Analysis of the gene cluster encoding carotenoid biosynthesis in *Erwinia herbicola* Eho13

Kin-Ying To,[†] Erh-Min Lai, Lan-Ying Lee,[‡] Tsuey-Pin Lin, Chien-Hui Hung, Chi-Liang Chen, Yu-Sun Chang and Shih-Tung Liu

Author for correspondence: Shih-Tung Liu. Tel: +886 3 328 0292. Fax: +886 3 328 3031.
e.mail: cgliu@sun670.imb.sinica.edu.tw

Molecular Genetics
Laboratory, Department of
Microbiology and
Immunology, Chang-Gung
Medical College, Kwei-Shan,
Taoyuan, Taiwan 333

***Erwinia herbicola* is known to synthesize carotenoids and gives an orange-coloured phenotype. These carotenoids play a role in the protection of the cells from the damage caused by near-UV irradiation in nature. The genes encoding these carotenoids in *E. herbicola* Eho13 are clustered in a 7 kb DNA fragment. The complete sequence of this fragment has been determined. DNA sequence analysis revealed that the entire sequence contains at least five genes, which are transcribed in the same direction. These five genes are organized in the order *crtE-crtX-crtY-crtI-crtB*. A gene fusion study showed that two different regions in this 7 kb gene cluster contain promoter activity. Primer-extension analysis identified two transcription start sites, located 147 bp upstream from the first gene of the cluster, *crtE*, and within the last gene of the cluster, *crtB*. An RNA-PCR study suggested that the five *crt* genes were organized in an operon and were transcribed from the promoter upstream from *crtE*.**

Keywords: *Erwinia herbicola*, carotenoid biosynthesis, operon, gene cluster

INTRODUCTION

Erwinia herbicola is a yellow-pigmented Gram-negative bacterium belonging to the family of *Enterobacteriaceae* (Pérombelon, 1992). This organism is ubiquitous in nature and is often found to colonize plant surfaces (Billing & Baker, 1963; Goodfellow *et al.*, 1976). Because of this colonization ability and the capability of producing antimicrobial substances (Winkelman *et al.*, 1980), many strains of *E. herbicola* have been used as biological control agents to prevent plant diseases (Beer *et al.*, 1984; Riggle & Klos, 1970). One of the reasons that *E. herbicola* can survive on the plant surface is probably the presence of pigments (carotenoids) in the cells. Laboratory study has shown that *Escherichia coli* transformed with the carotenoid synthesis genes (*crt*) of *E. herbicola* is more resistant to near-UV irradiation and to the phototoxic agents than the cells without the genes, suggesting that

synthesis of carotenoids by *E. herbicola* may be essential for the survival of this organism in nature (Tuveson *et al.*, 1988).

Carotenoids are synthesized and are biologically important in many organisms ranging from bacteria and fungi to higher plants (Bartley *et al.*, 1991, 1992; Carattoli *et al.*, 1991; Misawa *et al.*, 1990; Perry *et al.*, 1986). Many of the *crt* genes from these organisms have been cloned and studied (Bartley *et al.*, 1991, 1992; Carattoli *et al.*, 1991; Hundle *et al.*, 1992, 1993; Lee & Liu, 1991; Math *et al.*, 1992; Sandmann & Misawa, 1992; Shnurr *et al.*, 1991). Previous studies have shown that carotenoid synthesis in *E. herbicola* is encoded by plasmid (Gantotti & Beer, 1982; Thiry, 1984) or chromosomal (Lee & Liu, 1991; Misawa *et al.*, 1990; Perry *et al.*, 1986) genes. The genes encoding the biosynthesis of carotenoids in *E. herbicola* Eho10 are clustered in a 12.4 kb fragment. Genetic studies revealed that the expression of these genes requires cAMP (Perry *et al.*, 1986). In contrast, expression of carotenoid genes in *E. herbicola* Eho13 and in *Erwinia uredovora* 20D3 does not have this requirement (Lee & Liu, 1991; Misawa *et al.*, 1990). The Eho13 genes are clustered in a 7.0 kb chromosomal fragment (Lee & Liu, 1991). Tn1000-mutagenesis and restriction-deletion studies showed that this gene cluster contains at least three different regions.

[†] **Present address:** Department of Genetics, Harvard Medical School, and Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA.

[‡] **Present address:** Department of Plant Pathology, University of California, Davis, CA 95616, USA.

The GenBank accession number for the sequence reported in this paper is M90698.

Table 1. Bacterial strains and plasmids used in this study

Strain/plasmid	Characteristics	Source
<i>E. herbicola</i>		
Eho13	Wild-type, isolated from apple buds	This laboratory
<i>E. coli</i>		
HB101	F ⁻ <i>hsd20</i> (<i>r_B</i> ⁻ , <i>m_B</i> ⁻), <i>supE44 recA13 ara14 proA2 rpsL20</i> (Str ^R) <i>xyl-5 mtl-5 λ</i> ⁻	Boyer & Roulland-Dessoix (1969)
CSH50	F ⁻ , <i>ara Δ(lac-pro) strA thi</i>	Miller (1972)
Plasmid		
pKM005	A transcription-fusion plasmid containing a promoterless <i>lacZ</i>	Masui <i>et al.</i> (1983)
pSL525	A plasmid clone containing the entire <i>crt</i> operon of <i>E. herbicola</i> Eho13; 9.7 kb; orange-coloured phenotype	Lee & Liu (1991)

Insertion or deletion in one of these regions could result in mutants expressing different phenotypes, suggesting that multiple genes are involved in carotenoid synthesis (Lee & Liu, 1991). Genetic study also showed that a pink mutation of Eho10, which had an *AvaI* deletion in the gene cluster, could be complemented by a non-pigmented Eho13 deletion mutant and resulted in the synthesis of orange-coloured carotenoids, suggesting that the biochemical pathway involved in carotenoid biosynthesis in these two strains may be similar (Lee & Liu, 1991). Recently, genes encoding carotenoid synthesis (*crt*) in *E. uredoovora* 20D3 have been cloned and sequenced (Misawa *et al.*, 1990). The cloned fragment consists of six genes, *crtE*, *crtB*, *crtI*, *crtY*, *crtZ* and *crtX*, which are responsible for the synthesis of geranylgeranyl pyrophosphate, phytoene, lycopene, β -carotene, zeaxanthin and zeaxanthin- β -diglucoside, respectively (Misawa *et al.*, 1990; Sandmann & Misawa, 1992). The functions of several *crt* genes in *E. herbicola* Eho10 have also been determined (Hundle *et al.*, 1992, 1993; Math *et al.*, 1992). In this article, we report the sequence of the DNA fragment containing *crtEXYIB* of *E. herbicola* Eho13 and the location of the promoter responsible for the transcription of these *crt* genes.

METHODS

Bacterial strains, plasmids and growth conditions. Bacterial strains and plasmids used in this study are listed in Table 1. LB agar and LB broth (Miller, 1972) were used as general-purpose media. Ampicillin at 100 μ g ml⁻¹ was used for selection (LB-Ap).

DNA preparation. Plasmids were isolated and purified according to the CsCl centrifugation method described by Perry *et al.* (1986). Plasmid screening was carried out according to the method of Kado & Liu (1981).

Probe labelling. Oligonucleotide primers (500 ng) were end-labelled with 50 μ Ci (1.85 MBq) [γ -³²P]ATP (5000 Ci mmol⁻¹) (185 TBq mmol⁻¹), Amersham] and 16 U T4 polynucleotide kinase (Promega) according to the method described by Sambrook *et al.* (1989). Labelled primers were purified with a Sep-Pak C18 cartridge (Millipore).

DNA sequencing. Nucleotide sequences of both strands were

determined by the dideoxy chain termination method (Sanger *et al.*, 1977), using supercoiled plasmids as templates. A T7 DNA polymerase sequencing kit (Promega) and [α -³⁵S]dATP [1000 Ci mmol⁻¹ (37 TBq mmol⁻¹), Amersham] were used for DNA sequencing. DNA fragments were subcloned into pGEM vectors (Promega) or pUC vectors (Yanisch-Perron *et al.*, 1985). DNA sequencing was carried out by primer walking. PCR-direct sequencing was carried out according to the method of Krishnan *et al.* (1991). Primers were synthesized with a Cruchem Model PS250 DNA Synthesizer.

Fusion studies. Transcription fusion was carried out with a fusion vector pKM005 (Masui *et al.*, 1983) in *E. coli* CSH50 (Miller, 1972). Clones containing promoter activity were selected on MacConkey Agar (BBL) containing ampicillin.

RNA isolation and primer extension. Cells were cultured in 100 ml of LB-Ap broth to mid-exponential phase. Total RNA was isolated from cells according to the acid-phenol method of Aiba *et al.* (1981). RNA was precipitated with ethanol and suspended in TE (10 mM Tris/HCl, pH 7.6; 1 mM EDTA). The sequences of the primers were: *P1* (5' ATTTATGACA-ACGCTAAAAGTGCGATGCCC 3'); *P2* (5' CAGGCA-ATAACTGCTCAAGC 3'); and *P3* (5' CGGCTGATCGG-TTTGAAAAC 3'). These primers were complementary to the sequences between nucleotides 271–300 (*P1*), 403–422 (*P2*) and 5325–5344 (*P3*) (Fig. 1). Approximately 10⁸ c.p.m. of labelled primer was annealed with 100 μ g total bacterial RNA. Primer annealing was carried out in a solution containing 34 mM Tris/HCl (pH 8.2), 50 mM NaCl, 5 mM MgCl₂ and 5 mM DTT. The annealing mixture was incubated at 85 °C for 5 min and then at 42 °C for 12–16 h. Primer extension was carried out with 100 U AMV reverse transcriptase (Promega) and 40 U rRNasin (Promega) at 42 °C for 90 min. The cDNA synthesized was precipitated by ethanol and was analysed in an 8% (w/v) acrylamide sequencing gel (Sambrook *et al.*, 1989).

PCR. RNA PCR was carried out with AMV reverse transcriptase (Promega) and with *Taq* DNA polymerase (Promega) according to the method previously described (Liu *et al.*, 1991).

RESULTS AND DISCUSSION

Nucleotide sequence of the *E. herbicola crt* operon

We have previously cloned the *crt* genes of *E. herbicola* Eho13 (Lee & Liu, 1991). The *E. coli* clones containing these genes produced an orange colour. The DNA

GAAATTTGCTAAAGACACGGCGCTCAGTGCTAATGACGTTTCTCAGGTAATTGCGACATTACGTGCACTGGCAAGCTGAACCCCGGCGCTCAGGAAAAAC 100
 CTGTCACGCCACAGGCAATACGACGGAGGTTTATGAATTCACACCTTTTCCACAGGCTCAACTAGTATTAATTATGTGGATTAGCGAGAAAAACCT 200
 crtP1 -35 -10 +1
 GAAACCTTAATGTAAAAATAACGGGTAAGCCTGCCAACCATGTTATGCGAGATTAAGCGTTTCTTTAAAGGGCATCGCACTTTTAGCGTTGTCATAAAT 300
 GTATCCGTTGATAAGGACAGCCCGAATGACGGTCTGCGCAAAAAACACGTTCAACCCACTCGCAGTGTCTGCTGAGCAATTACTGGCTGATATTGATCGG 400
 M T V C A K K H V H P T R S A A E Q L L A D I D R
 RBS >crtE
 CGGCTTGAGCAGTTATTGCTGTCGAGGGAGAACGGGATTTGTAGGTGCCGCGATGCGTGAGGGCGCGCTGGCACCGGGTAAACGTTATCGCCCGATGC 500
 R L E Q L L P V E G E R D F V G A A M R E G A L A P G K R I R P M
 TGTGTTGCTCACAGCAGCGATCTGGCTGCGCTGTGACCCAGGAGGATTGCTGGATATCGCCTGCGCGGTGGAATGTGTCACGCGGCATCGTTGAT 600
 L L L L T A R D L G C A V S H E G L L D I A C A V E M V H A A S L I
 TCTTGATGATATGCCCTGTATGGATGATGCGCAGATCGTGGCGGGCGCGCTACCGTTCACTGTCTGAGTACGGCGAGCATGTGGCAATCTGCGCGGGCTC 700
 L D D M P C M D D A Q I V G G R P T V H C Q Y G E H V A I L A A V
 GCCTTACTGAGCAAGCCCTTTGGGGTCATTGCTGATGTCAGATGGCCTGACCGCGCTGGCAAAAAACCGCGAGTGTGAGAATTATCAACGCCATTGGTA 800
 A L L S K A F G V I A D A D G L T A L A K N R A V S E L S N A I G
 TGCAAGGTTTGGTACAGGGTCAGTTAAGGATCTGTGAGAAGGCGATAAGCGCGCAGCGCTGAAGCCATTTTGATGACGAACCACTTTAAACAGCAC 900
 M Q G L V Q G Q F K D L S E G D K R R S A E A I L M T N H F K T S T
 CCTGTTTTCGCGTCTATCAGATGGCCTCAATTGTGGCAATGCTCCAGCAACGACGCGATTACCTACACCGTTTTTCGCTTGATCTCGGCCAAGCG 1000
 L F C A S M Q M A S I V A N A S S E R R D Y L H R F S L D L G Q A
 TTTCAACTGCTGGACGACTTAACGATGGCATGGCTGACACGGGTAAGGACATCAATCAGGATGAAGCAAAATCGACCCCTGGTCAATTTATTAGGCTCCA 1100
 F Q L L D D L T D G M A D T G K D I N Q D E G K S T L V N L L G S
 GAGCAGTTGAAGAGCGTCTGCGGCACAGTCTGCATCTCGCCAGTGAGCACTCTCAGCGCGTGTCAAAACGGTCATTCTACTCAACAATTTATCCAGGC 1200
 R A V E E R L R H S L H L A S E H L S A A C Q N G H S T Q Q F I Q A
 CTGTTTGACAAAAAAGCTCGCTGCCGTCAGTTAAGGATGCTGCATGAGCCACTTCGCGCGATCGCTCCGCTTTTACAGCCATGTCGCTGCACTGGAA 1300
 W F D K K L A A V S M S H F A A I A P P F Y S H V R A L E
 crtE< >crtX
 RBS
 AACCTCGCTCAGGAGTTAGTGGCTCGCGGTCACTAGGTGACCTTTATTAGCAACAGGACACTAAACCACTCATTGAGCATCCCCACATCGGATTTTCATG 1400
 N L A E L V A R G H A D V T F I Q Q D T K P L I Q H P H I G F H
 CTGTAGGAGAAGAAAGCCATCCGCTCTGCTCATTATCGCGCTGTGCACTTAGCGGCCCTCTCTGGGCGCTCAATGCTTCGCGCTCATATGAAT 1500
 A V G E E S H P S G S L S R L L H L A A H P L G P S M L R L I N E M
 GCGCGCACAGCGATATGCTGTGCTGCGCTCCCCAGAGCCTTTAAGCATTTGGCGATAGATGGTGTATCGTTGATCAGATGGAGCCAGCAGCGCT 1600
 A R T S D M L C L A L P R A F N D L A I D G V I V D Q M E P A G A
 CTGGTCGCTGAAGCGCTGGGCTGCGGTTGTCTCGGTTGCTGTGCACTCCCGCTTAATCGCAACCGGATTTGCCCTGGCGGTGATGCTTTTGACT 1700
 L V A E A L G L P F V S V A C A L T L N R E P D L P L A V M P F D
 ATGGCACCGCATGCGCGAGTGTGAGCGCTATGCAAGCAGCGAAAAAATTTTACTGCGTGTGCGTGGCATGACCGCGTCAATGCGCATCAGTCA 1800
 Y G T S D A A R E R Y A A S E K I Y D W L M R R H D R V I A H H S H
 CAGGATGGGCTTAGCACCCCGGATAAACTCCATCAGTGTTTTTCGCCATTGGCACAAATCAGTCACTTATCCAGAAGTGGATTTCCCGGTAAAGCG 1900
 R M G L A P R D K L H Q C F S P L A Q I S Q L I P E L D F P R K A
 CTACCGGCTTGTTTTCATGCTGTGCGCCCTGCGAGAAACGACAGACATCAACCTCTTCGCGCTTTATTTTCGCGCTCGAAAAGCCCGCAATTT 2000
 L P A C F H A V G P L R E T Q T T S T S S P L Y F S P S E K P R I
 TCGCTCTGCTGGGACGTTGACAGGTCATGCTACGGTCTGTTTAAAGCCATCGTGAAGCCCTGTGAAGAGATTGATGGTCAGCTTCTGTGGCGCACTG 2100
 F A S L G T L Q G H R Y G L F K A I V K A C E E I D G Q L L L A H C
 TGGCGCGCTGACTGCTTTTTCAGGCGAAGAGCTGGCGGAAGCGCTCACACGAGGTTGTGGATTTTGCAGTCACTGCGCGCGCTGTCGAGGCGCAG 2200
 G R L T A F Q A E E L A R S R H T Q V V D F A D Q S A A L S Q A Q
 CTTGCCATTACCCATGGCGGGATGAATACGGTGTGATGCCATTAATATCGCACGCGCTTTTAGCCCTTCGTTGGCATTTGATCAGCCCGCGTGG 2300
 L A I T H G G M N T V L D A I N Y R T P L L A L P L A F D Q P G V
 CCTCAGTATTGTTTCTCATGGGATAGGAAAACGTGCGTCCGCTTTTACCACAGCCATGCGTGGCGCGCCACATGCGTGCATTGCTGACCAACGTCGA 2400
 A S R I V S H G I G K R A S R F T T S H A L A R H M R A L L T N V D
 CTACCGCAACGCATGACGAAAATCCAGACCGCCTTCGTTTAGCGCGCGCACAAATGGCGCTGCAGATATCGTTGAGCAGGCATGCGACCGCGCAG 2500
 Y P Q R M T K I Q T A L R L A G G T M A A A D I V E Q A M R T G Q

Fig. 1. For legend see page 335.

fragments containing these genes were subcloned into pGEM3 or pUC18 and the sequences of these fragments were determined by primer walking. DNA of both strands

was sequenced. The results showed that the entire sequence was 6965 bp long. Of these nucleotides, 6365 are shown in Fig. 1. The sequence contained five ORFs which

```

CCCCTTTTGAAGGGGAGCGGCTATGCGACCGCATTATGATCTGATTTTAGTGGGCGCGCGGCTGGCTAATGGCCTCATTGCCCTGCGCCTTCAGCAGCAG 2600
      M R P H Y D L I L V G A G L A N G L I A L R L Q Q Q
      RBS      >crtY
P V L K G S G Y A T A L
      crtX<
CAACCTGACATGCGTATTCTGCTTATCGAAGCGGCCCCAGGCGGGCGGTAATCATACCTGGTCTTTTACGACGCTGATTAAACGAGAGCCAACATC 2700
Q P D M R I L L I E A A P Q A G G N H T W S F H D A D L T E S Q H
GCTGGGTAGCACCGCTGGTGGTATATCACTGGCCCGACTATCAGGTCCGCTTTCCACGCGCGCTCGTAAGCTGAACAGCGGCTATTTTCTGTGACTTC 2800
R W V A P L V V Y H W P D Y Q V R F P T R R R K L N S G Y F S V T S
GCACGCTTTTGGCGAGGTTTACAAACAGAAGTTTGGTCAGCACTTATGGATTAGCCGCGCGGTGGCAGAGGTTACGCGCGACGCTGTTTCTGACTGAATAAT 2900
Q R F A E V L Q K F G Q L H L W I S R A V A E V H A D A V R L N N
GGTCAGGTATCAGTGCCAGCGCGGTGATTGATGGGCGAGGTTATACGCCAAATTACGCGCTGAACGTTGGGATTCCAGCGGTTTATCGGACAGGAATGGC 3000
G Q V I S A S A V I D G R G Y T P N S A L N V G F Q A F I G Q E W
GACTCAGTAAACCGCAGCGTTTATCTTACCCATTATTATGGATGCCACGGTCGATCAGCAAAATGGTTATCGCTTTGTTTACAGCCTGCCGCTATCTGC 3100
R L S K P H G L S S P I I M D A T V D Q Q N G Y R F V Y S L P L S A
GACAGAATCTTAATCGAAGATACCCACTATATCGATAACGCGACACTGGAACCTGAACGCGCGCAAAATATTCGCGATTATGCTGCCAGCAGGAT 3200
T E L L I E D T H Y I D N A T L E P E R A R Q N I R D Y A A Q Q D
TGGCAGCTTCAGACTCTGCTTCGTGAGGAACAGGGCGCATTGCCAATAACCTTAACGGCGATAGCACCGGTTTGGCAGCAACAGCCTTTAGCCTGCA 3300
W Q L Q T L L R E E Q G A L P I T L T G D S T A F W Q Q Q P L A C
GTGGACTGCGTGCCGACTGTTTCAACCCCAACCGGTTATTCACTCCCGCTTGGCGGTGGCGGATCGCCTGAGTGGCGCTTGATGCTTTACATC 3400
S G L R A P L V H P T T G Y S L P L A V A L A D R L S A L D V F T S
GTCTCAATTATCATCAGCGATTACCCACTTTGCCACGAGCGCTGGCAGCAGCAACGCTTTTCCGCTGTTGAATCGCATGCTGTTTATAGCGGGACCC 3500
S S I H Q A I T H F A H E R W Q Q Q R F F R M L N R M L F L A G P
GCCGATTACGCTGGCGTGTATGCGCGCTTCTATGGTTTACCTGAAGATTAACTCTCCGTTTATGCGGGCAAACTCACGCTGACCGATCGGCTAC 3600
A D S R W R V M Q R F Y G L P E D L I S R F Y A G K L T L T D R L
GTATTCTGAGCGGCAAGCCGCTGTTCCGCTATTAGCGGCTTGCAGGCAATTATGACGACTCATCGTTAAAGAGCGCAACATGAACCACTACGGTA 3700
R I L S G K P P V P V L A A L Q A I M T T H R M K P T T V
      crtY< RBS      >crtI
ATTGGTGCAGGCTTTGGTGGCTGGCATTAGCAATTCTGCTGCGCGCGGGGATCCCTGTCTTACTGCTTGAGCAACGCGCAAAACCGGTGGCGGG 3800
I G A G F G G L A L A I R L O A A G T P V L L L L E Q R D K P G G R
CTTATGCTATGAAGATCAGGGAATCACTTTTGTATGCGGGCTACCGTTATTACCGATCCCGCGCATTGAAGAACTCTTACCCTGGCGGGAACA 3900
A Y V Y E D Q G F T F D A G P T V I T D P S A I E E L F T L A G K Q
GTTAAAGATTACGTTGAAGTCTGCGCGTGGCGGCTTTTATCGCTGTTGGGAGTCAGGAAAGTTTCAATTACGATAACGATCAGGCACAGCTT 4000
L K D Y V E L L P V A P F Y R L C W E S G K V F N Y D N D Q A Q L
GAGGCGCAGATTACGAGTTAATCCACGCGATGTTGAAGGCTATCGTCAGTTTCTGGACTATTACGTCGCGGTGTTAAAGAAGGCTATCTGAAGCTTG 4100
E A Q I Q Q F N P R D V E G Y R Q F L D Y S R A V F K E G Y L K L
GCACCGTGCTTTTCTGCTTTCAGGATATGCTTCGCGCGCGCCCACTGGCGAAACTGCAGGATGCGGAACCGTTTACAGTAAAGTTGCCAGCTA 4200
G T V P F L S F R D M L R A A P Q L A K L Q A W R T V Y S K V A S Y
CATTGAGGATGAACATCTTCGTCAGGCGTTTCTTTTCACTCGCTGTAGTGGGCGGCAATCCCTTCGCCACCTCCTCATTATACGCTAATACATGCA 4300
I E D E H L R Q A F S F H S L L V G G N P F A T S S I Y T L I H A
CTCGAAGCGAATGGGCGCTCTGTTTCCACGCGGTGTTACCGGGCATAGTAAAGGATGATAAAGCTGTTTCAGGATCTGGGTGGCGAAGTGGTAC 4400
L E R E W G V W F P R G G T G A L V K G M I K L F Q D L G G E V V
TGAATCGGAAGGTGAGCCATGGAACACCGGGGATACCATTAAGCGGTGCAATTAGAGGACGGACGAGATTCCCGACCCGGGCTGTCGCTTCCAA 4500
L N A K V S H M E T T G D T I E A V H L E D G R R F P T R A V A S N
TGCGGATGTGGTTCACACCTATCGCGACCTGTTAAGTCAGCATCTGCGCGGTGAAGCAGTCGAAAAACTGCAAACTAAGCGCATGAGCAACTCATTG 4600
A D V V H T Y R D L L S Q H P A A V K Q S K K L Q T K R M S N S L
TTTGTCTCTATTTTGGTCTGAATCACCATCAGATCAGCTTGCCACCATACGGTTTGTTCGCGCTCGCTACCGTGAGCTGATTATGAAATTTCA 4700
F V L Y F G L N H H H D Q L A H H T V C F G P R Y R E L I H E I F
ATCATGACGACTCGCAGACGATTCTCACTTTACCTGCACGCACTTGTGTACGGATTTCGTCAGTGGCACCGGAAGGTTGCGGCGATTACTATGTGT 4800
N H D G L A D D F S L Y L H A P C V T D S S L A P E G C G S Y Y V L
AGCTCCCGTGCCACATTAGGCACAGCGAAGCTCGACTGGACGGTTGAAGGGCGCGAGTCGCGATCGTATTTTGTAGTACCTTGAGCAACATTACATG 4900
A P V P H L G T A N L D W T V E G P R L R D R I F E Y L E Q H Y M
CCTGGCTTCGCTAGCAGTTGGTTACGCGAGGAATGTTACGCCGTTTCGATTTCGCGACCGCTCAATGCCTATCAAGGCTCAGCCTTTTCTGTTGAGC 5000
P G L R S Q L V T Q R M F T P F D F R D Q L N A Y Q G S A F S V E

```

Fig. 1—cont. For legend see facing page.

were transcribed in one direction. These genes were organized in the order *crtE-crtX-crtY-crtI-crtB* based on sequence homology with the *crt* genes of *E. uredovora*

(Misawa *et al.*, 1990). The level of DNA homology of the *crt* genes between *E. herbicola* Eho13 and *E. uredovora* is higher than that between the *crt* genes of *E. herbicola*

CCGTTCTCACAGAGCGCTGGTTCCGACCGCACACCGCGATAAAACCATCAATAATCTTTACCTGGTAGGTGCAGGCACCCATCCCGCGCGGGCAT 5100
 P V L T Q S A W F R P H N R D K T I N N L Y L V G A G T H P G A G I
 TCCCGCGTGATCGGCTCGGCTAAGGCAACTGCAGGTTTAAATGCTGGAGGACCTGATTGAAATAATCGCTCGTTACTCAATCATCGCGTCGAAACGATGG 5200
 P G V I G S A K A T A G L M L E D L I
 crtI< RBS >crtB
 CTGTTGGCTCGAAAGTTTTCGACTGCCTCAAAGTTATTTGACGCAAAACCCGCGCAGTGTACTTATGCTCTACTCCTGGTGCCCATTTGTGACGA 5300
 crtP2 -35 -10 +1
 A V G S K S F A T A S K L F D A K T R R S V L M L Y S W C R H C D D
 CGTGATTGACGATCAAAACCTGGGTTTCAAACCGATCAGCGCTATTACAAACGCTGAGCAACGCTGATGCAGTTGGAGATGAAACGCGCCAGGCC 5400
 RBS >crtB'
 V I D D Q T L G F Q T D Q P S L Q T P E Q R L M Q L E M K T R Q A
 TATGCCGATCGCAGATGCAGAACCGGCTTTGCTGCTTTTCAGGAAGTGGCGATGGCCCATGATATTGCGCCGCTTACGCGTTTGATCATCTGGAAG 5500
 Y A G S Q M H E P A F A A F Q E V A M A H D I A P A Y A F D H L E
 GCTTTGCCATGGATGTGCTGAGGCGGAGTACATTGAGTGGACGATACATTACGCTATTGCTATCAGCTTGCAGGCGTGGTTCGCTTGATGATGGCGCA 5600
 G F A M D V R E A E Y I Q L D D T L R Y C Y H V A G V V G L M M A Q
 GATTATGGGCTACGCGATAACGCTACACTGGACCGCGCTGTGACCTTGGGCTGGCATTTCAGTTGACCAACATCGCCCGGATATTGTGGAAGATGCG 5700
 I M G V R D N A T L D R A C D L G L A F Q L T N I A R D I V E D A
 CATGTGGGCTGTTGTTATCTGCCGCGAGTTGGCTGGAAATGAAGGACTGAATAAAGAAAATTATGCGGCACCGGAAACCGCCAGGCGCTGAGCCGCA 5800
 H V G R C Y L P A S W L E N E G L N K E N Y A A P E N R Q A L S R
 TCGCCGACGCTGCTGGTGCAGGAAGCCGAACCTTACTATTGTCGCTACGCGGGGACTGGCGGGGCTGCCCTTGCGCTCAGCCTGGGCCATCGCCACAGC 5800
 I A R R L V Q E A E P Y Y L C A T A G L A G L P L R S A W A I A T A
 GAAGCAGGTTTATCGTAAATTTGGGTGAAGTTGAACAGGCAGGAACAGAAGCCTGGGACCATCGGCAGTCAACCACGACCCGAAAAATATCGCTT 6000
 K Q V Y R K I G V K V E Q A G T E A W D H R Q S T T T P E K L S L
 CTGTTGATGGCATCCGCTCAGGCCATTACTTCCCGGATGCGGCCTCATCCGCTCGCCCTGCCCCTCTCTGGCAGCGCCGATTAGCCGAATGTGCTTC 6100
 L L M A S G Q A I T S R M R P H P P R P A H L W Q R P I
 crtB, crtB'<
 TCTTAGCGTCGCTGAAGTTTGTATAGCGGCGCGCATAGAGAAAACAAAGAAACACAGCCTTCTTTGCTCTTACGGCATGATGCATGCGGTGGGCC 6200
 ATATACAACCGCTTCAGGTAGCCTTTGCGCGGAATATCCGGCAACGGCAACGCTGATGCACAGCCGCTCGTGAACCATGAAATAGAGCAATCCATAGG 6300
 CCGTCATGCCCGCAATCCACTGAAGCGGCCATTCCCGTACTGCCACATAAATAAGCAGG

Fig. 1. Sequence of the *crt* operon of *E. herbicola* Eho13. The transcription start sites are indicated by ' + 1'. The name for each gene in the *crt* operon was designated according to Misawa *et al.* (1990). Ribosome-binding sequences (RBS), the –35 region, the –10 region and the peptide sequence representing dinucleotide-binding folds in *crtI* are underlined.

Table 2. Characteristics of the *crt* genes of *E. herbicola* Eho13

Gene	Length (bp)	Molecular mass of peptide (Da)	Percentage homology with:			
			<i>E. uredoovora</i> 20D3*		<i>E. herbicola</i> Eho10†	
			DNA	Peptide	DNA	Peptide
<i>crtE</i>	909	32624	86	91	55	52
<i>crtX</i>	1296	47076	83	87	54	47
<i>crtY</i>	1149	43254	85	89	58	58
<i>crtI</i>	1479	55018	88	96	67	77
<i>crtB</i>	891	33193	87	94	62	64

* Misawa *et al.* (1990).

† GenBank no. M87280.

strains Eho10 and Eho13 (Table 2). Computer analysis revealed that the proteins translated from these genes contained high percentages of hydrophobic amino acids dispersed throughout the sequences, suggesting that these were membrane-associated proteins. Similar to the *crtI* gene of *E. herbicola* Eho10, Eho13 *crtI* also contains a 29

amino acid region of dinucleotide-binding folds (Fig. 1) (Armstrong *et al.*, 1990). The characteristics of these five genes are summarized in Table 2.

The enzymes encoded by these genes were: *crtE*, geranyl-geranyl pyrophosphate synthase (Sandmann & Misawa,

1992); *crtX*, 3-hydroxy- β -carotene glycosylase (Shnurr *et al.*, 1991; Hundle *et al.*, 1992); *crtY*, lycopene cyclase (Shnurr *et al.*, 1991; Hundle *et al.*, 1992); *crtI*, phytoene desaturase (Misawa *et al.*, 1990); and *crtB*, phytoene synthase (Sandmann & Misawa, 1992). Among the five genes in the cluster, the protein encoded by *crtX* is a 47076 Da enzyme. The function of this enzyme is to convert zeaxanthine to zeaxanthine- β -diglucoside (Shnurr *et al.*, 1991). Previous study has shown that the precursor of zeaxanthine was β -carotene and the synthesis of zeaxanthine required β -carotene hydroxylase (Shnurr *et al.*, 1991). This enzyme is encoded by *crtZ* in *E. uredoovora* (Misawa *et al.*, 1990). However, the *crtZ* gene was not present in the *crt* cluster that we have cloned. Therefore, *E. coli* HB101(pSL525) cannot synthesize zeaxanthine- β -diglucoside due to the lack of β -carotene hydroxylase in the cells. Therefore, the orange phenotype produced by *E. coli* HB101(pSL525) was due to the presence of β -carotene instead of zeaxanthine- β -diglucoside. The *crt* gene cluster that we have cloned also contained a 340 bp region downstream from *crtB* which is 87% homologous to the 3' end of the *E. uredoovora crtZ* gene (Misawa *et al.*, 1990), suggesting that the sequence encoding the N-terminal end of β -carotene hydroxylase was probably deleted during the subcloning process. In other words, the complete *crt* gene cluster of *E. herbicola* Eho13 may contain six genes and may be organized in a similar manner as the *crt* gene cluster of *E. uredoovora*.

Identification of the regions containing promoter activity

The regions containing promoter activity in the *crt* gene cluster were identified by cloning the DNA fragments into a promoter-probe vector pKM005 (Masui *et al.*, 1983). The red clones on MacConkey agar were selected. Results showed that two different regions in this gene cluster gave promoter activity. One was the 2.5 kb *Tth1111* fragment which contains the region upstream from the first gene of the gene cluster, *crtE*. The second region is the 2.6 kb *HindIII* fragment which is located toward the 3' end of the gene cluster (Fig. 2). Further subcloning experiments showed that the promoter activity was located in the 0.42 kb *EcoRI*-*EcoRV* fragment and in the 1.7 kb *PstI*-*HindIII* fragment (Fig. 2).

Mapping of the transcription start sites

Transcription start sites in the *crt* cluster were determined by primer extension. The results revealed that the 5' terminus was located at nucleotide position 178 (Figs 1 and 3a). The predicted -10 and -35 sequences (*crtP1*) are shown in Fig. 1. A second primer, *P2*, which is complementary to nucleotides 403-422 (Fig. 1), also confirmed this transcription start site. However, the signal of this cDNA product was weak due to inefficient synthesis of long cDNA fragments by AMV reverse transcriptase and was detectable only after a long exposure of the film (result not shown). We also detected an additional 5' terminus at nucleotide 300 with primer *P2*

Fig. 2. Transcription fusion study. DNA fragments encoding the carotenoid synthesis in pSL525 were subcloned into a fusion vector, pKM005 (Masui *et al.*, 1983). The clones containing promoter activity were selected on MacConkey Agar. Numbers indicate nucleotide positions. The 3' portion of *crtX*, the 5' portion of *crtI*, and *crtY* of the *crt* gene cluster are not shown. Promoter activity was found in a 0.42 kb *EcoRI*-*EcoRV* fragment (a) and in a 1.7 kb *PstI*-*HindIII* fragment (b). A, *AvaI*; H, *HindIII*; P, *PstI*; RI, *EcoRI*; RV, *EcoRV*; T, *Tth1111*.

(Fig. 3b). This location is probably not a true transcription start site, since a 134 bp PCR fragment (nucleotides 188-321) which includes this start site did not generate a transcription fusion with *lacZ* in pKM005. This terminus probably represents a specific RNA-processing site. Since our gene fusion study revealed that the 1.7 kb *PstI*-*HindIII* fragment contained promoter activity, we have used computer program TARGSEARCH to search for the potential promoter sequence in this fragment. The results revealed a potential promoter in the region of nucleotide 5240-5270 as shown in Fig. 1. We used primer *P3* (complementary to nucleotides 5325-5344) to detect the potential transcription start site in this region. The results revealed that this terminus was located at nucleotide 5275 (Figs 1 and 3c). The predicted -10 and -35 sequences (*crtP2*) are shown in Fig. 1. Computer analysis showed that an ORF, which is in-frame and overlaps with *crtB*, starts at nucleotide 5371 and ends at 6087. A putative ribosome-binding sequence was found 10 bp upstream from the initiation codon ATG (Fig. 1). If this ORF, *crtB'*, indeed encodes a protein, this protein is 58 amino acids shorter, at the N-terminal end, than the protein encoded by *crtB*. Our previous minicell experiment did not detect a protein with a size similar to the protein encoded by *crtB'* (Lee & Liu, 1991). There was a possibility that this protein was translated at a low level or not translated in the cells.

Fig. 3. Mapping of the transcription start sites of the *crt* cluster. Primer extensions were carried out with primers *P1* (a), *P2* (b) and *P3* (c) (see Methods). The same primers were used for DNA sequencing. The letters above each lane indicate the dideoxynucleotide used to terminate each reaction. DNA sequencing with primer *P3* was carried out by the PCR-direct sequencing method (Krishnan *et al.*, 1991). The extension products were analysed on an 8% sequencing gel. * indicates the 5' terminus of RNA.

Fig. 4. PCR amplification of the *crt* transcript. RNA PCR was carried out with primer sets covering the ranges of nucleotides 217–832 (lanes 1, 2), 1074–1447 (lanes 3, 4), 2284–2580 (lanes 5, 6), 3433–3717 (lanes 7, 8) and 4731–5696 (lanes 9, 10). Lane M, pGEM molecular mass marker (Promega). The RNA sample was treated with DNase-free RNase (Promega) (lanes 2, 4, 6, 8, 10) prior to PCR or left untreated (lanes 1, 3, 5, 7, 9). PCR was carried out for 30 cycles under the following conditions: 1 min of denaturation at 94 °C, 2 min of annealing at 50 °C, and 2 min of extension at 72 °C.

PCR analysis of the *crt* transcript

We have used a PCR method to determine whether the five genes in the gene cluster were organized in an operon. Each primer pair was designed to demonstrate that the adjacent genes were actually linked in an RNA molecule. A total of five primer pairs was able to amplify the RNA molecules containing the junctions: non-translated

Fig. 5. The function of *crt* genes of *E. herbicola* Eho13. The enzyme encoded by each of the genes is indicated in parentheses.

region—*crtE*, *crtE*—*crtX*, *crtX*—*crtY*, *crtY*—*crtI* and *crtI*—*crtB* (Fig. 4), suggesting that the five *crt* genes may be organized in an operon in the order *crtE*—*crtX*—*crtY*—*crtI*—

crtB and are transcribed from the promoter *crtP1*. The amplified products were indeed derived from RNA, since treatment of samples with DNase-free RNase failed to give any PCR products (Fig. 4).

Conclusion

The gene cluster encoding the biosynthesis of carotenoids in *E. herbicola* Eho13 may contain a total of six genes. Among these genes, *crtE*, *crtX*, *crtY*, *crtI* and *crtB* are transcribed in one direction and are probably organized in an operon. The enzymes encoded by these genes and the biochemical reactions catalysed by these enzymes are shown in Fig. 5. The results presented in this paper should provide useful information for the understanding of the mechanism of carotenoid biosynthesis.

ACKNOWLEDGMENTS

We want to thank T. Simonitch and K. Harrison-Lavoie for their technical assistance. This research was supported by a medical research grant CMRP 348 from the Chang-Gung Memorial Hospital and a biological research grant NSC 81-0412-B-182-12 from the National Science Council of the Republic of China. K.Y.T. was a recipient of a NSC post-doctoral fellowship. E.M.L., L.Y.L., T.P.L. and C.L.C. were recipients of NSC graduate-student fellowships.

REFERENCES

- Aiba, H., Adhya, S. & de Crombrughe, B. (1981). Evidence for two functional *gal* promoters in intact *Escherichia coli* cells. *J Biol Chem* **256**, 11905–11910.
- Armstrong, G. A., Alberti, M. & Hearst, J. E. (1990). Conserved enzymes mediate the early reactions of carotenoid biosynthesis in nonphotosynthetic and photosynthetic prokaryotes. *Proc Natl Acad Sci USA* **87**, 9975–9979.
- Bartley, G. E., Viitanen, P. V., Pecker, I., Chamovitz, D., Hirschberg, J. & Scolnik, P. A. (1991). Molecular cloning and expression in photosynthetic bacteria of a soybean cDNA coding for phytoene desaturase, an enzyme of the carotenoid biosynthesis pathway. *Proc Natl Acad Sci USA* **88**, 6532–6536.
- Bartley, G. E., Viitanen, P. V., Bacot, K. O. & Scolnik, P. A. (1992). A tomato gene expressed during fruit ripening encodes an enzyme of the carotenoid biosynthesis pathway. *J Biol Chem* **267**, 5036–5039.
- Beer, S. V., Rundle, J. R. & Norelli, J. L. (1984). Recent progress in the development of biological control for fire blight – a review. *Acta Hort (The Hague)* **151**, 195–201.
- Billing, E. & Baker, L. A. E. (1963). Characterization of *Erwinia*-like organisms found in plant material. *J Appl Bacteriol* **26**, 58–65.
- Boyer, H. W. & Roulland-Dessoix, D. (1969). A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. *J Mol Biol* **41**, 459–472.
- Carattoli, A., Romano, N., Ballario, P., Morelli, G. & Macino, G. (1991). The *Neurospora crassa* carotenoid biosynthetic gene (albino 3) reveals highly conserved regions among prenyltransferase. *J Biol Chem* **266**, 5854–5859.
- Gantotti, B. V. & Beer, S. V. (1982). Plasmid-borne determinants of pigmentation and thiamine prototrophy in *Erwinia herbicola*. *J Bacteriol* **151**, 1627–1629.
- Goodfellow, M., Austin, B. & Dickinson, C. H. (1976). Numerical taxonomy of some yellow-pigmented bacteria isolated from plants. *J Gen Microbiol* **97**, 219–233.
- Hundle, B. S., O'Brien, D. A., Alberti, M., Beyer, P. & Hearts, J. E. (1992). Functional expression of zeaxanthin glucosyltransferase from *Erwinia herbicola* and a proposed uridine diphosphate binding site. *Proc Natl Acad Sci USA* **89**, 9321–9325.
- Hundle, B. S., O'Brien, D. A., Beyer, P., Kleinig, H. & Hearst, J. E. (1993). *In vitro* expression and activity of lycopene cyclase and β -carotene hydroxylase from *Erwinia herbicola*. *FEBS Lett* **315**, 329–334.
- Kado, C. I. & Liu, S. T. (1981). Rapid procedure for detection and isolation of large and small plasmids. *J Bacteriol* **145**, 1365–1373.
- Krishnan, B. R., Blakesley, R. W. & Berg, D. E. (1991). Linear amplification DNA sequencing directly from single phage plaques and bacterial colonies. *Nucleic Acids Res* **19**, 1153.
- Lee, L.-Y. & Liu, S.-T. (1991). Characterization of the yellow-pigment genes of *Erwinia herbicola*. *Mol Microbiol* **5**, 217–224.
- Liu, S.-T., Li, S.-N., Wang, D.-C., Chang, S.-F., Chiang, S.-C., Ho, W.-C., Chang, Y.-S. & Lai, S.-S. (1991). Rapid detection of hog cholera virus in tissues by the polymerase chain reaction. *J Virol Methods* **35**, 227–236.
- Masui, Y., Coleman, J. & Inouye, M. (1983). Multipurpose expression cloning vehicles in *Escherichia coli*. In *Experimental Manipulation of Gene Expression*, chapter 2, pp. 15–22. London: Academic Press.
- Math, S. K., Hearst, J. E. & Poulter, C. D. (1992). The *crtE* gene in *Erwinia herbicola* encodes geranylgeranyl diphosphate synthase. *Proc Natl Acad Sci USA* **89**, 6761–6764.
- Miller, J. H. (1972). *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Misawa, N., Nakagawa, M., Kobayashi, K., Yamano, S., Izawa, Y., Nakamura, K. & Harashima, K. (1990). Elucidation of the *Erwinia ureidovorae* carotenoid biosynthesis pathway by functional analysis of gene products expressed in *Escherichia coli*. *J Bacteriol* **172**, 6704–6712.
- Pérombelon, M. C. M. (1992). The genus *Erwinia*. In *The Prokaryotes*, 2nd edn, pp. 2899–2921. Edited by A. Balows, H. G. Trüper, M. Dworkin, W. Harder, K. H. Schleifer. New York: Springer-Verlag.
- Perry, K. L., Simonitch, T. A., Harrison-Lavoie, K. J. & Liu, S.-T. (1986). Cloning and regulation of *Erwinia herbicola* pigment genes. *J Bacteriol* **168**, 607–612.
- Riggle, J. H. & Klos, E. J. (1970). Inhibition of *Erwinia amylovora* by *Erwinia herbicola*. *Phytopathology* **60**, 1310.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Sandmann, G. & Misawa, N. (1992). New functional assignment of the carotenogenic gene *crtB* and *crtE* with constructs of these genes from *Erwinia* species. *FEMS Microbiol Lett* **90**, 253–258.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* **74**, 5463–5467.
- Shnurr, G., Schmidt, A. & Sandmann, G. (1991). Mapping of a carotenogenic gene cluster from *Erwinia herbicola* and functional identification of six genes. *FEMS Microbiol Lett* **78**, 157–162.
- Thiry, G. (1984). Plasmids of the epiphytic bacterium *Erwinia ureidovorae*. *J Gen Microbiol* **130**, 1623–1632.
- Tuveson, S. V., Larson, R. A. & Kagan, J. (1988). Role of cloned carotenoid genes expressed in *Escherichia coli* in protecting against

inactivation by near-UV light and specific phototoxic molecules. *J Bacteriol* **170**, 4675–4680.

Winkelman, G., Lupp, R. & Jung, G. (1980). Herbicolins. New peptide antibiotics from *Erwinia herbicola*. *J Antibiot (Tokyo)* **33**, 353–358.

Yanisch-Perron, G., Vieira, J. & Messing, J. (1985). Improved M13

phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119.

Received 24 June 1993; revised 20 August 1993; accepted 25 August 1993)