in a 16%T 3%C SDS-polyacrylamide gel as described by Schägger and Jagow (1987). Lane 1, X1411/pUC12; lane 2, X1411/pSH243; lane 3, X1411/pSH245. Numbers to the right indicate the molecular masses in kDa (K) of the size markers. The positions of plasmid-encoded proteins are indicated to the left. **B** Minicells harboring the indicated plasmids were labeled with [^{35}S]methionine and proteins were separated in a 8% SDS-polyacrylamide gel. Lane 1, X1411/pMC1403; lane 2, X1411/pFR6. The *arrowhead* indicates the position of the TdcR-LacZ fusion protein. Numbers to the left indicate the molecular masses in kDa (K) of the size markers (116K marks the position of *Escherichia coli* β-galactosidase)

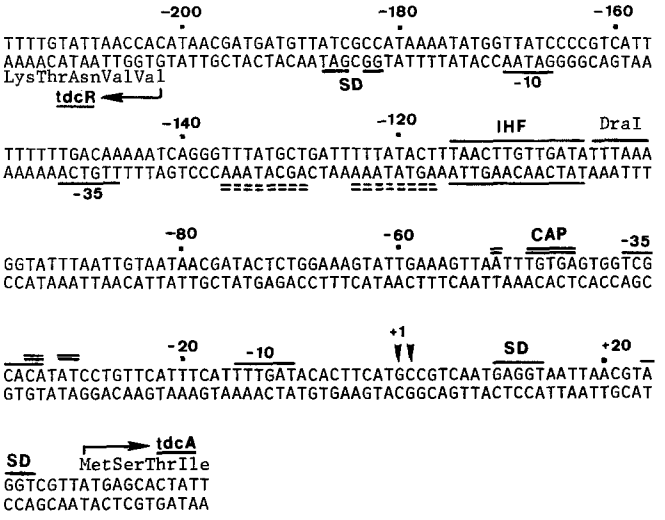

Fig. 5. Structure of the *tdc* regulatory region. The location of several regulatory sequences known to affect *tdc* expression are shown. Nucleotides are numbered relative to their positions to the major *tdc* mRNA start site (+1). The beginning of the *tdcR* and *tdcA* coding regions are indicated. Abbreviations: CAP, catabolite gene activator protein binding site; IHF, integration host factor binding site; = =, AGPyATAAA consensus sequence. *Arrowheads* denote transcription start sites of *tdc* mRNA. The putative −35 and −10 sequences, and the ribosome binding sites (SD) are indicated

pUC4K (Vieira and Messing 1982) abolished the synthesis of the 12 kDa polypeptide but had no effect on the synthesis of the 21 kDa and 22 kDa polypeptides (Fig. 4A, lane 3). These results suggest that the three open reading frames are expressed in vivo, and that *tdcR* and *orfA-orfB* belong to different transcriptional units. The functions of the *orfA* and *orfB* open reading frames are not known.

As mentioned above, the weak expression of *tdcR* presumably reflects the poor Shine-Dalgarno sequence as well as poor codon usage. Furthermore, regulatory molecules exist only in a few copies per cell. Another experimental strategy was employed further to examine *tdcR* expression. Two *tdcR-lacZ* protein fusions (pFR1 and pFR6) were constructed as described in Materials and methods. As judged by growth on LB+XGal plates, both fusions directed the synthesis of low but clearly detectable levels of β-galactosidase. Plasmid pFR6 was chosen for further analysis. In a *Δlac* strain it directed the synthesis of about 80 units of β-galactosidase. It also synthesized in minicells a small amount of a fusion protein of about 125 kDa (Fig. 4B, lane 2) in excellent agreement with the predicted size of 124 kDa. In SDS-polyacrylamide gels, the 125 kDa polypeptide co-migrated with affinity-purified TdcR-LacZ protein (data not shown). This result provides further evidence for in vivo expression of *tdcR* and for the correctness of the assigned reading frame of the *tdcR* gene.

Discussion

The experiments described above provide evidence for a regulatory gene, *tdcR*, required for efficient expression of the *tdc* operon. Nucleotide sequence of the cloned *tdc* DNA and deletion studies with *tdcB-lacZ* fusion plasmids clearly indicate that *tdcR* is located 232 base pairs upstream of the *tdcA* gene and is transcribed in opposite orientation to that found for the *tdc* operon. The amino acid sequence of TdcR, deduced from the nucleotide sequence of the *tdcR* gene, shows it to be a small polypeptide consisting of 99 amino acid residues with a calculated M_r of 11816. The relatively weak expression of TdcR is generally consistent with the presence of a poor Shine-Dalgarno sequence preceding the GUG codon for translation initiation, and the random codon usage statistic correlating with weakly expressed *E. coli* genes. These results are reminiscent of the D-serine deaminase activator gene, *dsdC*, encoding a polypeptide of 33 kDa, needed for transcriptional activation of the *dsdA* gene of *E. coli*. It is interesting to note that *dsdC* is located 731 base pairs upstream of the translation initiation codon of *dsdA* and is transcribed divergently from *dsdA* (Palchaudhuri et al. 1988).

Previous reports from this laboratory suggested that efficient expression of the *tdc* operon is dependent on the anaerobic metabolism of the four amino acids threonine, serine, valine and isoleucine, in the presence of an electron acceptor such as fumarate or nitrate, and requires cAMP (Hobert and Datta 1983). Although the existence of a consensus CAP binding site on the *tdc* DNA between *tdcR* and *tdcA* (Fig. 5; Goss and Datta 1985), and the in vitro gel-shift and transcription experiments (R.V. Patil and P. Datta, unpublished observations) have confirmed the role of cAMP for *tdc* gene expression, the nature of the real inducer generated during anaerobic metabolism remains to be established. If TdcR were to function as a positive activa-

Fig. 6. Sequence of a 19 residue TdcR polypeptide and its relationship to the sequence of other DNA binding proteins. Below the sequence of TdcR (residues 43–61) are the carboxy-terminal residues from 74–98 of Fis and from 442–470 of NtrC from *Klebsiella pneumoniae* (Johnson et al. 1988). The helix-turn-helix regions from TnpR, Cro and Hin are taken from Pabo and Sauer (1984). *Dots* indicate identical amino acids, and the *dash* designates the gap used to maximize alignment

tor protein, it must, according to the accepted view, bind the inducer, and the protein-inducer complex must subsequently recognize and bind to specific DNA sequence(s) to facilitate transcription. Most prokaryotic transcription-regulatory proteins examined thus far interact with their respective target DNAs by a helix-turn-helix motif (Pabo and Sauer 1984; Kelley and Yanofsky 1985). Analysis of the TdcR sequence according to the method of Dodd and Egan (1987), developed for the detection of potential DNA-binding regions in proteins, reveals a polypeptide segment (amino acid residues 43–61) that shows significant homology to helix-turn-helix motifs of other DNA binding proteins (Fig. 6). TdcR has the appropriate amino acids, alanine and leucine, at the two most conserved residues (positions 5 and 15) although not at the first hinge amino acid residue (at position 9) which is usually a glycine; in addition, residues serine and arginine occur at high frequencies at positions 11 and 12, respectively, and a nonpolar residue at position 10 (Pabo and Sauer 1984; Dodd and Egan 1987). The entire 20-amino acid region of TdcR displays 25% and 35% identity to the NtrC and Fis polypeptides which activate, respectively, σ^{54}-dependent promoters associated with nitrogen regulation (Buikema et al. 1985; Drummond et al. 1986) and Hin-mediated DNA inversion in vivo and in vitro (Johnson et al. 1988). Thus, it appears likely that TdcR could potentially function as a DNA binding protein and, in association with the putative inducer, activate *tdc* transcription.

The putative binding site(s) on *tdc* DNA for TdcR has not been identified. The untranslated DNA segment between the *tdcR* and *tdcA* reading frames (see Fig. 5) contains sequences implicated in *tdc* gene expression: the CAP site for transcription initiation by the cAMP-CAP complex which is inhibited by the *E. coli* integration host factor (IHF), which binds to the DNA at the IHF site (R.V. Patil and P. Datta, unpublished observations). Experiments with the *tdcB-lacZ* plasmids (Figs. 1, 3) suggested that this region of DNA might also include a *cis*-sequence needed for the action of the *tdcR* gene product. The two direct repeat elements, with the consensus sequence 5′-AGPyATAAA-3′ at nucleotides −117/−124 and −129/−136 (Fig. 5), are possible candidates for binding of TdcR. This notion is consistent with the observations that the endpoint of the plasmid pSH202 (Fig. 1), which expresses basal level of β-galactosidase activity in the presence or absence of the *tdcR* gene product, maps in between the two direct repeats (G at position −128); whereas plasmid pSH240, which includes these direct repeats but lacks a functional *tdcR* gene product, exhibits low TdcB-LacZ activity in the absence of plasmid-encoded TdcR but shows maximal activity in the presence of TdcR.

The results clearly indicate a genetic role of *tdcR* for efficient expression of the *tdc* operon. Further experiments are necessary to understand the biochemical mechanism of transcriptional activation by TdcR, identify its DNA binding site, and establish the identity of the physiological inducer.

Acknowledgements. This work was supported by Public Health Service Grant GM21436 from the National Institutes of Health.

References

Biggin MD, Gibson TJ, Hong SF (1983) Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc Natl Acad Sci USA 80:3963–3965

Buikema WJ, Szeto WW, Lemley PV, Orme-Johnson WH, Ausubel FM (1985) Nitrogen fixation specific regulatory genes of *Klebsiella pneumoniae* and *Rhizobium meliloti* share homology with the general nitrogen regulatory gene *ntrC* of *K. pneumoniae*. Nucleic Acids Res 13:4539–4555

Casadaban MJ, Chou J, Cohen SN (1980) In vitro gene fusions that join an enzymatically active β-galactosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for the detection and cloning of translation initiation signals. J Bacteriol 143:971–980

Chang ACY, Cohen SN (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J Bacteriol 134:1141–1156

Clark-Curtiss JE, Curtiss R, III (1983) Analysis of recombinant DNA using *Escherichia coli* minicells. Methods Enzymol 101:347–362

Datta P, Goss TJ, Omnaas JR, Patil RV (1987) Covalent structure of biodegradative threonine dehydratase of *Escherichia coli*: homology with other dehydratases. Proc Natl Acad Sci USA 84:393–397

Davis RW, Botstein D, Roth JR (1980) Advanced Bacterial Genetics. Cold Spring Harbor Laboratory Press, New York

Dodd IB, Egan JB (1987) Systematic method for the detection of potential λ cro-like DNA-binding regions in proteins. J Mol Biol 194:557–564

Drummond M, Whitty P, Wootton J (1986) Sequence and domain relationships of *ntrC* and *nifA* from *Klebsiella pneumoniae:* homologies to other regulatory proteins. EMBO J 5:441–447

Goss TJ, Datta P (1985) Molecular cloning and expression of the biodegradative threonine dehydratase gene (*tdc*) of *Escherichia coli*. Mol Gen Genet 201:308–314

Goss TJ, Schweizer HP, Datta P (1988) Molecular characterization of the *tdc* operon of *Escherichia coli* K-12. J Bacteriol 170:5352–5359

Grosjean H, Fiers W (1982) Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene 18:199–209

Hobert EH, Datta P (1983) Synthesis of biodegradative threonine dehydratase of *Escherichia coli*: role of amino acids, electron acceptors, and certain intermediary metabolites. J Bacteriol 155:586–592

Johnson R, Ball CA, Pfeffer D, Simon MI (1988) Isolation of the gene encoding the Hin recombinational enhancer binding protein. Proc Natl Acad Sci USA 85:3484–3488

Kelley RL, Yanofsky C (1985) Mutational studies with the *trp* repressor of *Escherichia coli* support the helix-turn-helix model of repressor recognition of operator DNA. Proc Natl Acad Sci USA 82:483–487

Ludtke D, Larson TJ, Beck C, Boos W (1982) Only one gene is required for *glpT*-dependent transport of *sn*-glycerol-3-phosphate in *Escherichia coli*. Mol Gen Genet 186:540–547

Messing J (1983) New M13 vectors for cloning. Methods Enzymol 101:20–78

Miller JH (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratory Press, New York

Pabo CO, Sauer RT (1984) Protein DNA recognition. Annu Rev Biochem 53:293–321

Palchaudhuri S, Patel V, McFall E (1988) DNA sequence of the D-serine deaminase activator gene *dsdC*. J Bacteriol 170:330–334

Schägger H, von Jagow G (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem 166:368–379

Schweizer HP, Datta P (1988) Genetic Analysis of the *tdcABC* operon of *Escherichia coli* K-12. J Bacteriol 170:5360–5363

Umbarger HE (1978) Amino acid biosynthesis and its regulation. Annu Rev Biochem 47:533–606

Vieira J, Messing J (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259–268

Wieslander L (1979) A simple method to recover intact high molecular weight RNA and DNA after electrophoretic separation in low gelling temperature agarose gels. Anal Biochem 98:305–309

Communicated by H. Hennecke

Received March 3, 1989