Genetic, Functional and Sequence Analysis of the \textit{xylR} and \textit{xylS} Regulatory Genes of the TOL Plasmid pWW0

By ROBERT A. SPOONER,* KAREN LINDSAY AND F. CHRISTOPHER H. FRANKLIN

Department of Genetics, University of Birmingham, PO Box 363, Birmingham B15 2TT, UK

(Received 23 September 1985; revised 16 December 1985)

Mutant derivatives of a plasmid, pCF20, which carries the \textit{XhoI-D} fragment of the TOL plasmid pWW0 have been isolated using Tn5 transposon mutagenesis. Insertion mutations of the \textit{xylR} and \textit{xylS} regulatory genes of the catabolic pathway have been isolated and characterized and their ability to induce catechol 2,3-oxygenase activity determined. Analysis of the insertion mutants and also segments of the \textit{XhoI-D} fragment cloned into plasmid pUC8 in maxicells has identified a 68 kDa polypeptide product encoded by the \textit{xylR} gene. No clear candidate for the \textit{xylS} polypeptide was observed. The nucleotide sequence of the \textit{xylS} region, the intergenic region and part of the \textit{xylR} region has been determined and open reading frames (ORFs) assigned for both genes. The ORF designated \textit{xylS} appears capable of encoding a polypeptide of \(\sim 37\) kDa.

\textbf{INTRODUCTION}

A set of inducible enzymes specified by genes on the TOL plasmid pWW0 enables \textit{Pseudomonas putida} mt-2 to grow on the aromatic hydrocarbons toluene, \textit{m-} and \textit{p-}xylene, \textit{3-}ethyltoluene, and 1,2,4-trimethylbenzene and their alcohol, aldehyde and carboxylic acid derivatives (Worsey & Williams, 1975; Kunz & Chapman, 1981). The genes encoding these catabolic enzymes are organized into two regulatory units on the plasmid (Franklin et al., 1983). Two regulatory genes, \textit{xylR} and \textit{xylS}, control the expression of the pathway genes, and a model has been proposed for the regulatory action of the \textit{xylR} and \textit{xylS} gene products (Worsey et al., 1978). In this model the \textit{xylR} gene product interacts with the hydrocarbon and alcohol substrates to induce expression of both the operon encoding the early enzymes (the `upper pathway' \textit{xylCAB}) and the operon encoding the later enzymes (the `lower or \textit{meta} pathway' \textit{xylDLEGF}) of the degradative pathway, whereas the \textit{xylS} gene product interacts with the carboxylic acid substrates to induce expression of the lower pathway genes only. Both \textit{xylR} (Franklin & Williams, 1980; Nakazawa \textit{et al.}, 1980) and \textit{xylS} (Inouye \textit{et al.}, 1981; Franklin \textit{et al.}, 1981) appear to be positively acting regulators, and have been mapped within the 6-4 kb \textit{XhoI-D} fragment of the TOL plasmid (Franklin \textit{et al.}, 1983; Inouye \textit{et al.}, 1983).

However, recent evidence has indicated that the model requires some modification, although it remains basically correct. Firstly, it seems that the \textit{xylR} gene product and its co-inducer are insufficient for the induction of the lower pathway, but induction will occur if a functional \textit{xylS} gene product is also present, suggesting that some interaction of these components is involved (Franklin \textit{et al.}, 1983; Inouye \textit{et al.}, 1983). Secondly, there is some evidence that a third regulatory element may exist which interacts specifically with unsubstituted substrates (Franklin \textit{et al.}, 1983). Before they can be incorporated into the regulatory scheme, however, these observations need to be further clarified.

\textit{Abbreviation: ORF, open reading frame.}

0001-2965 © 1986 SGM
In this paper we report the detailed analysis of the DNA segment (XhoI-D fragment) encoding the regulatory genes, the identification of the xylR polypeptide product, and the designation of the xylR and xylS open reading frames.

METHODS

Bacterial strains and vectors. All strains used in this study were Escherichia coli K12 derivatives: SK1592, F' EndA- Gal- hsdR41 HsdM' sbcB15 thi-1 (Kushner, 1978); KT759, thr-1 leu-6 thi-1 supE44 lacY1 Tn5 (a C600 derivative; Franklin et al., 1981); CSR603, recA1 uraA6 phr-1 (Sancar & Rupert, 1978); JM101, SupE-Ti+ Δ(lac-proAB) [F' traD36 ProAB- lacF' lacZAM15]; JM83, Ara- Δ(lac-proAB) Str+ g80 lacZAM15.

Plasmids used were pWW0 (TOL) (Williams & Murray, 1974); pRK2501 (Kahn et al., 1979); pUC8 (Vieira & Messing, 1982); pKT240 (Bagdasarian et al., 1983). M13mp18 and M13mp19 (Norlander et al., 1983) were used as vectors for sequencing.

Tn5 mutagenesis. KT759 was transformed with pCF20 and the resulting Tc₈ transformants were incubated at room temperature for 7 d. Plasmid DNA was extracted (Birnboim & Doly, 1979) and was used to transform SK1592. Tc₈ Km₈ colonies were selected and DNA was prepared, and the positions of Tn5 insertions were determined by restriction endonuclease site mapping.

Enzyme assays and protein determinations. Catechol 2,3-oxygenase activity was assayed and concentration of protein in cell-free extracts was measured according to Murray et al. (1972).

Maxicell preparations. The method of Sancar et al. (1979) was followed, with the following modifications. After UV irradiation, CSR603 and plasmid-containing derivatives were incubated (37 °C, 1 h), d-cycloserine was added to 200 μg ml⁻¹ final concentration and the cells were shaken for 15 h at 37 °C. Samples (2 ml) were washed twice in 5-4%, (w/v) Methionine Assay Medium (Difco) and suspended in 1 ml of the same medium. After incubation (1 h, 37 °C), [³⁵S]Methionine (Amersham) was added according to Sancar et al. (1979). Products were analysed by SDS-PAGE and the gels were treated with Amplify (Amersham) before drying and autoradiography.

DNA sequencing. Restriction fragments to be sequenced were ligated into either M13mp18 or M13mp19 (Norlander et al., 1983; Yanisch-Perron et al., 1985) and were propagated in E. coli JM101. Sequencing procedures were essentially as described by Sanger et al. (1977) and Schreier & Cortese (1979), using 15 base primer (BRL) and deoxyadenosine 5'-α-[³⁵S]thiotriphosphate (Amersham), and were done in 1-5 ml polypropylene tubes rather than capillaries. Buffer gradient gels (Biggin et al., 1983) were used routinely. Areas of doubtful sequence were confirmed by including 25% (v/v) or 50% (v/v) formamide in the urea-acrylamide mix. Sequence data were assembled into a complete sequence with the aid of a computer program developed by Dr Georges Ware at the University of Bristol Medical School, Bristol, UK.

RESULTS

Isolation and characterization of Tn5 insertions in the xylR and xylS genes

To facilitate isolation of Tn5 insertions in xylR and xylS, pCF20 was constructed by cloning the pWW0 XhoI-D fragment into the XhoI site of the broad-host-range vector pRK2501 (see Fig. 1). The plasmid was then transformed into an E. coli K12 C600 strain carrying a chromosomal Tn5 insertion; 38 Tn5 insertion derivatives were then isolated, of which 17 were found to have insertions in the XhoI-D fragment region of pCF20. These were selected for further study.

Fig. 2 shows the positions of the 17 Tn5 insertions as determined by restriction endonuclease digestion analysis, using information from Jorgenson et al. (1979) and Auerwald et al. (1980) to provide accurate sizes of transposon-derived restriction fragments. Earlier studies (Franklin et al., 1983; Inouye et al., 1983) had determined the approximate locations of xylR and xylS on the XhoI-D fragment. Using this information, it was possible to tentatively identify six of the Tn5 mutants as insertions within xylR and eight within xylS. Of these, nine were selected for further study, five from the xylR region and four from the xylS region.

Enzyme induction analysis

To test the biochemical activity of the insertion mutations, pCF22 was constructed by cloning the pWW0 SsrI-D fragment, which specifies the meta pathway genes xylD, xylL and xylE together with their promoter, into pKT240. This vector was chosen as there is no transcription from vector promoters through the SsrI site (Bagdasarian et al., 1983). Each mutant derivative of pCF20 and pCF20 itself was transformed into E. coli SK1592 containing pCF22. Cultures of
TOL regulatory genes xylR and xylS

FIG. 1. Derivation of plasmids constructed in this study. pCF20 is the XhoI-D fragment of pWW0 cloned in the pRK2501 XhoI site. pCF31 and pCF27 are XhoI–HindIII subfragments of pKT570 (Franklin et al., 1983) cloned in pUC8 digested with SalI–HindIII. pCF30 is a XhoI–PstI subfragment of pKT570 cloned in pUC8 digested with SalI–PstI. The positions of xylR and xylS are included for reference. 0, TOL DNA; —, vector DNA; b, BglII; B, BamHI; E, EcoRI; h, HincII; H, HindIII; P, PstI; S, SalI; X, XhoI.

FIG. 2. Physical and genetic map of the XhoI-D fragment of pWW0. (a) pCF20 consists of the pWW0 XhoI-D fragment (■) cloned in pRK2501 (—). The map is compiled from Franklin et al. (1983), Inouye et al. (1983), Kahn et al. (1979) and this study. Symbols as in Fig. 1. (b) Sites of Tn5 insertions defining xylS and xylR (▲) and those outside these genes (∆). The position of xylS and the probable 5'-end of xylR have been determined in this study. Arrows indicate the direction of transcription of these two genes.

these strains were then grown under non-inducing (L-broth) and inducing conditions (L-broth plus m-toluate, benzoate or m-methylbenzyl alcohol) and the levels of catechol 2,3-oxygenase activity measured.

According to the current interpretation of the regulatory model (Worsey et al., 1978; Franklin et al., 1983) the strains carrying xylR insertion mutants should show induced catechol 2,3-oxygenase activity following growth in the presence of m-toluate or benzoate but not after
R. A. SPONNER, K. LINDSAY AND F. C. H. FRANKLIN

Table 1. Catechol 2,3-oxygenase activities in cell extracts of E. coli SK1592(pCF22) carrying pCF20 or transposon mutant derivatives

The host strain was E. coli K12 SK1592(pCF22) (pCF22 = pKT240 + SsrI-D fragment of pWW0; see Results) into which pCF20 or a transposon mutant derivative was transformed (pCF20 = pRK2501 + XhoI-D fragment of pWW0; see Results). Uninduced cell cultures were grown in L-broth, and induced cultures in L-broth containing 5 mM m-toluate, 5 mM benzoate or 5 mM m-methylbenzyl alcohol. Cell extracts were prepared and assays done as described in Methods.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>No inducer</th>
<th>m-Toluate</th>
<th>Benzoate</th>
<th>m-Methylbenzyl alcohol</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCF20</td>
<td>93</td>
<td>800</td>
<td>ND</td>
<td>210</td>
</tr>
<tr>
<td>pCF20::Tn5-45.1</td>
<td>95</td>
<td>460</td>
<td>ND</td>
<td>190</td>
</tr>
<tr>
<td>pCF20::Tn5-7.3</td>
<td>72</td>
<td>300</td>
<td>ND</td>
<td>78</td>
</tr>
<tr>
<td>pCF20::Tn5-41.2</td>
<td>76</td>
<td>390</td>
<td>ND</td>
<td>80</td>
</tr>
<tr>
<td>pCF20::Tn5-30.1</td>
<td>100</td>
<td>330</td>
<td>ND</td>
<td>110</td>
</tr>
<tr>
<td>pCF20::Tn5-2.1</td>
<td>130</td>
<td>560</td>
<td>ND</td>
<td>230</td>
</tr>
<tr>
<td>pCF20</td>
<td>130</td>
<td>1700</td>
<td>780</td>
<td>ND</td>
</tr>
<tr>
<td>pCF20::Tn5-34.2</td>
<td>140</td>
<td>1100</td>
<td>460</td>
<td>ND</td>
</tr>
<tr>
<td>pCF20::Tn5-7.1</td>
<td>190</td>
<td>150</td>
<td>110</td>
<td>ND</td>
</tr>
<tr>
<td>pCF20::Tn5-31.1</td>
<td>120</td>
<td>200</td>
<td>140</td>
<td>ND</td>
</tr>
<tr>
<td>pCF20::Tn5-26.1</td>
<td>140</td>
<td>170</td>
<td>180</td>
<td>ND</td>
</tr>
</tbody>
</table>

ND, Not determined.

induction with m-methylbenzyl alcohol. In strains carrying xylS mutants, catechol 2,3-oxygenase should not be induced by any of the substrates, although, if the proposed third regulatory gene (Franklin et al., 1983) were present on pCF20, induction by benzoate would occur.

Table 1 shows the catechol 2,3-oxygenase activities of derivatives of SK1592(pCF22) containing the nine pCF20 insertion mutants selected. Of those containing the five putative xylR mutants, three, namely pCF20::Tn5-7.3, -41.2 and -30.1, show induction with m-toluate but not with m-methylbenzyl alcohol, which is the predicted phenotype of xylR mutants. The others, pCF20::Tn5-45.1 and -2.1 show induction with both substrates and therefore are not insertions within the xylR gene. Of the strains containing pCF20::Tn5 insertions that map around the xylS region, three, pCF20::Tn5-7.1, -31.1 and -26.1 show no induction by either m-toluate or benzoate, consistent with the expected phenotype of xylS mutants. pCF20::Tn5-34.2 permits induction with benzoate and m-toluate and thus is not an xylS mutant.

Identification of the xylR polypeptide products using E. coli maxicells

To attempt identification of the xylR gene product, the pCF20::Tn5 derivatives described above, pCF20 and the vector pRK2501 were transformed separately into the E. coli maxicell strain CSR603. In addition, two pUC8 derivatives were constructed, pCF31 specifying the entire xylR coding region, and pCF30 in which a segment of the xylR gene has been deleted (see Fig. 1). These, and pUC8, were also introduced into CSR603. Maxicells were prepared from these CSR603 derivatives and were incubated with [35S]methionine. The plasmid-specified polypeptide products were analysed by SDS-PAGE and autoradiography.

Comparison of the products from pUC8, pCF30 and pCF31 (Fig. 3) reveals the presence in the latter of a 68 kDa polypeptide which is absent in the other tracks. This polypeptide is also produced by maxicells containing pCF20 and the xylR+ derivatives pCF20::Tn5-2.1 and -45.1, but is absent in those containing pCF20::Tn5-7.3, -30.1 and -41.2, leading to the conclusion that this is indeed the xylR gene product.

Similar experiments designed to identify the xylS gene product failed to produce a visible candidate on the maxicell gels. One possible explanation is that the polypeptide co-migrates with one of the other polypeptides specified by the plasmids analysed. Deletion of the neomycin
Fig. 3. Autoradiograms of SDS-PAGE (11%, w/v; 55:1, acrylamide:NN'-methylene bisacrylamide) of [35S]methionine-labelled maxicells. Maxicells were prepared as described in Methods. Mₐ values were determined by comparison of mobilities with protein molecular weight standards (BRL, high molecular weight range). (a) Analysis of polypeptides produced by pCF31 and pCF30. 1, CSR603(pUC8); 2, CSR603(pCF31); 3, CSR603(pCF30). (b) Analysis of polypeptides produced by pCF20 and pCF20::Tn5 insertion mutants in the xylR region. 1, CSR603(pRK2501); 2, CSR603(pCF20); 3, CSR603(pCF20::Tn5-45.1); 4, CSR603(pCF20::Tn5-41.2); 5, CSR603(pCF20::Tn5-30.1); 6, CSR603(pCF20::Tn5-2.1); 7, CSR603(pCF20::Tn5-2.1).
Fig. 4. Sequencing strategy. Arrows show the direction and length of sequences used in the compilation of Fig. 5. Region shown in Fig. 5; other symbols as in Fig. 1.

**xylR.** The 570 bp *BglII* fragment of pCF30 was ligated into the *BamHI* site of M13mp18; clones in opposite orientations were used to determine the sequence at each end (1; full arrows). The sequence was completed by constructing deletion derivatives using *ExoIII* and S1 nuclease, according to Henikoff (1984) (1, 2; half arrows), and by subcloning of *Sau3A* fragments (3).

**xylS.** The 1630 bp *BglII–EcoRI* fragment of pCF27 was cloned into M13mp18 cut with *BamHI* and *EcoRI* to generate M13bE2 (4; full arrow). Deletions were made as above to provide a continuous sequence (4; half arrows). The 1630 bp *BamHI–PstI* fragment of M13bE2 was prepared from the double-stranded replicative form and was ligated to M13mp19 cut with *BamHI* and *PstI*. Deletions were made as above to generate clones for sequencing (5).

In order to link the sequences of *xylS* and *xylR* correctly, pCF30 was cut with *HincII* and ligated in a large volume to make pCF30Ah2, the 1·1 kb *HindIII–EcoRI* fragment of which was cloned into M13mp19 for sequencing (6). To further extend the sequence of the 5′ end of *xylR*, pCF30Ah2 was cut with *BglII* and ligated in a large volume to generate pCF30Ah2Ab2, the 590 bp *HindIII–EcoRI* fragment of which was sequenced after cloning into M13mp19 (7).

Phosphotransferase II encoded by Tn5 also failed to reveal a putative *xylS* polypeptide (data not presented). However, a candidate did emerge from the sequencing studies and is presented below.

**DNA sequencing**

The nucleotide sequence presented in Fig. 5 was obtained using the strategy outlined in Methods and Fig. 4.

**xylS.** The position of *xylS* had been previously determined by mapping of transposon insertion mutations. Sequencing reveals a 963 bp open reading frame (ORF), terminating in the codon TGA, that is capable of encoding an unmodified product of 36·5 kDa, and is in the region expected to contain *xylS*. No other ORF of similar length is present on either strand in this region. That this ORF is indeed the *xylS* gene is further suggested by the sequence of pCF20::Tn5-26.1; the transposon interrupts this ORF after base 258 (data not presented). Upstream of this ORF is a purine-rich region (Fig. 6) with some similarities to that upstream of
Fig. 5. (a) DNA sequence of the 5' end of xylR and the upstream region. Putative promoters are indicated by overlining, inverted repeats by underlining and direct repeats by dots. (b) DNA sequence of the xylS region. Derived polypeptide sequences are displayed in a one-letter notation (IUPAC-IUB, 1969).
(a)

E. coli 16S rRNA

3' -AUUCCUC A

xylS RNA

5' -AAAAUAAGAGAACCGUGAACUAUG

P. aeruginosa 16S rRNA

3' -AUUCCPy

(b)

xylR

5' -AUAAAAAAAAAGAGAACCGUGAACUAUG

xylS

5' -AAAAUAAGAGAACCGUGAACUAUG

xylE

5' -GACUAUGAACAGGUGACGCUG

Fig. 6. (a) RNA sequence of the 5' region from the initiator codon of xylS. Potential base pairing with the 3' ends of 16S rRNA of E. coli (Shine & Dalgarno, 1974) and P. aeruginosa (Shine & Dalgarno, 1975) are indicated. (b) Homology of ribosome-binding regions of xylR, xylS and xylE (Nakai et al., 1983). Regions of homology are underlined.

xylE (Nakai et al., 1983), in particular a pentanucleotide with complementarity to the 3' end of 16S rRNA of Pseudomonas aeruginosa and a tetramer complementary to part of the 3' end of 16S rRNA of E. coli: thus, this region may be a ribosome-binding site. There are no clear candidates, on the basis of comparing nucleotide sequences, of either E. coli-like (Hawley & McClure, 1983) or Pseudomonas-like (Mermod et al., 1984) promoters.

xylR. The approximate position of xylR and its direction of transcription can be deduced from the following points. (i) It must lie between the sites of insertion of pCF20::Tn5-45.1 and -2.1, and must lie over the mapped positions of pCF20::Tn5-7.3, -41.2, -3.1 and -30.1, giving maximum and minimum sizes of 2-2 kb and 0-6 kb, respectively. (ii) The polypeptide product of xylR is approximately 68 kDa, giving an estimated gene size of 1.5-2.0 kb, in reasonable agreement with the estimate of point (i). (iii) An RNA polymerase binding site likely to be the xylR promoter has been mapped on the XhoI-D fragment 1.98 ± 0.06 kb from its left-hand terminus (on Fig. 2) (Franklin et al., 1983). (iv) The 450 bp HindIII subfragment of the XhoI-D fragment (Fig. 2) is required for xylR expression (Inouye et al., 1983); thus, one terminus of xylR lies near the insertion of pCF20::Tn5-2.1. (v) pCF30 produces a truncated product of ~30 kDa; this in conjunction with (iii) indicates transcription from left to right in Fig. 1.

Thus, xylR is expected to lie such that the end encoding the amino-terminus of the product is ~2 kb from the XhoI site of pCF31, and with the end encoding the carboxy-terminus of the product in the 450bp HindIII fragment of pCF31. In this position, nucleotide sequencing reveals the 5' end of an ORF that probably corresponds to xylR. Immediately upstream of this is a clear candidate for a ribosome-binding site (Shine & Dalgarno, 1974); indeed, the region flanking the AUG initiator codon is very similar in base-constitution to typical E. coli genes (W101 = 35; Stormo et al., 1982). W101 is a weighting function that can distinguish E. coli translational starts in linear DNA sequences and which has good predictive ability. Furthermore, there are two candidates for a Pseudomonas-like promoter upstream, near the position expected from mapping of RNA polymerase binding sites (Franklin et al., 1983).

Intergenic region. Direct and inverted repeats have been identified and are shown in Fig. 5. Further experiments are required to determine whether these have any biological functions.

**Discussion**

In this study, we describe the isolation of Tn5 insertion mutant derivatives of the TOL XhoI-D fragment, the functional characterization of such insertion mutations within the xylR (Tn5-7.3, -41.2, -30.1) and xylS (Tn5-7.1, -31.1, -26.1) regulatory genes of the TOL degradative pathway.
and their utilization along with pUC8 derivatives pCF30 and pCF31 to attempt identification of the polypeptide products of the regulatory genes. The mapped locations of the insertions within the genes and of those flanking the genes reveal more precisely the locations and extents of xylR and xylS within the XhoI-D fragment of the TOL plasmid pWW0. Since the submission of this paper for publication, Inouye et al. (1985) have published essentially similar results to those of this study for the xylR region.

The xylR mutants described show induction of the lower pathway, measured by catechol 2,3-oxygenase activity, in the presence of m-toluate but not m-methylbenzyl alcohol, as expected from the model for regulation of the TOL catabolic pathway and from previous reports. A polypeptide product of ~68 kDa has been identified as the xylR gene product. The direction of transcription of xylR can be inferred as being from left to right in Fig. 1, by comparing the polypeptides produced in pCF30 and pCF31, since the XhoI-PstI fragment contained in pCF30 produces a truncated polypeptide. DNA sequencing has identified an ORF of at least 215 bases on the DNA strand encoding xylR, and in the position expected. A polypurine tract immediately upstream probably contains a ribosome-binding site since it shows complementarity to the 3'-end of P. aeruginosa 16s rRNA (Shine & Dalgarno, 1975) and to the sequences thought to bind ribosomes in the transcripts of xylC (Inouye et al., 1984) and xylE (Nakai et al., 1983). Further upstream of this are two candidates for Pseudomonas-like promoters. Preliminary experiments show that one of these (positions -137 to -163, Fig. 5) binds E. coli RNA polymerase and is thereby protected from DNAase digestion. Although E. coli and P. putida RNA polymerases make identical contacts with E. coli promoters (Gragerov et al., 1984), Pseudomonas promoters may bind each of these enzymes differently. Studies on this are currently in progress.

The xylS mutants described here show no induction of the lower pathway in the presence of either m-toluate or benzoate. It has been reported (Franklin et al., 1983) that in xylS Tn5 mutants in pWW0-161, the lower pathway can be induced by benzoate, indicating the possibility of a third regulatory element. The absence of induction by benzoate in the three xylS mutants studied here strongly suggests that the proposed third element does not reside on the pWW0 XhoI-D fragment.

Although no clear candidate for the xylS polypeptide can be seen on maxicell gels, we expect it to be ~36 kDa in size since an ORF capable of encoding such a product is present, and is interrupted by Tn5-26.1. This ORF, designated xylS, is transcribed from right to left in Fig. 1. xylS and the region upstream have some unusual features. (i) Notwithstanding the regions of complementarity identified with E. coli 16s rRNA (see Results) the region is probably a poor ribosome-binding site (at least in E. coli), since the spacing between the Shine–Dalgarno candidate and the initiator codon is too small to fulfil the simple 'rule' of Gold et al. (1981). Furthermore, the region flanking the AUG initiator has a base constitution very unlike those flanking the starts of typical E. coli genes (W101 = -188, using the algorithm of Stormo et al., 1982). (ii) There are no clear candidates for an E. coli-like promoter or, indeed, for a Pseudomonas-like promoter. Both of these factors may contribute to the low level of expression of xylS noted in E. coli.

In addition the ORF designated xylS has an unusual codon usage that differs from that of the mercuric reductase gene on Tn501 (Brown et al., 1983), xylE (Nakai et al., 1983) and the carboxypeptidase G2 gene (Minton et al., 1984) in two ways that may be significant (Table 2). Firstly codons UUA (Leu), AUA (Ile), and AGA and AGG (Arg) are used, and secondly there is no excess of G or C in the third position for amino acids His, Gln, Asn, Lys and Asp. A parallel may be seen between xylS, with its use of very rare codons, and the dnaG gene (and other regulatory genes) of E. coli (Konigsberg & Godson, 1983), which uses a high percentage of rare codons and is subject to low expression. It has been suggested (Minton et al., 1984) that the bias of Pseudomonas genes towards codons ending in C or G contributes to the low expression of these genes in E. coli. Whether this is true of xylR is unknown; of the ORF most likely to represent the 5' end of xylR, only 71 codons have been determined in this study, none of which are unusual. As might be expected for a regulatory gene, xylS has an excess (+15) of basic amino acids (Lys, Arg, His) over acidic (Glu, Asp).
Table 2. Codon usages of Pseudomonas genes

a, The mercuric reductase gene of Tn501 (Brown et al., 1983); b, *xyE* (Nakai et al., 1983); c, the carboxypeptidase G2 gene (Minton et al., 1984); d, *xyLS* (this study; unusual choices of codons are indicated by asterisks).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>c</th>
<th></th>
<th></th>
<th>a</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>UUU</td>
<td>Phe</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>UCU</td>
<td>Ser</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>UAU</td>
<td>Tyr</td>
<td>2</td>
<td>7</td>
<td>0</td>
<td>2</td>
<td>UGU</td>
<td>Cys</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>UUC</td>
<td></td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>10</td>
<td>UCC</td>
<td></td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>UAC</td>
<td></td>
<td>6</td>
<td>5</td>
<td>13</td>
<td>1</td>
<td>UGC</td>
<td></td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>UUA</td>
<td>Leu</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5*</td>
<td>UCA</td>
<td></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>UAA</td>
<td>Och</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>UGA</td>
<td>Opl</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>UUG</td>
<td></td>
<td>9</td>
<td>5</td>
<td>0</td>
<td>7</td>
<td>UCG</td>
<td></td>
<td>5</td>
<td>1</td>
<td>7</td>
<td>3</td>
<td>UAG</td>
<td>Amb</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>UGG</td>
<td>Trp</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>CUU</td>
<td>Leu</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>CCU</td>
<td>Pro</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>CAU</td>
<td>His</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>6</td>
<td>CGU</td>
<td>Arg</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>CUC</td>
<td></td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>CCC</td>
<td></td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>CAC</td>
<td></td>
<td>9</td>
<td>10</td>
<td>4</td>
<td>8</td>
<td>CGC</td>
<td></td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>CUA</td>
<td></td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>CCA</td>
<td></td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>CAA</td>
<td>Gln</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>CGA</td>
<td></td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>CUG</td>
<td></td>
<td>30</td>
<td>18</td>
<td>30</td>
<td>11</td>
<td>CCG</td>
<td></td>
<td>19</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>CAG</td>
<td></td>
<td>16</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>CGG</td>
<td></td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>AUU</td>
<td>Ile</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>ACU</td>
<td>Thr</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>AAU</td>
<td>Asn</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>9</td>
<td>AGU</td>
<td>Ser</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>AUC</td>
<td></td>
<td>21</td>
<td>6</td>
<td>21</td>
<td>12</td>
<td>ACC</td>
<td></td>
<td>22</td>
<td>13</td>
<td>15</td>
<td>4</td>
<td>AAC</td>
<td></td>
<td>11</td>
<td>6</td>
<td>17</td>
<td>8</td>
<td>AGC</td>
<td></td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>AUA</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4*</td>
<td>ACA</td>
<td></td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>AAA</td>
<td>Lys</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>7</td>
<td>AGA</td>
<td>Arg</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AUG</td>
<td>Met</td>
<td>12</td>
<td>11</td>
<td>7</td>
<td>5</td>
<td>ACG</td>
<td></td>
<td>12</td>
<td>1</td>
<td>9</td>
<td>2</td>
<td>AAG</td>
<td></td>
<td>15</td>
<td>10</td>
<td>31</td>
<td>8</td>
<td>AGG</td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GUU</td>
<td>Val</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>GCU</td>
<td>Ala</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>GAU</td>
<td>Asp</td>
<td>5</td>
<td>8</td>
<td>1</td>
<td>7</td>
<td>GGU</td>
<td>Gly</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>GUC</td>
<td></td>
<td>21</td>
<td>4</td>
<td>10</td>
<td>7</td>
<td>GCC</td>
<td></td>
<td>40</td>
<td>10</td>
<td>37</td>
<td>7</td>
<td>GAC</td>
<td></td>
<td>23</td>
<td>19</td>
<td>24</td>
<td>7</td>
<td>GGC</td>
<td></td>
<td>29</td>
<td>13</td>
</tr>
<tr>
<td>GUA</td>
<td></td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>GCA</td>
<td></td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>GAA</td>
<td>Glu</td>
<td>15</td>
<td>12</td>
<td>10</td>
<td>9</td>
<td>GGA</td>
<td></td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>GUG</td>
<td></td>
<td>24</td>
<td>13</td>
<td>20</td>
<td>4</td>
<td>GCG</td>
<td></td>
<td>40</td>
<td>3</td>
<td>14</td>
<td>7</td>
<td>GAG</td>
<td></td>
<td>17</td>
<td>9</td>
<td>16</td>
<td>15</td>
<td>GGG</td>
<td></td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>
The identification of the xylR polypeptide product in the maxicell system permits further experimentation to construct plasmids with higher levels of expression of this regulatory gene. Similarly, the identification of the xylS ORF should allow construction of plasmids expressing high levels of the xylS gene product. It is hoped that these will lead to the purification and further characterization of the xylR and xylS products. Such experiments are currently in progress.

We wish to thank Drs C. Thomas, C. A. Smith and M. Pinkney for helpful discussions; Dr G. Ware for the gift of computer programs; Dr D. Marshall for instructions in computing; Mrs C. Price for technical assistance; Mrs C. Parry for photography; and Mrs E. A. Badger for typing of the manuscript. The work was supported by SERC Grant no. GR/C/59703.

REFERENCES


