D-Xylose (D-glucose) isomerase from *Arthrobacter* strain N.R.R.L. B3728

Gene cloning, sequence and expression

Thérèse LOVINY-ANDERTON,* Pang-Chui SHAW,† Myung-Kyo SHIN and Brian S. HARTLEY‡

Centre for Biotechnology, Imperial College of Science, Technology and Medicine, London SW7 2AZ, U.K.

Arthrobacter strain N.R.R.L. B3728 superproduces a D-xylose isomerase that is also a useful industrial D-glucose isomerase. The gene (*xylA*) that encodes it has been cloned by complementing a *xylA* mutant of the ancestral strain, with the use of a shuttle vector. The 5' region shows strong sequence similarity to *Escherichia coli* consensus promoters and ribosome-binding sequences and allows high levels of expression in *E. coli*. The coding sequence shows similarity to those for other D-xylose isomerases and is followed by 22 nucleotide residues with stop codons in each reading frame, a good 'consensus' ribosome-binding site and an open reading frame showing similarity to those of known D-xylulokinases (*xylB*). Studies on the expression of the cloned gene in *Arthrobacter* and in *E. coli* suggest that the two genes are part of a *xyl* operon regulated by a repressor that is defective in strain B3728. Codon usage in these two genes, and in another open reading frame (*nxi*) that was adventitiously isolated during early cloning attempts, shows some characteristic omissions and a strong G + C preference in redundant positions.

INTRODUCTION

This work is part of a programme to improve the thermostability and lower the pH optimum of the potentially useful industrial enzyme D-xylose (D-glucose) isomerase (EC 5.3.1.5), so as to allow production of sweeter high-fructose syrups. As steps to this end, we have studied in detail the purification and properties of the wild-type enzyme (Smith *et al.*, 1991) and developed a host–vector system for the optimal industrial strain that would allow mutant genes to be re-introduced and expressed in it (Shaw & Hartley, 1988). Meanwhile our colleagues in the Department of Biophysics have crystallized the enzyme (Akins *et al.*, 1986), determined its tertiary structure (Henrick *et al.*, 1989) and made studies of its ligand-binding and mechanism (Collyer *et al.*, 1990).

The present paper reports the strategies used to clone and sequence the *Arthrobacter* gene that encodes the D-xylose isomerase, referred to below as *xylA*, and construction of expression systems that allow the enzyme to be produced as about 10% of total soluble protein either in *Escherichia coli* or in the natural *Arthrobacter* host.

MATERIALS AND METHODS

Chemicals and enzymes

These were purchased as follows: radioactive nucleotides, restriction endonucleases, Klenow fragment of *E. coli* DNA polymerase from Amersham International (Amersham, Bucks., U.K.); calf intestinal phosphatase, other restriction endonuclease, glycogen, isopropyl β -D-thiogalactopyranoside, 5-bromo-4-chloroindol-3-yl β -galactoside, DNA molecular-mass markers and d(C')GTP from Boehringer Mannheim (Mannheim, Germany); [–40]DNA sequencing primer from BioLabs (Beverly, MA, U.S.A.). The eight deoxy- and dideoxy-

nucleotides used for DNA sequencing were from P-L Pharmacia (Uppsala, Sweden). Ultra-pure D-xylose was from Sigma Chemical Co. (Poole, Dorset, U.K.). All other chemicals and enzymes were as in Smith *et al.* (1991).

Bacterial strains, plasmids and culture conditions

Arthrobacter strains N.R.R.L. B3724 and B3728 were from the American Type Culture Collection; the xylose-minus strain PC1 used to isolate the *xylA* gene was a spontaneous mutant of strain B3724 (Smith, 1980). The *E. coli* *xylA* strain JA221 used for *xylA* complementation was from Briggs *et al.* (1984). *E. coli* strains HB101 (Boyer & Roulland-Dussoix, 1969) and JM101 (Messing, 1983) were hosts for propagation of plasmids and bacteriophages respectively.

The hybrid plasmid pCG2100 is described by Shaw & Hartley (1988); plasmid pTZ19U (Mead *et al.*, 1986) was purchased from Bio-Rad Laboratories (Watford, Herts., U.K.) and bacteriophages M13mp8 and M13mp9 (Messing, 1983) were from Amersham International.

Media for culture, preparation and propagation of *Arthrobacter* protoplasts were as in Shaw & Hartley (1988). Media for *E. coli* (LB, 2×TY and minimal media and plates with or without antibiotics) were as in Miller (1970). *Arthrobacter* cultures were at 30 °C and *E. coli* cultures at 37 °C.

Methods

Enzyme purifications, assays and characterizations were as in Smith *et al.* (1991).

Recombinant DNA techniques involving *E. coli* were carried out as described by Maniatis *et al.* (1982) and those involving *Arthrobacter* as described by Shaw & Hartley (1988) unless otherwise stated.

Synthesis of redundant oligonucleotide probes was by the cyanoethyl phosphoramidite method of Sinha *et al.* (1984) with

* Present address: Department of Neuroscience, Institute of Psychiatry, De Crespigny Park, London SE5 8AS, U.K.

† Present address: Department of Biochemistry, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong.

‡ To whom correspondence should be addressed.

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X59466.

the use of an Allied Biosystems DNA synthesizer, followed by purification from polyacrylamide gels.

Chromosomal library of *Arthrobacter* strain B3728 DNA in plasmid pBR322

This was constructed by digesting 200 µg of B3728 DNA with 30 units of *Sau*3A enzyme in 3 ml at 37 °C for 1 h, and the 5–10 kb size fragments were collected after extraction and sucrose-density-gradient centrifugation. These fragments (300 ng) were ligated to 100 ng of *Bam*HI-digested and dephosphorylated pBR322 DNA, and the product was transformed into 0.2 ml of competent cells of *E. coli* strain JA221 (*xylA*) prepared by the $\text{CaCl}_2/\text{RbCl}$ procedure of Maniatis *et al.* (1982). The recombinant cells were selected by spreading on LB plates containing ampicillin (100 µg/ml). To create a library, this procedure was repeated until a total of 28.8 ml containing 2×10^8 *E. coli* JA221 cells had been transformed with 21.6 µg of fragment DNA ligated to 7.2 µg of pBR322 DNA, resulting in 16000 amp^r colonies of which 12400 were tet^s and therefore presumably contained inserts at the *Bam*HI site.

Chromosomal libraries of *Arthrobacter* strain B3728 DNA in pUR expression vectors

A partial *Sau*3A digest of *Arthrobacter* B3728 DNA (1.5 µg; average fragment size 0.5–2.2 kb) was ligated to a mixture (0.6 µg total) of *Bam*HI-digested and dephosphorylated DNA from expression vectors pUR290, pUR291 and pUR292, so as to allow expression in all three reading frames. The product was transformed into the competent host *E. coli* BMH71–18 (Ruther & Muller-Hill, 1983); the whole library contained 180000 colony-forming units.

To confirm that the *xylA* gene was present in the 1.9 kb insert in plasmid pAX2, a small library of this DNA (50 ng) was digested with *Sau*3A, ligated into the *Bam*HI site of expression vectors pUR290, pUR291 and pUR292 (40 ng) and transformed into *E. coli* BMH71–18.

Chromosomal libraries of *Arthrobacter* strain B3728 DNA in plasmid pCG2100

These were constructed by ligating 1.2 µg of *Sal*I-digested (or *Cla*I-digested) B3728 DNA, sized 3–10 kb, to 2 µg of *Sal*I-digested (or *Cla*I-digested) and dephosphorylated pCG2100 DNA. The mixtures were used to transform about 10^{11} protoplasts of *Arthrobacter* strain PC1 (*xylA*) in 2 ml by the protocol of Shaw & Hartley (1988). The transformed protoplasts were regenerated on rich agar plates containing the selection antibiotic kanamycin sulphate (4 mg/ml). After 7 days at 30 °C, the regenerated kan^r colonies were collected and concentrated.

The clones containing functional *xylA* genes were then detected by their ability to grow on minimal agar plates containing kanamycin and 0.2% (w/v) D-xylose as sole carbon source. No colonies were found from the *Cla*I digest library, but after 2 days at 30 °C eight colonies were found from the *Sal*I digest library. Their plasmid DNAs were isolated and found to be unique by restriction-endonuclease mapping.

Complementation of *E. coli* *xylA* strain JA221

The plasmid DNA selected as above (100 ng) was transformed into 0.2 ml of competent cells of this *E. coli* strain prepared by the $\text{CaCl}_2/\text{RbCl}$ procedure of Maniatis *et al.* (1982). The recombinant cells were spread on minimal agar plates containing xylose (2 mg/ml) and ampicillin (100 µg/ml) plus leucine and tryptophan to satisfy the auxotrophic requirements of strain JA221.

A similar technique was used to subclone the *Arthrobacter* *xylA* gene, by using plasmids containing various deletions in the

insert DNA isolated as above. This eventually yielded a functional 1.9 kb insert.

DNA sequence analysis

Several recombinant M13mp8 and M13mp9 bacteriophages containing fragments of the 1.9 kb insert were constructed by using the procedure of Messing (1983) and sequenced by the method of Sanger *et al.* (1977) with [γ -³⁵S]thio]dATP for radioactive detection (Biggin *et al.*, 1983).

In order to resolve some strong compressions in G+C-rich parts of the sequence, d(C⁺)GTP was used in similar proportion in place of GTP in each dideoxy chain-termination reaction mixture (Mizusawa *et al.*, 1986), and the reactions with the Klenow DNA polymerase were incubated at higher temperature (48 °C) adding more polymerase in the 'chase step' of the reactions. The 6% acrylamide denaturing gels contained 40% deionized formamide as well as the 7 M-urea used in routine sequencing gels.

RESULTS AND DISCUSSION

Relevant conclusions can be drawn in hindsight from some unsuccessful strategies to clone the *Arthrobacter* B3728 *xylA* gene encoding D-xylose isomerase, as follows.

Cloning by complementation of an *E. coli* *xylA* mutation

The partial *Sau*3A digest library of *Arthrobacter* B3728 DNA contained 12400 inserts sized 5–10 kb. Statistically about 3500 should contain a gene of this size, but no colonies grew on ampicillin and D-xylose as sole carbon source or gave purple colonies on EMB plates supplemented with D-xylose and ampicillin (Briggs *et al.*, 1984). This is not because *E. coli* fails to recognise *Arthrobacter* expression signals (see below). Some of the large number of *Sau*3A sites (GATC) in this gene (39, 372, 600, 720, 856, 1180, 1375, 1772, 1805, 1862) may be preferentially cleaved.

Use of redundant oligonucleotide probes based on peptide sequences

Three mixed oligonucleotides were synthesized according to partial amino acid sequences (Smith *et al.*, 1991), with G and C as the redundant third codon since *Arthrobacter* was reported to have high G+C content (Keddie, 1974). As can be seen from the corresponding fragments of the actual *xylA* sequence, Probes 1 and 3 each contain one mismatch (underlined) and Probe 2 contains two mismatches:

```
Peptide 1 sequence: -Asp-Ala-Glu-Ala-Ala-Glu-Arg-
Probe 1 sequence :  GAT GCC GAA GCC GCC GCC GA
                   C  G  G  G  G  G
Xyl A (1495-1514) :  GAC GCC GAG GCC GCC GCA GA
```

```
Peptide 2 sequence: (NT) Ser-Val-Gln-Pro-Thr-Pro-Ala-Asp-
Probe 2 sequence :  GTC CAA CCC ACC CCC GCC GA
                   G  G  G  G  G  G
Xyl A (397-416) :  GTT CAG CCG ACC CCT GCA GA
```

```
Peptide 3 sequence: -Asp-Ala-Thr-Glu-Ala-Glu-Arg-
Probe 3 sequence :  GAC GCC ACC GAA GCC GAA AG
                   G  G  G  G  G  C
Xyl A (574-593) :  GAC GCC ACC GAG GCA GAG CG
```

These probes were 5'-end-labelled with [γ -³²P]ATP and used in Southern-blot hybridizations of various restriction-endonuclease digests of *Arthrobacter* B3728 DNA and with lysed colony filters (Grunstein & Hogness, 1975) of the genomic library described above.

Probe 2 gave only weak Southern-blot hybridization and only

```

-36   -30       -20       -10       -1 1
AGCAAAATTCCTAGGAAGACAAGGTAGGGTCAAATT ATG GCA TCA AAC GCA AGC GAT
                                         Met Ala Ser Asn Ala Ser Asp 7
22
GAA CTG ATC GGC ACC TGG GTA AGC GGC TGG GCA GGA GCC CGT GGC TAT
Glu Leu Ile Gly Thr Trp Val Ser Gly Trp Ala Gly Ala Arg Gly Tyr 23
70
GAA ACT CGC AAT GAG GGA CGA GTC CAC GCT GCC CTG CGT CAC GAC ACC
Glu Thr Arg Asn Glu Gly Arg Val His Ala Ala Leu Arg His Asp Thr 39
118
ACT GAA GAC TGG GAA TAC GTT ATT TAC GGC CCG TCC AAA GAG GAA CTA
Thr Glu Asp Trp Glu Tyr Val Ile Tyr Gly Pro Ser Lys Glu Glu Leu 55
166
GCC GCG GTC GCC GAG ACC CTC AAA AAG CAC CCT AAC CGT CGA CTA ACG
Ala Ala Val Ala Glu Thr Leu Lys Lys His Pro Asn Arg Arg Leu Thr 71
214
GCC TTC GAT GAT TCG GCC GAA AAC TTG GTC GTC ATC GCC AAT GAG GTT
Ala Phe Asp Asp Ser Ala Glu Asn Leu Val Val Ile Ala Asn Glu Val 87
262
GGC CTT CAA GTT ACG GCC GAT GAC GAA GCA CTG ATG GTC ACC GAA CTT
Gly Leu Gln Val Thr Ala Asp Asp Glu Ala Leu Met Val Thr Leu Glu 103
310
GCT GTG CAT GAC GTC GAG GTC CCA CTT CCT GCT GAC GGT TTT GTT TTC
Ala Val His Asp Val Glu Val Pro Leu Pro Ala Asp Gly Phe Val Phe 119
358
CAG ATT GAA CGC GAT GGC ACC CAC GCG TAC GTC TCA CTA CAC CCT GAG
Gln Ile Glu Arg Asp Gly Thr His Ala Tyr Val Ser Leu His Pro Glu 135
406
GAC AAC GAA GAG CTT GTT GCC GCT TCG GGT CAT GTT TCG GCC GTG AAT
Asp Asn Glu Glu Leu Val Ala Ala Ser Gly His Val Ser Ala Val Asn 151
454
GGC TTT GCG ATC TTC GAC CGT ATT ATC ACC GGC GCA GAT TTC CGT CGC
Gly Phe Ala Ile Phe Asp Arg Ile Ile Thr Gly Ala Asp Phe Arg Arg 167
502
CGT GGC CTC ACC CTG ATC ATG CGC GCT TGG CTT CCT GGA TGC CAC AGC
Arg Gly Leu Gly Thr Leu Ile Met Arg Ala Trp Leu Pro Trp His Arg 183
560
AC....(ca. 100 b.p.)....TCGACGCT CCCGCCAGC CAAGGATGAA ATTACGAGCG
663
701
ACTGGCCTCA ATGGCTCAGC AGTGAACTCTCA TCGACATTTT CCAAGGCTTG GCATCCAGCG
761
CCCGTGTTG CACCAGGTCC AAGCGGCGAA TCTCATCCAC GAAGGCCAGC ATACGATGGT
821
GTCTACCGGC ACCGCTTCCG GCAAGTCCAT GGCCTACCTC ATGCCGAGTT TGGATGCCCT
881
ATTCAAGTCA CGCGATAGCC TTTAATGATG CCGATTCCGG CGCCTCAATT CTCTACATT
941
GTCCAACCAA GGCGCTACGG CAGACCAACT ATCTGCGGTGC A

```

Fig. 1. DNA sequence and coding sequence of the '*nxi* gene'

This was isolated as a colony from a partial *Sau3A* digest library that hybridized strongly with the redundant oligonucleotide (GAT^T/GCC^C/GGA^A/GCC^C/GCC^C/GGA^A), which is complementary to nucleotide residues 907–936 in the 3' 'untranslated' region, with mismatches shown in *italics*. The residues with lines above in the 3'-end region show termination codons in each reading frame. The residues in **bold** in the 5' 'untranslated' region fit the *E. coli* promoter consensus (Harley & Reynolds, 1987), and those underlined fit the *E. coli* consensus ribosome-binding site (Shine & Dalgarno, 1974).

to a single colony in the library, which on subsequent growth appeared to lose its plasmid insert, but Probe 3 hybridized strongly to several bands and to several colonies, and to a 3 kb *EcoRI* fragment of plasmid DNA extracted from these. However, when this was subcloned and partially sequenced, no open reading frame was found.

Probe 1 also hybridized strongly to several bands and to several colonies, and to a 2.1 kb *SalI* fragment from these. After subcloning and sequencing, it showed a 5' open reading frame that encoded some small peptides previously identified from partial sequence studies (Hartley *et al.*, 1987). Hence the adjacent restriction fragment was subcloned and sequenced to give the whole open reading frame (Fig. 1). Although this has plausible transcription and translation signals that match the *E. coli* consensus (see the legend to Fig. 1), the *N*-terminal sequence does not correspond to that determined for the purified D-xylose isomerase, and no similarity was observed when the DNA sequence from nucleotide residue –36 to 561 was compared with

six different D-xylose isomerase gene sequences (*Actinoplanes missouriensis*, *Ampullariella* sp., *Bacillus subtilis*, *E. coli*, *Streptomyces violaceoniger* and *Arthrobacter* sp.) by the proportional algorithm of the Staden program (Staden, 1982) with a spin length of 11 and a proportional score of 8.

The 183-amino acid-residue sequence of the putative protein was also searched against the sequences of the above xylose isomerases and other protein sequences within the composite protein database OWL version 6.0 with the program SWEEP, which is an extension of the Lipman–Pearson algorithm (Lipman & Pearson, 1985), but no sequence similarities were found.

Hence this is not the *xylA* gene, but the promoter sequence similarities and the codon usage (see below) suggest that it expresses a small protein that cannot exceed 300 residues, since there are stop codons in each reading frame in the 3'-end sequence shown in Fig. 1; in the absence of known function or sequence similarity we have named this gene *nxi* (not xylose isomerase).

Fig. 2. Construction of plasmids containing the *xylA* gene

Plasmid pCG2100 (Shaw & Hartley, 1988) contains a cryptic plasmid from *Corynebacterium glutamicum* N.C.I.B. 10026 and the kanamycin-resistance gene from plasmid pNCAT4 (Herrera-Estrella *et al.*, 1983) (bold lines) ligated into the tetracycline-resistance gene of plasmid pBR322 (Bolivar *et al.*, 1977) (thin line). Plasmid pAXI1 contains a 4.8 kbp *SalI* fragment of *Arthrobacter* B3728 DNA (bold line) cloned into the *SalI* site in plasmid pCG2100 (thin line). The boxed 1.9 kbp *SalI*–*BssHII* fragment was sequenced as shown in Fig. 3 and shown to contain the complete *xylA* gene and part of the *xylB* gene as indicated. Plasmid pAXI2 was constructed to allow expression of the *xylA* gene in *E. coli* under control of the *lacZ'* promoter, by fusing the *SmaI* fragment (bold line) from the insert in pAXI2 to the expression vector pTZ19U described by Mead *et al.* (1986) (thin line) with the use of multiple site linkers containing a universal sequencing primer site as indicated (thick lines). The stippled and white boxes again represent the *xylA* and *xylB* genes respectively.

Antibody screening of colonies producing β -galactosidase fusion proteins

Rabbit antibodies against purified *Arthrobacter* D-xylose isomerase were purified by affinity purification on nitrocellulose-bound antigen (Robinson *et al.*, 1988) and used to screen about 30000 colonies on nitrocellulose filters of a partial *Sau3A* digest library in *E. coli* BMH71–18 of *Arthrobacter* B3728 DNA in vectors pUR290, pUR291 and pUR292 (Ruther & Muller-Hill, 1983). After inducing the active β -galactosidase-antigen fusion proteins with isopropyl β -thiogalactopyranoside, three colonies were strongly positive. Preliminary mapping and sequencing showed that they contained fragments of the coding sequence shown in Fig. 4, proving that this was indeed the *xylA* gene, but further study was discontinued when the full-length gene was cloned as follows.

Cloning by complementation in *Arthrobacter* strain PC1 (*xylA*)

This strategy was adopted because of doubts about recognition of *Arthrobacter* expression signals by *E. coli*. It relies on the hybrid plasmid pCG2100, which has useful cloning sites, replicates in both organisms and has selection markers suitable for each (Shaw & Hartley, 1988). The transformation protocols for

Arthrobacter rely on protoplast fusion and regeneration, and are less efficient than those for *E. coli* but adequate for the direct selection strategy pursued here.

This selection relies on the *xylA* mutation in *Arthrobacter* strain PC1, isolated by Smith (1980) as a spontaneous mutant of the xylose-inducible strain B3724 that was unable to grow on D-xylose as sole carbon source. Cells grown on rich medium in the presence of xylose contain no xylose isomerase, but have high concentrations of D-xylulokinase, the second enzyme in the xylose pathway that in *E. coli* is encoded by the *xylB* gene, which is part of an *xyl* operon. This suggested that strain PC1 was an *xylA* mutant rather than an *xylB* or a regulatory or transport mutant.

Two libraries of *Arthrobacter* DNA from the xylose isomerase-constitutive strain B3728 were constructed by cloning 3–10 kb *BamHI* or *SalI* fragments into the respective cloning sites in plasmid pCG2100. After transformation of protoplasts of strain PC1 and regeneration, no colonies were found from the *BamHI* transformation that were able to grow on xylose as sole carbon source, but eight colonies from the *SalI* digest library were able to do so. Restriction-endonuclease mapping showed that each contained the same 4.8 kb insert, indicated as plasmid pAXI1 in Fig. 2.

Fig. 3. Location, subcloning and sequencing of the *xylA* gene in plasmid pAXI1

The boxes show the 4.8 kbp insert of *Arthrobacter* DNA in plasmid pCG2100 and deletion derivatives thereof, made by re-ligating restriction fragments of plasmid pAXI1 as indicated (restriction sites not to scale; S = *Sau*3A). The *xylA* gene (stippled) was located by transforming *Arthrobacter* strain PC1 with these and testing for growth on xylose (indicated + or -). The arrows show restriction fragments of the 1.9 kbp *SalI*-*Bss*HII insert in plasmid pAXI2, cloned and sequenced in bacteriophages M13mp8 and M13mp9.

Localization of the *xylA* gene within this 4.8 kb fragment was simplified by the observation that plasmid pAXI1 could complement the *xylA* lesion in *E. coli* strain JA221. Various constructs containing deletions in this insert DNA were tested for their ability to transform strain JA221 to allow growth on xylose, as shown in Fig. 3. Plasmid pAXI2 contained a 1.9 kb *SalI*-*Hind*III insert that retains complementing activity. Thereafter plasmids were normally propagated in *E. coli* strain HB101 for ease of transformation and DNA isolation.

The presence of the *xylA* gene in the 1.9 kb insert was confirmed by constructing a small library of *Sau*3A fragments of that DNA in the expression vectors pUR290, pUR291 and pUR292 in *E. coli* BMH71-18, and screening these with anti-XI as above; about 20 strongly positive colonies were found.

Sequence of the *Arthrobacter xylA* gene

The DNA sequence of the 1.9 kb insert in pAXI2, shown in Fig. 4, was deduced from various fragments subcloned into bacteriophages M13mp8 and M13mp9, as indicated in Fig. 3. Most of it has been sequenced more than once and in both orientations. Other fragments (results not shown) confirm the overlap across the *Bam*HI site.

The ATG at positions 391-393 encodes a methionine residue preceding an open-reading frame that begins with the known *N*-terminal sequence of the D-xylose isomerase and ends with stop codons in all three reading frames following position 1576. After the presumed removal of the *N*-terminal methionine residue, the derived amino acid sequence of 394 residues, corresponding to a subunit weight of 43 245 Da, fits the composition of the purified

enzyme and also the sequences of peptides derived from it (Smith *et al.*, 1991). This sequence was used to fit the crystallographic data to derive the tertiary structure, and to proposed sequence similarities to other known D-xylose isomerases (Henrick *et al.*, 1989).

Preceding the coding sequence is a 5'-end region that contains putative RNA polymerase- and ribosome-binding sequences that show almost perfect match to the equivalent *E. coli* 'consensus sequences', including 'spacer distances' (Shine & Dalgarno, 1974; Harley & Reynolds, 1987). The importance of these sequences is shown by noting that a 5'-end deletion of the pXI2 insert down to residues 298-300 retained its ability to complement *xylA* in *E. coli* strain JA221, whereas deletions down to positions 348-350 completely abolished complementation. This reinforces previous conclusions that expression signals in *Corynebacteria* are very similar to those in *E. coli* (Roberts *et al.*, 1985; Kaczorek *et al.*, 1985).

Following the *xylA* stop codon (1576-1578) is another open reading frame containing a plausible ribosome-binding sequence and ATG initiation codon at positions 1617-1620. The coding sequence thereafter shows similarity to the *N*-terminal sequence of D-xylulokinases from *Ampullaria* (Saari *et al.*, 1987), *E. coli* (Lawlis *et al.*, 1984) and *Klebsiella aerogenes* (Neuberger *et al.*, 1981), as shown in Fig. 5. Hence the D-xylose isomerase (*xylA*), D-xylulokinase (*xylB*) and xylose permease (*xylT*) genes appear to be in that order in a operon, as in the above species and in *Salmonella typhimurium* (Shamanna & Sanderson, 1979) and *Bacillus subtilis* (Wilhelm & Hollenberg, 1984).

Codon usage in the *Arthrobacter xylA* gene, the *xylB* gene

Fig. 5. Amino acid sequence similarities in D-xylulokinases
The N-terminal sequences of D-xylulokinases from *Ampullaria* (Saari *et al.*, 1987), *E. coli* (Lawlis *et al.*, 1984) and *K. aerogenes* (Neuberger *et al.*, 1981) are compared with the coding sequence following the *Arthrobacter xylA* gene.

Table 1. Codon usage in *Arthrobacter* strain B3728

Data for *xylA* gene (395 residues) and the N-terminus of *xylB* gene (96 residues) are from Fig. 2 and for the putative *nxi* gene (183 residues) from Fig. 1.

<i>xylA</i> <i>xylB</i> <i>nxi</i>					<i>xylA</i> <i>xylB</i> <i>nxi</i>					<i>xylA</i> <i>xylB</i> <i>nxi</i>					<i>xylA</i> <i>xylB</i> <i>nxi</i>				
TTT	Phe	4	0	2	TCT	Ser	1	1	0	TAT	Tyr	0	0	1	TGT	Cys	0	0	0
TTC	Phe	22	2	4	TCC	Ser	4	3	1	TAC	Tyr	9	0	3	TGC	Cys	0	1	0
TTA	Leu	0	0	0	TCA	Ser	0	1	2	TAA	*	—	—	—	TGA	*	—	—	—
TTG	Leu	0	0	1	TCG	Ser	4	1	3	TAG	*	—	—	—	TGG	Trp	5	2	5
CTT	Leu	2	2	5	CCT	Pro	4	1	4	CAT	His	0	1	2	CGT	Arg	4	2	6
CTC	Leu	5	1	2	CCC	Pro	0	0	0	CAC	His	13	2	6	CGC	Arg	14	3	4
CTA	Leu	0	1	3	CCA	Pro	10	0	1	CAA	Gln	0	1	1	CGA	Arg	0	0	2
CTG	Leu	31	3	4	CCG	Pro	3	2	1	CAG	Gln	9	3	1	CGG	Arg	0	0	0
ATT	Ile	2	0	3	ACT	Thr	0	0	2	AAT	Asn	3	1	3	AGT	Ser	0	1	0
ATC	Ile	13	6	5	ACC	Thr	22	4	7	AAC	Asn	15	1	4	AGC	Ser	6	1	2
ATA	Ile	0	0	0	ACA	Thr	0	0	0	AAA	Lys	0	1	2	AGA	Arg	0	0	0
ATG	Met	10	2	3	ACG	Thr	1	1	2	AAG	Lys	18	0	1	AGG	Arg	0	0	1
GTT	Val	1	1	6	GCT	Ala	3	1	5	GAT	Asp	5	5	6	GGT	Gly	4	1	2
GTC	Val	10	1	8	GCC	Ala	22	7	10	GAC	Asp	26	5	7	GGC	Gly	33	5	9
GTA	Val	4	2	1	GCA	Ala	11	0	5	GAA	Glu	14	2	10	GGA	Gly	2	2	3
GTG	Val	3	5	2	GCG	Ala	7	6	3	GAG	Glu	13	1	7	GGG	Gly	2	1	0

fragment and the adventitiously cloned ‘*nxi* gene’ is shown in Table 1. The absence of some codons such as TTA, ATA, CCC, ACA and AGA is noteworthy, and there is strong preference for codons ending in G or C in redundant positions, as expected from the estimated G + C content of the genomic DNA (57.4 %; Shaw, 1987) and observed by Roberts *et al.* (1985) for an *Arthrobacter ermA* gene.

Fig. 4. DNA sequence and coding sequence of the *xylA* gene

This shows the sequence of the 1.9 kbp *SalI*–*Bss*HIII insert in plasmid pAXI2. The region 5′ of the assumed mRNA start (residue + 1) shows residues (with asterisks) representing a putative CRP-binding site and others (underlined) that fit the *E. coli* promoter consensus (Farley & Reynolds, 1987). Residues marked S/D (lines over) preceding the presumed initiation codons (bold) fit the *E. coli* consensus ribosome-binding site (Shine & Dalgarno, 1974). Between the end of the *xylA* gene and the start of the *xylB* gene, three termination codons (over- and under-lined) and another consensus ribosome-binding site are indicated.

Expression and regulation of the *Arthrobacter xyl* operon

The fact that the *Arthrobacter xylA* gene could complement an *E. coli xylA* lesion implied that it was expressing from its own promoter, but the need to develop high-level expression vectors for mutant enzymes made it important to establish the relative levels and regulation in both species.

Table 2 shows that the chromosomal *xylA* gene is almost

Table 2. Comparative expression levels of the cloned *xylA* gene in xylose-inducible and xylose-constitutive strains of *Arthrobacter* and *E. coli* and from its own promoter or the *E. coli lacZ* promoter

The strains were transformed with plasmid pAXI1 containing the *Arthrobacter xylA* gene and its promoter from strain B3728 or with plasmid pAXI3 in which the *Arthrobacter xylA* coding sequence is fused in phase to the *E. coli lacZ* promoter in plasmid pTZ19U. Cultures were grown in shake flasks to stationary phase on minimal medium supplemented with 1% (w/v) glucose (G) or xylose (X) or in rich LB medium. Lysed cleared cell extracts were assayed against 1 M-D-fructose by the cysteine/carbazole assay method of Dische & Borenfreund (1951); line (b) shows the activity of resident host genes and line (a) minus line (b) reflects the level of expression from the plasmid-encoded gene. Abbreviations: IPTG, isopropyl β -D-thiogalactopyranoside; N.G., no growth.

Strain	Plasmid	Specific activity of isomerase in cells grown on medium:			
		G minimal	X minimal	G + X minimal	Rich LB
<i>Arthrobacter</i> B3724 (<i>xylA</i> -inducible)	(a) pAXI1	0.26	1.46	0.44	0.05
	(b) None	0.01	1.10	0.02	0.02
	(a) - (b)	0.25	0.36	0.42	0.03
<i>Arthrobacter</i> B3728 (<i>xylA</i> -constitutive)	(a) pAXI1	1.47	1.40	—	0.72
	(b) None	1.03	1.00	1.09	0.65
	(a) - (b)	0.44	0.40	—	0.07
<i>Arthrobacter</i> PC 1 (<i>xylA</i> , <i>xylB</i> -inducible)	(a) pAXI1	0.05	0.56	—	—
	(b) None	0.00	N.G.	—	—
	(a) - (b)	0.05	0.56	—	—
<i>E. coli</i> JA221 (<i>xylA</i> , <i>xylB</i> -inducible)	(a) pAXI1	0.26	1.46	0.44	0.05
	(b) None	0.01	1.10	0.02	0.02
	(a) - (b)	0.25	0.36	0.42	0.03
<i>E. coli</i> JM101 (Δ <i>lac pro</i> , <i>lacF</i> ⁺), IPTG-induced	(a) pAXI3	1.00	—	—	0.19
	(b) None	0.09	—	—	0.06
	(a) - (b)	0.91	—	—	0.13

completely repressed by growth on glucose rather than xylose in both the wild-type *Arthrobacter* strain B3724 and the mutant *xylA*, *xylB*-inducible strain PC1 derived from it; but its expression is unaffected in the *xylA*-constitutive strain B3728, which must therefore contain either an *xyl* repressor or *xyl* operator lesion. In the former case the cloned *xylA* gene in plasmid pAXI1 that was obtained from strain B3728 would be repressed by growth on glucose in strains B3724 and PC1, which contain functional repressors; but not in the presumed repressor-deficient strain B3728. Table 2 shows that this is so; the residual levels of activity in strains B3724 and PC1 on glucose can be explained by saturation of the repressor content by multi-copies of the cloned operator sites. Specific activities of isomerase were uniformly lower in cells grown on rich LB medium, but this may be because the total protein content of such cells is higher.

The *xyl* operon in *E. coli* is regulated by catabolite repression rather than a specific *xyl* repressor (Shamanna & Sanderson, 1979). To study expression of the cloned *Arthrobacter* gene in *E. coli*, plasmid pAXI1 was transformed into *E. coli* strain JA221 (*xylA*, *xylB*-inducible). For comparison, plasmid pAXI3 was constructed in which the *Arthrobacter xylA* coding sequence is fused to the *E. coli lacZ* promoter. This was then transformed into *E. coli* JM101, which has its own wild-type xylose-inducible *xyl* operon, but on glucose and in presence of isopropyl β -D-thiogalactopyranoside as *lacZ* inducer expression is only from the plasmid-encoded *xylA* gene.

Table 2 shows that expression from the *lacZ* promoter in plasmid pAXI3 reaches levels comparable with the best obtained in *Arthrobacter*, i.e. about 5–10% of soluble protein. We were concerned that the enzyme purified from *E. coli* might differ from that isolated from *Arthrobacter*, by virtue of post-translational processing such as incomplete removal of *N*-terminal methionine. The specific activity for fructose of such purified enzyme was up to 25% lower, but the *N*-terminal sequences proved to be identical (results not shown), and in all other respects such as

urea stability and thermostability the enzymes were identical. We believe that the discrepancy in specific activities is artifactual, due perhaps to garnering of inhibitory metal ions or oxidation of methionine residues during the slightly different purification protocols.

Moreover the cloned *Arthrobacter* gene in plasmid pAXI1 expresses well from its own promoter in *E. coli* strain JA221, and this expression is significantly diminished by growth on glucose. This suggests that the *E. coli* cyclic AMP-binding protein may recognize a *crp* regulation sequence in the 5'-end region of the *Arthrobacter* gene. There is such a sequence 5'-TGTG-3' (Morita *et al.*, 1988) at positions 268–271 in Fig. 4. The adjacent sequence shows less fit to the '*crp* consensus' (Busby, 1986) but may be adequate to recognize the *E. coli* cyclic AMP-binding protein (Shirabe *et al.*, 1985). This in turn suggests that there may be dual control of the *xyl* operon in *Arthrobacter*.

This work was performed in part with a grant from the Science and Engineering Research Council Protein Engineering Club. We thank the Croucher Foundation (P.-C.S.) and the British Commonwealth and Foreign Office (M.-K.S.) for research studentships.

REFERENCES

- Akins, J., Brick, P., Jones, H. B., Hirayama, N., Shaw, P.-C. & Blow, D. M. (1986) *Biochim. Biophys. Acta* **874**, 375–377.
- Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3963–3965.
- Bolivar, F., Rodriguez, R. L., Belach, M. C. & Boyer, H. W. (1977) *Gene* **2**, 75–94.
- Boyer, H. W. & Roulland-Dussoix, D. (1969) *J. Mol. Biol.* **212**, 211–235.
- Briggs, K. A., Lancashire, W. E. & Hartley, B. S. (1984) *EMBO J.* **3**, 611–616.
- Busby, S. W. (1986) in *Regulation of Gene Expression: 25 Years* (Booth, I. R. & Higgins, C. F., eds.), pp. 51–77, Cambridge University Press, Cambridge.

- Collyer, C. A., Henrick, K. & Blow, D. M. (1990) *J. Mol. Biol.* **212**, 211–235
- Dische, Z. & Borenfreund, E. (1951) *J. Biol. Chem.* **192**, 583–587
- Grunstein, M. & Hogness, D. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 7333–7337
- Harley, C. B. & Reynolds, R. P. (1987) *Nucleic Acids Res.* **15**, 2343–2361
- Hartley, B. S., Anderton, T. & Shaw, P.-C. (1987) in *Chemical Aspects of Food Enzymes* (Andrews, A. T., ed.), pp. 120–136, Royal Society of Chemistry, London
- Henrick, K., Collyer, C. A. & Blow, D. M. (1989) *J. Mol. Biol.* **208**, 129–157
- Herrera-Estrella, L., Depicker, A., Van Montagu, M. & Schell, J. (1983) *Nature (London)* **303**, 209–213
- Kaczorek, M., Zettlmeissl, G., Delpeyroux, F. & Streeck, R. E. (1985) *Nucleic Acids Res.* **13**, 3147–3159
- Keddie, R. M. (1974) in *Bergey's Manual of Determinative Bacteriology*, 8th edn. (Buchanan, R. E. & Gibbons, N. E., eds.), pp. 618–625, Williams and Wilkins, Baltimore
- Lawlis, V. B., Dennis, M. S., Chen, E. Y., Smith, D. E. & Heimer, D. J. (1984) *Appl. Environ. Microbiol.* **47**, 15–21
- Lipman, D. J. & Pearson, W. R. (1985) *Science* **227**, 1435–1441
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor
- Mead, D. A., Szczesna-Skorupa, E. & Kemper, B. (1986) *Protein Eng.* **1**, 67–74
- Messing, J. (1983) *Methods Enzymol.* **101**, 20–78
- Miller, J. H. (ed.) (1970) *Experiments in Molecular Genetics*, pp. 431–435, Cold Spring Harbor Laboratory, Cold Spring Harbor
- Mizusawa, S., Nishimura, S. & Seela, F. (1986) *Nucleic Acids Res.* **16**, 1319–1324
- Morita, T., Shigedass, K., Kimizuka, F. & Aiba, H. (1988) *Nucleic Acids Res.* **14**, 7315–7332
- Neuberger, M. S., Hartley, B. S. & Walker, J. A. (1981) *Biochem. J.* **193**, 513–524
- Roberts, A. N., Hudson, J. S. & Brenner, S. (1985) *Gene* **35**, 259–270
- Robinson, P. A., Anderton, B. H. & Loviny-Anderton, T. L. F. (1988) *J. Immunol. Methods* **108**, 115–122
- Ruther, U. & Muller-Hill, B. (1983) *EMBO J.* **2**, 1791–1794
- Saari, G. C., Kumar, A. A., Kawasaki, G. H., Insley, M. Y. & O'Hara, P. J. (1987) *J. Bacteriol.* **169**, 612–618
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Shamanna, D. K. & Sanderson, K. E. (1979) *J. Bacteriol.* **139**, 64–70
- Shaw, P.-C. (1987) Ph.D. Thesis, University of London
- Shaw, P.-C. & Hartley, B. S. (1988) *J. Gen. Microbiol.* **134**, 903–911
- Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1342–1346
- Shirabe, K., Ebina, Y., Miki, T., Nakazawa, T. & Nakazawa, A. (1985) *Nucleic Acids Res.* **13**, 4687–4698
- Sinha, N. D., Biernat, J., McManus, J. & Koster, H. (1984) *Nucleic Acids Res.* **12**, 4533–4557
- Smith, C. A. (1980) Ph.D. Thesis, University of London
- Smith, C. A., Rangarajan, M. & Hartley, B. S. (1991) *Biochem. J.* **277**, 255–261
- Staden, R. (1982) *Nucleic Acids Res.* **10**, 2951–2961
- Wilhelm, M. & Hollenberg, C. P. (1984) *EMBO J.* **3**, 2555–2560

Received 3 October 1990/14 December 1990; accepted 8 January 1991