Mol Gen Genet (1994) 245:406–416

ORIGINAL PAPER

B. Hundle · M. Alberti · V. Nievelstein · P. Beyer
H. Kleinig · G. A. Armstrong · D. H. Burke
J. E. Hearst

Functional assignment of *Erwinia herbicola* Eho10 carotenoid genes expressed in *Escherichia coli*

Received: 2 March 1994 / Accepted: 6 May 1994

Abstract *Erwinia herbicola* is a nonphotosynthetic bacterium that is yellow pigmented due to the presence of carotenoids. When the *Erwinia* carotenoid biosynthetic genes are expressed in *Escherichia coli*, this bacterium also displays a yellow phenotype. The DNA sequence of the plasmid pPL376, carrying the entire *Erwinia* carotenoid gene cluster, has been found to contain 12 open reading frames (ORFs). Six of the ORFs have been identified as carotenoid biosynthesis genes that code for all the enzymes required for conversion of farnesyl pyrophosphate (FPP) to zeaxanthin diglucoside via geranylgeranyl pyrophosphate, phytoene, lycopene, β-carotene, and zeaxanthin. These enzymatic steps were assigned after disruption of each ORF by a specific mutation and analysis of the accumulated intermediates. Carotenoid intermediates were identified by the absorption spectra of the colored components and by high pressure liquid chromatographic analysis. The six carotenoid genes are arranged in at least two operons. The gene coding for β-carotene hydroxylase is transcribed in the opposite direction from that of the other carotenoid genes and overlaps with the gene for phytoene synthase.

Communicated by A. Kondorosi

B. Hundle[1] · M. Alberti · V. Nievelstein[2] · P. Beyer[2]
H. Kleinig[2] · G.A. Armstrong[3] · D.H. Burke[4] · J.E. Hearst (✉)
Department of Chemistry,
University of California and Division of Structural Biology,
Lawerence Berkeley Laboratory, Berkeley, CA 94720, USA

Present addresses:
[1] Gallo Center, Department of Neurology,
University of California San Francisco, San Francisco,
CA 94110, USA
[2] Institute of Biology II, Albert-Ludwigs-Universität,
Schänzlestr. 1, 79104 Freiburg, Germany
[3] Institute for Plant Sciences, Department of Plant Genetics,
Swiss Federal Institute of Technology (ETH), CH-8092 Zurich,
Switzerland
[4] University of Colorado, Department of Chemistry, Boulder,
ColoradoTo whom reprint requests should be addressed

Key words Carotenoid genes · *Erwinia herbicola*
Fumarate binding domain · Melibiose carrier protein
Cystathione γ-synthase

Introduction

Erwinia herbicola is a nonphotosynthetic, gram-negative bacterium which is yellow in color due to the presence of carotenoids. Carotenoids are methyl-branched C_{40} compounds composed of poly-isoprene units that have been further desaturated to produce a chromophore with conjugated double bonds. These compounds are powerful antioxidants and serve as important protectants against photooxidative damage (Tuveson et al. 1988). Plants and various photosynthetic and nonphotosynthetic bacteria synthesize a wide variety of carotenoids. At least 563 carotenoids had been described by 1987 (Straub 1987) and many more have since been isolated.

The carotenoid biosynthetic pathway, starting with the first isoprenoid compound isopentenyl pyrophosphate (IPP), is common to all plants and photosynthetic bacteria. Isomerization of IPP results in the formation of dimethylallyl pyrophosphate (DMAPP). One molecule of IPP is condensed with one molecule of DMAPP by a prenyl transferase to form geranyl pyrophosphate (GPP, C_{10}). Further addition of IPP molecules leads to the formation of farnesyl pyrophosphate (FPP, C_{15}), and subsequently geranylgeranyl pyrophosphate (GGPP, C_{20}). GGPP is then dimerized to the first C_{40} carotenoid, phytoene. In anoxygenic photosynthetic bacteria, such as *Rhodobacter capsulatus* and *R. sphaeroides*, phytoene is converted to neurosporene via three desaturation steps. These photosynthetic bacteria synthesize only linear carotenoids, primarily spheroidenone and hydroxyspheroidenone. In plants, however, phytoene is converted via four desaturation reactions to lycopene, which upon terminal cyclization and hydroxylation is converted to zeaxanthin via β-carotene. In *E. herbicola* (Hundle et al. 1991) and *E.*

uredovora (Misawa et al. 1990), zeaxanthin is converted to zeaxanthin diglucoside through two on-ring glucosylation reactions (Hundle et al. 1992).

The *R. capsulatus* carotenoid biosynthesis genes were cloned as part of a 46 kb photosynthetic gene cluster (Marrs 1981) and were the first carotenoid genes to be sequenced from any organism (Armstrong et al. 1989). The genes coding for zeaxanthin diglucoside production from farnesyl pyrophosphate are also clustered in *Erwinia* species. The carotenoid gene cluster of *E. herbicola* Eho10 was initially cloned on a 12.7 kb plasmid, pPL376 (Perry et al. 1986). The carotenoid biosynthesis genes from *E. uredovora* have been cloned and sequenced more recently (Misawa et al. 1990). The carotenoid genes of *E. herbicola* Eho13 have also been cloned as a 6.7 kb fragment (Lee and Liu 1991).

Nucleotide sequence analysis of pPL376 initially revealed the presence of three ORFs encoding gene products homologous to the enzymes for the early portion of the carotenoid biosynthetic pathway in *R. capsulatus* (Armstrong et al. 1990). In this paper, we present the remaining DNA sequence and the functional assignments for all the carotenoid biosynthesis genes involved in conversion of FPP to zeaxanthin diglucoside in *E. herbicola* Eho10, as determined by site-specific mutagenesis and deletion mutations. Tentative assignments of other ORFs present in pPL376 are also made based on sequence comparisons with protein sequence data bases. The gene organization observed in *E. herbicola* Eho10 differs from that in *E. uredovora* and *E. herbicola* Eho13.

Materials and methods

Organisms and growth conditions

The *Escherichia coli* strains containing the *E. herbicola* Eho10 carotenoid gene cluster on plasmid pPL376 (Perry et al. 1986) and mutant derivatives were grown at 37°C in LC media, using selection for resistance to 100 μg/ml ampicillin (Sigma Chemical Co., St. Louis, Mo., USA). *E. coli* DH5αcompetent cells were obtained from BRL (Gaithersburg, Md., USA). Restriction enzymes were purchased from BRL and NEB (Beverly, Mass., USA).

Plasmids, cloning techniques and nucleotide sequence determination

Plasmid pPL376 was subcloned and sequenced as described previously (Armstrong et al. 1990). DNA sequencing was performed by the dideoxy chain termination method (Sanger et al. 1977). All nucleic acid and enzymatic manipulations were performed according to standard published procedures (Maniatis et al. 1982) or manufacturers' protocols. In order to interrupt various ORFs three types of mutations were created.

Deletion mutations

Specific segments of the plasmid pPL376 were deleted using various restriction enzymes in order to eliminate ORFs not required for yellow pigment production (Fig. 3A).

Interposon mutations

ORFs 3 and 5 were interrupted by the insertion of a kanamycin resistance cartridge (Barany 1985) at unique *XhoI* (1.520) and *SstI* (3.989) sites respectively (Fig. 3B).

Frame shift mutations

These mutations were generated by restricting individual genes with specific restriction enzymes that leave a 5′ overhang. Cleaved overhanging ends were filled in using the Klenow fragment of DNA polymerase and blunt ends religated using DNA ligase. This procedure resulted in a frame shift in the amino acid translation and thus altered the downstream portion of the protein (Fig. 3C).

Extraction and separation of carotenoids

In order to analyze accumulated carotenoid precursors, *E. coli* DH5αcells (10 mg) containing various plasmid constructs were extracted (2 × 1 ml) with chloroform/methanol (2:1, v/v). For separation of individual components the extracts were dried under a stream of nitrogen. Carotenes were isolated from silica gel plates, which were developed with the solvent system petroleum benzene/diethyl ether/acetone (40:10:5, v/v/v) and applied to an ET 300/8/4 Nucleosil 5 μm C_{18} reversed phase high pressure liquid chromatographic (HPLC) column (Macherey-Nagel). The column was developed isocratically with acetonitrile at a flow rate of 1.4 ml/min. Xanthophylls were analyzed on silica gel plates using the same solvent system as above. Glucosylated xanthophylls were peracetylated for 30 min at room temperature using a 2:1 (v/v) mixture of benzene and acetic anhydride in the presence of catalytic amounts of 4-(dimethylamino)-pyridine. After a short treatment with methanol and partition against a mixture of petroleum benzene/diethyl ether 2:1 (v/v), the derivatized samples were analyzed on silica gel plates under the same conditions.

Protein sequence comparisons

Database searches were performed at the National Center for Biotechnology Information (NCBI) using the BLAST network service. Databanks searched were Brookhaven Protein Data Bank (June 1993); SWISS-PROT 27.0 (Nov. 93); PIR 38.0 (Sept. 93); genpept, CDS translations from GenBank (R) 79.0 (Oct. 93). Sequences were aligned using the BLASTP gene alignment program (Altschul et al. 1990). Final sequence alignments and the insertion of gaps were performed manually.

Nucleotide sequence accession number

The Genbank accession number of the nucleotide sequence from pPL376 presented in this paper is M87280.

Results and discussion

DNA sequence of plasmid pPL376

The DNA sequence of plasmid pPL376 is presented in Fig. 2. Sequence analysis resulted in identification of 12 open reading frames. ORF2, ORF3, ORF11 (*crtz*) and ORF12 are transcribed in the opposite direction (Fig. 1), in contrast to the other 8 open reading frames. All of the essential genes required for carotenoid production from FPP to zeaxanthin diglucoside (see below) are present in

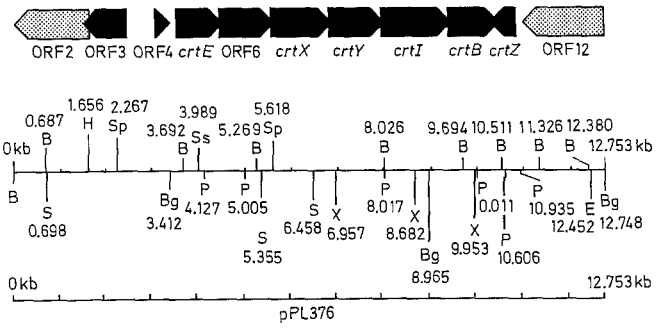


Fig. 1 Linear map of pPL376 showing orientation of open reading frames (ORFs) and various genes. DNA sequence analysis revealed twelve open reading frames. ORF1 (not shown) was found to be the vector β-lactamase gene and ORF12 had an incomplete start. Restriction enzyme site abbreviations are as follows: B, *Bam*HI; S, *Sal*I; Sp, *Sph*I; H, *Hind*III; P, *Pst*I; Ss, *Sst*I; X, *Xmn*I; Bg, *Bgl*I; and E, *Eco*RI

at least two separate operons, as ORF11 (*crtz*; Fig. 1) is transcribed in the opposite direction from the other essential genes. Analysis of the DNA sequence (Fig. 2) seems to suggest the presence of *E. coli*-like σ70 promoters (McClure 1985) preceding ORF3, ORF7, and ORF11 (data not shown). ORFs 5,9, and 10 had previously been predicted to be homologous to the *R. capsulatus crt*E, *crt*I and *crt*B genes, respectively (Armstrong et al. 1990).

Functional assignments to various ORFs

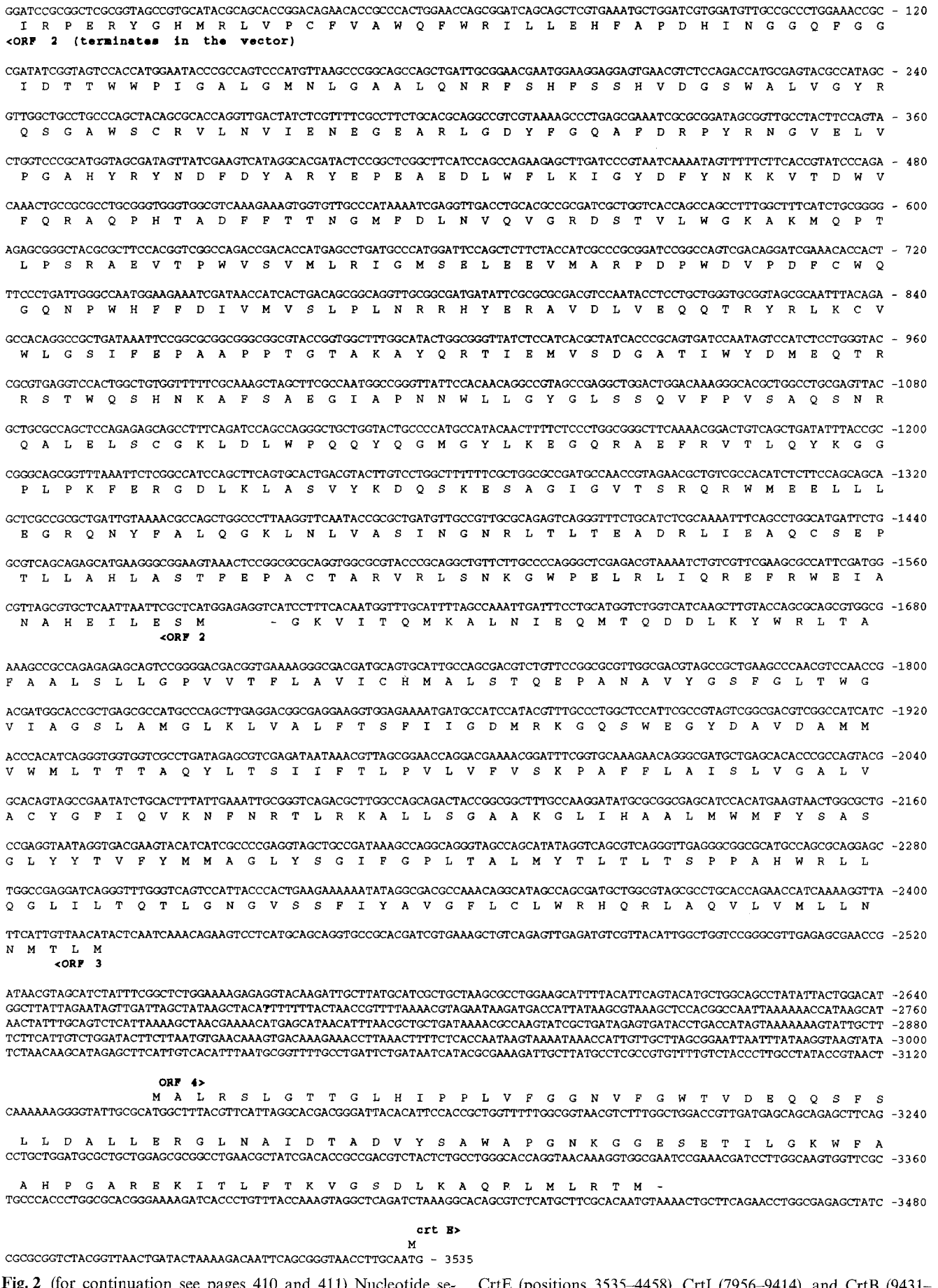
ORF1 (not shown) is the β-lactamase gene, comprising part of the vector pHC79 (Perry et al. 1980) sequence. In order to determine which of the other ORFs are involved in carotenoid production, mutants (Fig. 3) were generated by one of several mutational techniques and analyzed for the accumulation of carotenoid products, as shown in Table 1. The functions of the respective genes as determined on plasmid pPl376 are summarized in Table 2. Deletion of ORF2 (pAPU311H) or disruption of ORF3 with an interposon (pGXX; Fig. 3), or deletion of both ORF2 and ORF3 (pAPU211) caused no change in accumulated carotenoid products. This indicates that ORF2 and ORF3 are not required for carotenoid biosynthesis, or possibly that *E. coli* homologues may replace their functions. The deduced amino acid sequence of ORF2 contains two domains which are

Table 1 The various strains constructed and their accumulated products (*ORF* open reading frame)

Strain	Color	ORFs and genes disrupted	Phytoene	Lycopene	β-Carotene	Hydroxy-carotenoids	Carotenoid glucosides
pPL376	Yellow	None	Trace	–	Trace	Trace	95±4
pAPU311H	Yellow ORF2,3	Trace	–	Trace	Trace	90±2.5	
pAPU201	White	ORF2,3,4,6, *crtE, X*	–	–	–	–	–
pAPU211	Yellow	ORF2,3Trace	–	Trace	Trace	95±4	
pAPU211R	Yellow	ORF2,3 *crtX*	Trace	–	Trace	94±2	–
pGXX	Light yellow	ORF3	Trace	–	Trace	–	90±3
pGSS	Light yellow	*crtE*	Trace	–	Trace	–	–
pAPU2114	Yellow	ORF2,3,4	Trace	–	Trace	–	95±4
pAPU2116	Yellow	ORF2,3,6	Trace	–	7±15	Trace	89±2.8
pAPU2116X	Yellow	ORF2,3,6, *crtX*	4±0.8	–	10±1.1	75±2	–
pAPU211Y	Pink	ORF2,3, *crtY*	53±3	32±4	–	–	–
pAPU211I	White	ORF2,3, *crtI*	100	–	–	–	–
pAPU211B	White	ORF2,3, *crtB*	–	–	–	–	–
pAPU211Z	Orange	ORF2,3, *crtZ*	24±0.5	Trace	71±1	–	–
pAPU211BZ	White	ORF2,3, *crtB, Z*	–	–	–	–	–

Table 2 Function of genes on pPL376

Gene	Function or homologies of gene product	Nucleotide numbering on pPL376	Number of amino acids	Molecular weight (kDa)
ORF2	Prosucrase/isomaltase?	1588–vector	529	61
ORF3	Melibiose carrier Protein?	2414–1596	272	30
ORF4	–	3140–3451	103	11
crtE	Geranylgeranyl synthase	3535–4458	307	33
ORF6	–	4521–5564	347	37
crtX	Zeaxanthin-glucosyltransferase	5561–6802	413	45
crtY	Lycopene cyclase	6799–7959	386	43
crtI	Phytoene desaturase	7956–9414	492	54.5
crtB	Phytoene synthase	9431–10360	309	34
crtZ	β-Carotene hydroxylase	10826–10296	176	20
ORF12	Cystathione γ-synthase	vector–10916	611	66



ATAACGTAGCATCTATTTCGGCTCTGGAAAAGAGAGGTACAAGATTGCTTATGCATCGCTGCTAAGCGCCTGGAAGCATTTTACATTCAGTACATGCTGGCAGCCTATATTACTGGACAT -2640
GGCTTATTAGAATAGTTGATTAGCTATAAGCTACATTTTTTTACTAACCGTTTTAAAACGTAGAATAAGATGACCATTATAAGCGTAAAGCTCCACGGCCAATTAAAAAACCATAAGCAT -2760
AACTATTTGCAGTCTCATTAAAAGCTAACGAAAACATGAGCATAACATTTAACGCTGCTGATAAAACGCCAAGTATCGCTGATAGAGTGATACCTGACCATAGTAAAAAAAGTATTGCTT -2880
TCTTCATTGTCTGGATACTTCTTAATGTGAACAAAGTGACAAAGAAACCTTAAACTTTTCTCACCAATAAGTAAAATAAACCATTGTTGCTTAGCGGAATTAATTTATAAGGTAAGTATA -3000
TCTAACAAGCATAGAGCTTCATTGTCACATTTAATGCGGTTTTGCCTGATTCTGATAATCATACGCGAAAGATTGCTTATGCCTCGCCGTGTTTTGTCTACCCTTGCCTATACCGTAACT -3120

CAAAAAAGGGGTATTGCGCATGGCTTTACGTTCATTAGGCACGACGGGATTACACATTCCACCGCTGGTTTTTGGCGGTAACGTCTTTGGCTGGACCGTTGATGAGCAGCAGAGCTTCAG -3240

CCTGCTGGATGCGCTGCTGGAGCGCGGCCTGAACGCTATCGACACCGCCGACGTCTACTCTGCCTGGGCACCAGGTAACAAAGGTGGCGAATCCGAAACGATCCTTGGCAAGTGGTTCGC -3360

TGCCCACCCTGGCGCACGGGAAAAGATCACCCTGTTTACCAAAGTAGGCTCAGATCTAAAGGCACAGCGTCTCATGCTTCGCACAATGTAAAACTGCTTCAGAACCTGGCGAGAGCTATC -3480

CGCGCGGTCTACGGTTAACTGATACTAAAAGACAATTCAGCGGGTAACCTTGCAATG - 3535

Fig. 2 (for continuation see pages 410 and 411) Nucleotide sequences of pPL376 showing open reading frames and genes together with their encoded amino acid sequences. Sequences for CrtE (positions 3535–4458), CrtI (7956–9414), and CrtB (9431–10360) have been previously published, from our laboratory (Armstrong et al. 1990) and are not shown

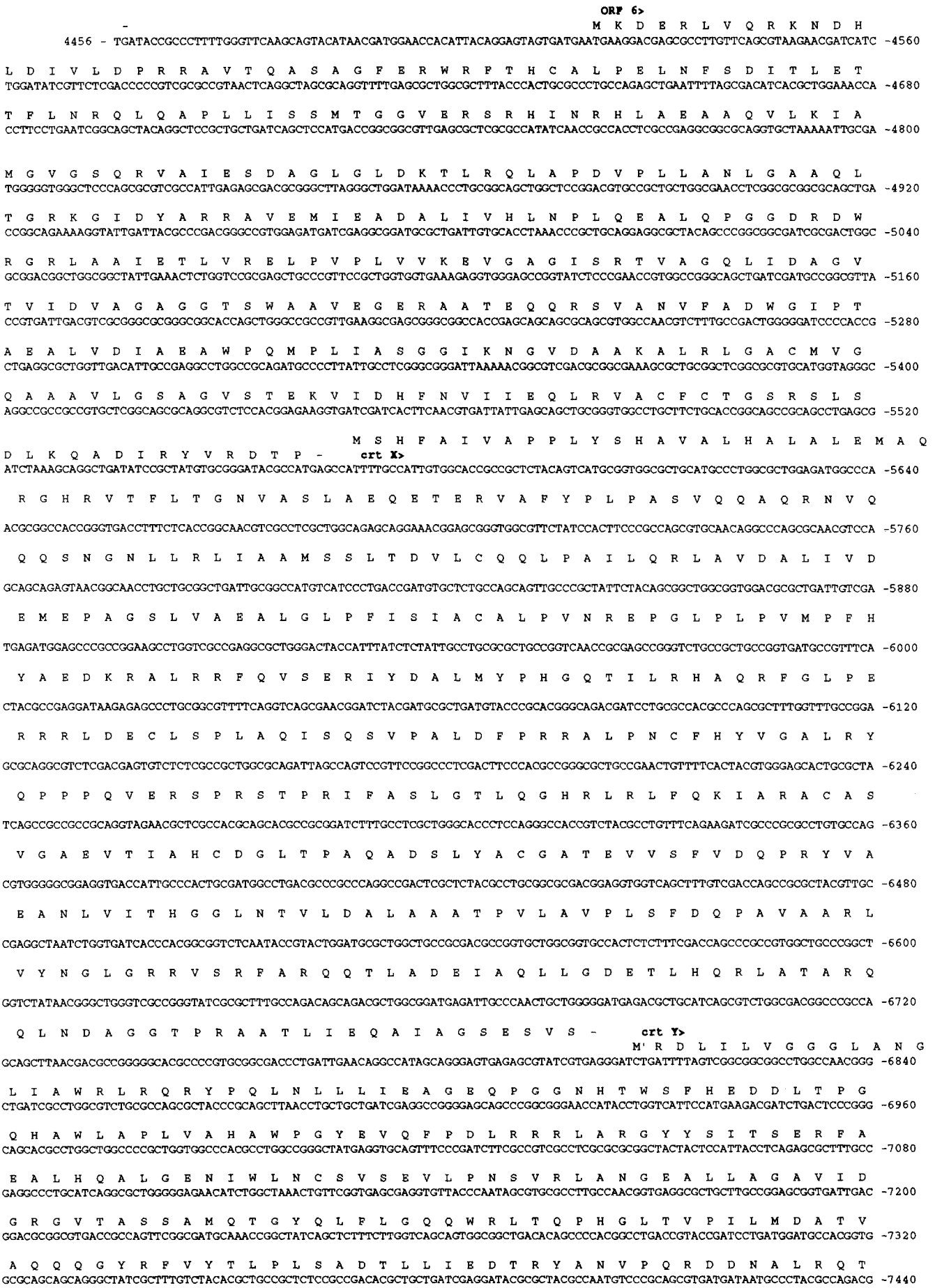


Fig. 2 (continued)

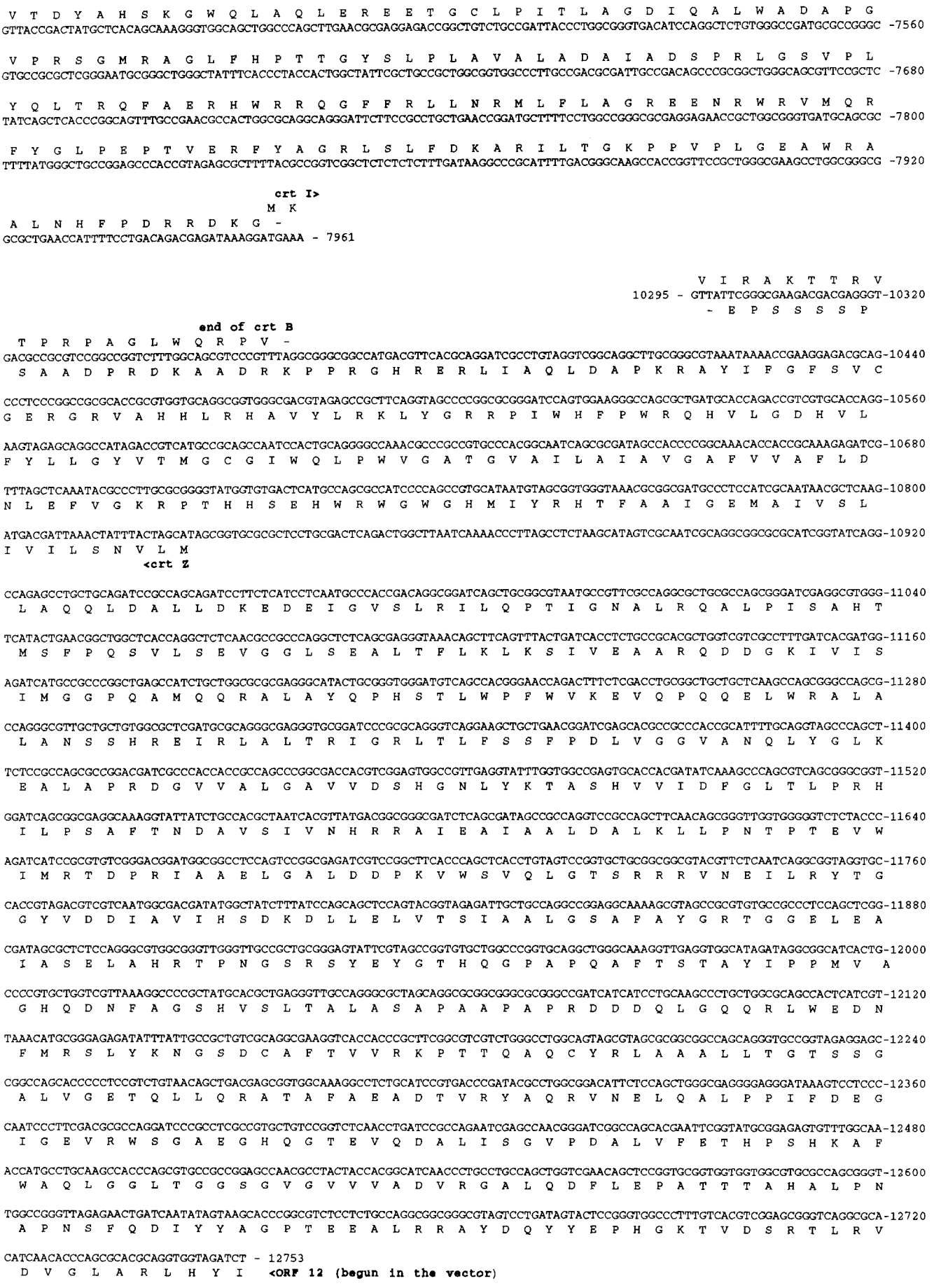


Fig. 2 (continued)

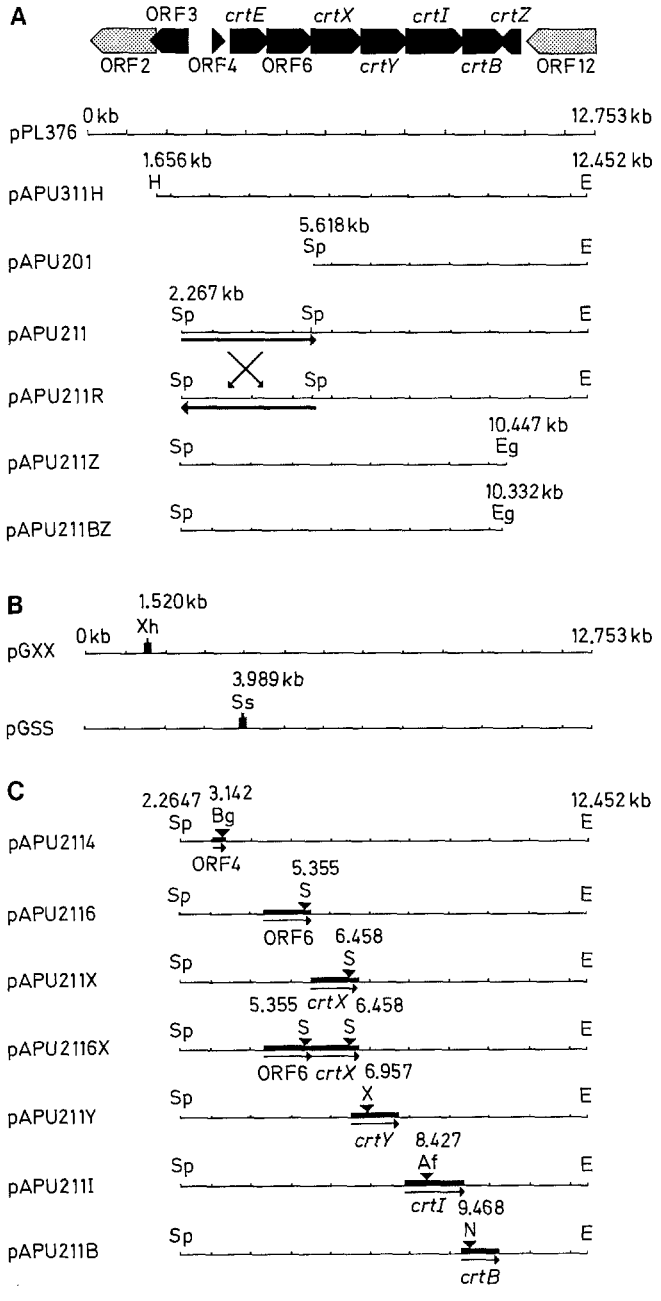


Fig. 3A–C Linear plasmid maps showing various mutation sites generated to interrupt the ORFs. **A** Deletion mutations, **B** interposon mutations, and **C** frame shift mutations

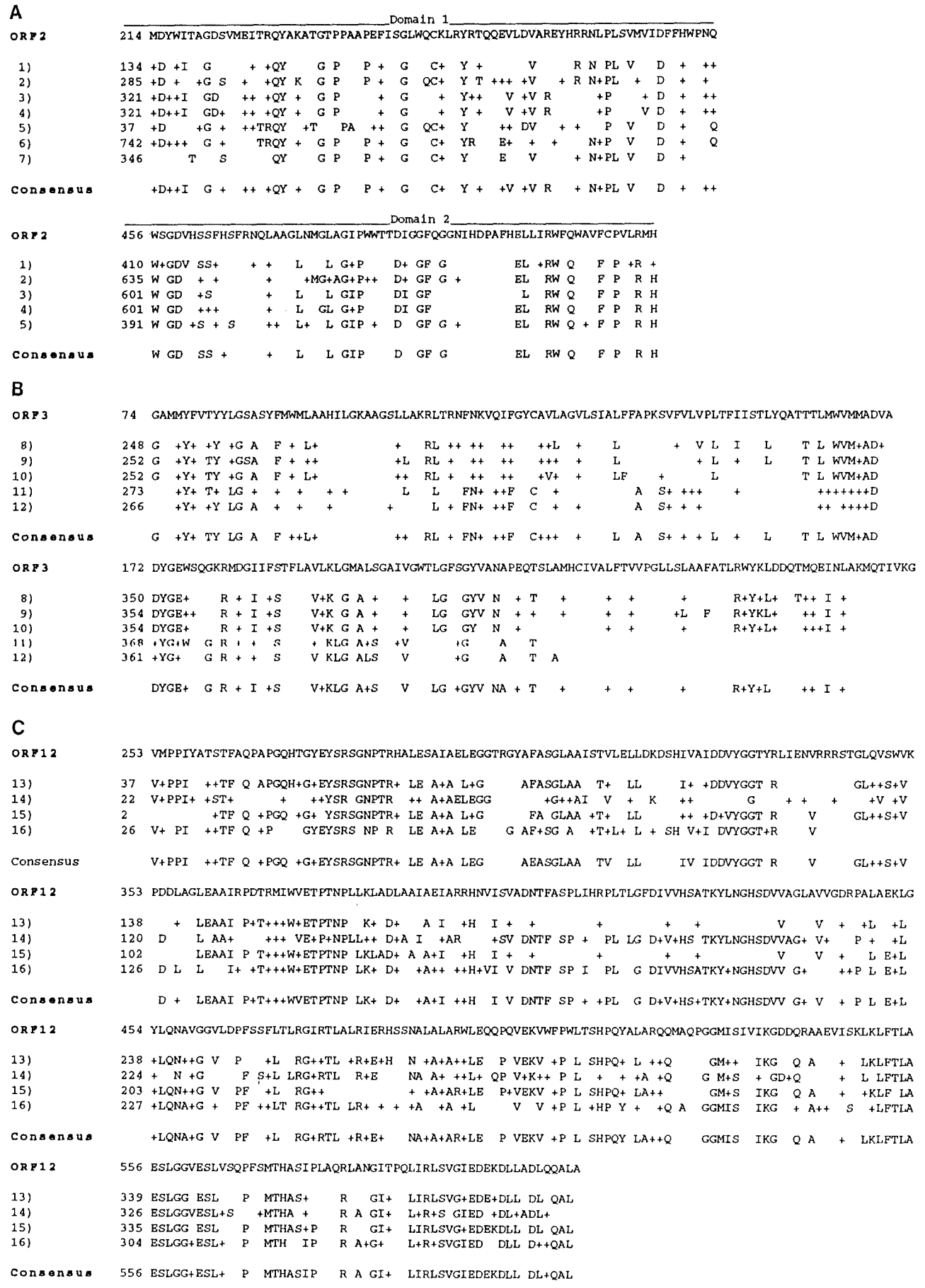
shared by various enzymes involved in carbohydrate metabolism (Fig. 4). ORF2 is truncated at the 3′ end as a result of the pPL376 cloning procedure. ORF3 has two conserved domains homologous with the lactose permease and melibiose carrier proteins from several bacterial sources.

The pAPU201 clone, in which ORFs 2,3,4,5, and 6 were deleted, gave a white phenotype and accumulated no carotenoids. Addition of the 3.351 kb, *Sph*I (2.267)-*Sph*I (5.618) fragment in pAPU211 to give pAPU201 restored the wild-type yellow phenotype, primarily due to the production of zeaxanthin and zeaxanthin glucosides (Table 1). This indicates that ORF4, ORF5, and/or ORF6 code for genes required for carotenoid biosynthesis. Interestingly, the pAPU211R clone, in which this *Sph*I-*Sph*I DNA fragment was inserted in the opposite orientation also had the same yellow phenotype and carotenoid composition as the wild type. Frame shift mutations in ORF4 (pAPU2114) and ORF6 (pAPU2116) gave constructs that produced wild-type pigments. The sequence comparison of ORF6 with the data bases did not yield any significant matches for most of the sequence, except in one small region (Fig. 5). The amino acid sequence of ORF6 contains a small segment of 26 residues which is highly conserved between bacterial and eukaryotic iosine-5′-monophosphate (IMP) dehydrogenases, spinach glycolate oxidase, and yeast dihydroorotate dehydrogenases (Fig 5). All of these enzymes can use NAD^+ as a terminal electron acceptor with the exception of the dihydroorotate dehydrogenases, which only use fumarate (Nagy et al. 1992). The conserved domain of 26 amino acids is not identical to an NAD^+ binding domain (Armstrong et al. 1993), but might be a putative fumarate binding domain. It is not known whether IMP dehydrogenase and spinach glycolate oxidase can use fumarate as an electron acceptor.

CrtE (ORF5)

An interposon insertion in pPL376 at the *Sst*I (3.989) site in ORF5 gave a white phenotype (pGSS). After 3–4 days storage at 4°C, these clones developed a very faint yellow color indicating that the mutant is slightly leaky, or that basal levels of GGPP present in *E. coli* cells can also be utilized for carotenoid biosynthesis. The protein encoded by ORF5 contains 307 amino acids and has a predicted molecular weight of 33 kDa. The protein has a 55% amino acid identity to GGPP synthase of *E. uredovora* and a 30% identity to the homology in *R. capsulatus* (Armstrong et al. 1990). The product of this gene from pAPU211 was expressed in *E. coli* and its in vitro activity was identified as GGPP synthase, using [^{14}C]IPP and [^{3}H]FPP as substrates (Math et al. 1992). Kinetic studies on *E. uredovora* CrtE indicate that while

Fig. 4A–C Sequence comparison of ORFs. **A** ORF2:1, gpY00839, human, acid maltase; 2, PIRJN0102, *Schwanniomyces occidentalis*, 1-glucoamylase; 3, spP14410, human, sucrase-isomaltase; 4, spP07768, rabbit, sucrase-isomaltase; 5, pirJC1200, Aspergillus occidentallis, α-glucosidase; 6, spP23739, rat, intestinal sucrase-isomaltase; 7, SPP29064, *Candida tsukubaensis*, α-glucosidase. **B** ORF3:8, spP02921, *Escherichia coli*, melibiose carrier protein; 9, pirB44166, *Klebsiella pneumoniae*, melibiose carrier protein; 10, pirS23576, *Salmonela typhimurium*, melibiose carrier protein; 11, spP23936, *Streptococcus thermophilus*, lactose permease; 12, spP222733, *Lactobacillus delbrueckii*, lactose permease. **C** ORF12:13, gpS52784, human, cystathionine γ-synthase; 14, spP00935, *E. coli*, cystathionine γ-lyase; 15, spP18757, *Rattus norvegicus*, cystathionine γ-lyase; 16, spP31373, *Saccharomyces cerevisiae*, cystathionine γ-lyase. Only conserved amino acids are explicitly specified, gaps are indicated as such


A
Domain 1
ORF2 214 MDYWITAGDSVMEITRQYAKATGTPPAAPEFISGLWQCKLRYRTQQEVLDVAREYHRRNLPLSVMVIDFFHWPNQ
1) 134 +D +I G + +QY G P P + G C+ Y + V R N PL V D + ++
2) 285 +D + +G S + +QY K G P P + G QC+ Y T +++ +V + R N+PL + D + +
3) 321 +D++I GD ++ +QY + G P + G + Y++ V +V R +P + D + ++
4) 321 +D++I GD+ ++ +QY + G P + G + Y + V +V R +P V D + ++
5) 37 +D +G + ++TRQY +T PA ++ G QC+ Y ++ DV + + P V D + Q
6) 742 +D+++ G + TRQY + G P P + G C+ YR E+ + + N+P V D + Q
7) 346 T S QY G P P + G C+ Y E V + N+PL V D +
Consensus +D++I G + ++ +QY + G P P + G C+ Y + +V +V R + N+PL V D + ++
Domain 2
ORF2 456 WSGDVHSSFHSFRNQLAAGLNMGLAGIPWWTTDIGGFQGGNIHDPAFHELLIRWFQWAVFCPVLRMH
1) 410 W+GDV SS+ + + L L G+P D+ GF G EL +RW Q F P +R +
2) 635 W GD + + + +MG+AG+P++ D+ GF G + EL RW Q F P R H
3) 601 W GD +S + L L GIP DI GF L RW Q F P R H
4) 601 W GD +++ + L GL G+P DI GF EL RW Q F P R H
5) 391 W GD +S + S ++ L+ L GIP + D GF G + EL RW Q + F P R H
Consensus W GD SS + + L L GIP D GF G EL RW Q F P R H
B
ORF3 74 GAMMYFVTYYLGSASYFMWMLAAHILGKAAGSLLAKRLTRNFNKVQIFGYCAVLAGVLSIALFFAPKSVFVLVPLTFIISTLYQATTTLMWVMMADVA
8) 248 G +Y+ +Y +G A F + L+ + RL ++ ++ ++ ++L + L + V L I L T L WVM+AD+
9) 252 G +Y+ TY +GSA F + ++ +L RL + ++ ++ +++ + L + L L T L WVM+AD
10) 252 G +Y+ TY +G A F + L+ ++ RL + ++ ++ +V+ + LF + L T L WVM+AD
11) 273 +Y+ T+ LG + + + + + L L FN+ ++F C + A S+ +++ + +++++++D
12) 266 +Y+ +Y LG A + + + + L + FN+ ++F C + + A S+ + + ++ ++++D
Consensus G +Y+ TY LG A F ++L+ ++ RL + FN+ ++F C+++ + L A S+ + + L + L T L WVM+AD
ORF3 172 DYGEWSQGKRMDGIIFSTFLAVLKLGMALSGAIVGWTLGFSGYVANAPEQTSLAMHCIVALFTVVPGLLSLAAFATLRWYKLDDQTMQEINLAKMQTIVKG
8) 350 DYGE+ R + I +S V+K G A + + LG GYV N + T + + + + R+Y+L+ T++ I +
9) 354 DYGE++ R + I +S V+K G A + + LG GYV N + + + + + + +L F R+YKL+ ++ I +
10) 354 DYGE+ R + I +S V+K G A + + LG GY N + + + + R+Y+L+ +++I +
11) 368 +YG+W G R + + S + KLG A+S +V +G A T
12) 361 +YG+ G R + + S V KLG ALS V +G A T A
Consensus DYGE+ G R + I +S V+KLG A+S V LG +GYV NA + T + + + + R+Y+L ++ I +
C
ORF12 253 VMPPIYATSTFAQPAPGQHTGYEYSRSGNPTRHALESAIAELEGGTRGYAFASGLAAISTVLELLDKDSHIVAIDDVYGGTYRLIENVRRRSTGLQVSWVK
13) 37 V+PPI ++TF Q APGQH+G+EYSRSGNPTR+ LE A+A L+G AFASGLAA T+ LL I+ +DDVYGGT R GL++S+V
14) 22 V+PPI+ +ST+ + ++YSR GNPTR ++ A+AELEGG +G++AI V + K ++ G + + + +V +V
15) 2 +TF Q +PGQ +G+ YSRSGNPTR+ LE A+A L+G FA GLAA +T+ LL ++ +D+VYGGT R V GL++S+V
16) 26 V+ PI ++TF Q +P GYEYSRS NP R LE A+A LE G AF+SG A +T+L+ L + SH V+I DVYGGT+R V
Consensus V+PPI ++TF Q +PGQ +G+EYSRSGNPTR+ LE A+A LEG AEASGLAA TV LL IV IDDVYGGT R V GL++S+V
ORF12 353 PDDLAGLEAAIRPDTRMIWVETPTNPLLKLADLAAIAEIARRHNVISVADNTFASPLIHRPLTLGFDIVVHSATKYLNGHSDVVAGLAVVGDRPALAEKLG
13) 138 + LEAAI P+T+++W+ETPTNP K+ D+ A I +H I + + + + V V + +L +L
14) 120 D L AA+ +++ VE+P+NPLL++ D+A I +AR +SV DNTF SP + PL LG D+V+HS TKYLNGHSDVVAG+ V+ P + +L
15) 102 LEAAI P T+++W+ETPTNP LKLAD+ A A+I +H I + + + V V + L E+L
16) 126 D L L I+ +T+++W+ETPTNP LK+ D+ +A++ ++H+VI V DNTF SP I PL G DIVVHSATKY+NGHSDVV G+ ++P L E+L
Consensus D + LEAAI P+T+++WVETPTNP LK+ D+ +A+I ++H I V DNTF SP + +PL G D+V+HS+TKY+NGHSDVV G+ V + P L E+L
ORF12 454 YLQNAVGGVLDPFSSFLTLRGIRTLALRIERHSSNALALARWLEQQPQVEKVWFPWLTSHPQYALARQQMAQPGGMISIVIKGDDQRAAEVISKLKLFTLA
13) 238 +LQN++G V P +L RG++TL +R+E+H N +A+A++LE P VEKV +P L SHPQ+ L ++Q GM++ IKG Q A + LKLFTLA
14) 224 + N +G F S+L LRG+RTL R+E NA A+ ++L+ QP V+K++ P L + + +A +Q G M+S + GD+Q + L LFTLA
15) 203 +LQN++G V PF +L RG++ + +A+AR+LE P+VEKV +P L SHPQ+ LA++ GM+S IKG Q A + +KLF LA
16) 227 +LQNA+G + PF ++LT RG++TL LR+ + + +A +A +L V V +P L +HP Y + +Q A GGMIS IKG + A++ S +LFTLA
Consensus +LQNA+G V PF +L RG+RTL +R+E+ NA+A+AR+LE P VEKV +P L SHPQY LA++Q GGMIS IKG Q A + LKLFTLA
ORF12 556 ESLGGVESLVSQPFSMTHASIPLAQRLANGITPQLIRLSVGIEDEKDLLADLQQALA
13) 339 ESLGG ESL P MTHAS+ R GI+ LIRLSVG+EDE+DLL DL QAL
14) 326 ESLGGVESL+S +MTHA + R A GI+ L+R+S GIED +DL+ADL+
15) 335 ESLGG ESL P MTHAS+P R GI+ LIRLSVG+EDEKDLL DL QAL
16) 304 ESLGG+ESL+ P MTH IP R A+G+ L+R+SVGIED DLL D++QAL
Consensus 556 ESLGG+ESL+ P MTHASIP R A GI+ LIRLSVGIEDEKDLL DL+QAL

```
ORF6        267  MPLIASGGIKNGVDAAKALRLGACMV  292

1)          359  +PLIA GGI    D AKA+  GA  +
2)          359  +P+IA GGI+N    AKAL LGA  V
3)          359  +P+IA GGI+N    AKAL LGA  V
4)          279  +P+   GG++ G D  KAL LGA  V
5)          380  + +I +GGIK+G DA + L  GA
6)          269  +P++  GGI +G DA +  R GA MV

consensus   267  +P+IA GGI+NG D AKAL LGA  V  292
```

Fig. 5 Sequence alignment of ORF6:1, spP31002, *Acinoacter calcoacticus*, inosine-5′-monophosphate (IMP) dehydrogenase; 2, spP12268, human, IMP dehydrogenase; 3, spP12269, Chinese hamster, IMP dehydrogenase, 4, pDB/IGOX, spinach, glycolate oxidase; 5, spP28272, *S. cerevisiae*, dihydroorotate dehydrogenase; 6, spP32747, *Schizosaccharomyces pombe*, dihydroorotate dehydrogenase

GPP and FPP are the genuine substrates for producing GGPP, DMAPP is not a good substrate (Weidemann et al. 1993). *E. herbicola* CrtE shares two highly conserved domains with several FPP synthases, hexaprenyl pyrophosphate synthase, and GGPP synthases (Math et al. 1992). Domain I consists of the sequence ^{82}(V/I)EX_6LX_2DD$X_{2\text{-}4}$DX_4RRG^{106} and domain II contains ^{222}GX_2FQX_2DDX_2D$X_{5\text{-}8}$GK^{240} (Armstrong et al. 1993).

CrtX (ORF7)

A frame shift mutation in the *Sal*I site (6.458) in ORF7 gave a yellow strain, pAPU211X. Pigment analysis of pAPU211X showed accumulation of zeaxanthin. This gene product is thus identified as zeaxanthin glucosyltransferase, which converts zeaxanthin to zeaxanthin mono- and diglucoside (Hundle et al. 1991). The predicted molecular weight of this enzyme is 45 kDa. The in vitro enzyme activity of the *E. herbicola* gene product, which requires activated glucose (UDP-glucose) as substrate was demonstrated recently by our laboratory (Hundle et al. 1992). *E. herbicola crtX* starts at nucleotide 5561 and the deduced amino acid sequence shows 45% similarity to *E.* urodevora CrtX. In the pAPU211R clone, the 5′-terminus of *crtX* is truncated from nucleotides 5561 to 5681. This suggests that CrtX can function without the first 18 amino acid residues. UDP-requiring glucosyltransferases, including CrtX, share a conserved region, the UDP binding domain QX_{13}TX_2GX_7LX_4PX_4PX3DQX_4A (Hundle et al. 1992).

CrtY (ORF8)

This gene was identified by a frame shift mutation in the *Xmn*I (6.957) site in ORF8. The resulting strain, pAPU211Y, has a pink phenotype and accumulates lycopene, indicating that this enzyme is required for the conversion of lycopene to β-carotene. The predicted molecular weight of *E. herbicola* lycopene cyclase is 43 kDa and it has 58% sequence similarity to the *E. uredovora* lycopene cyclase. The identity of this enzyme has been further established in our laboratory through observation of its activity in *E. coli* and in vitro conversion of lycopene to β-carotene (Hundle et al. 1993).

CrtI (ORF9)

A frame shift interruption of this ORF at the *Afl*1 (8.427) site resulted in a white strain, pAPU211I. This strain accumulated phytoene, and so this ORF is assumed to encode phytoene dehydrogenase. The predicted 54.5 kDa product of this gene has 76% sequence identity to *E. uredovora* CrtI and 42% identity to *R. capsulatus* CrtI protein (Armstrong et al. 1990). The phytoene dehydrogenases from *E. herbicola* and *E. uredovora* catalyze the conversion of phytoene to lycopene (four desaturation steps), while *R. capsulatus* CrtI only converts phytoene to neurosporene (three desaturation steps). Nevertheless, *R. sphaeroides crtI* mutants are complemented *in trans* by the *E. herbicola crtI* gene (Hunter et al. 1994). This indicates that the two enzymes function in fundamentally similar ways in a carotenoid biosynthesis cascade despite the difference in the dehydrogenase sequence. The *Erwinia* carotenoid desaturases, *R. capsulatus* desaturases (CrtI and CrtD) and *Neurospora crassa al*-1 gene product, a CrtI homolog, all share a common 30-amino acid N-terminal βαβfold responsible for FAD/NAD(P) binding: (V/I)(V/I)G(A/G)GXGG(L/I)X_2(L/A)X_3GX_2(V/T)X(V/L)X(E/D)X_5GG (Wierenga et al. 1986). The *E. uredovora* CrtI homologue has been expressed in *E. coli* and renatured from inclusion bodies. The reconstituted enzyme demonstrated a specific cofactor requirement for flavin adenine dinucleotide (FAD) while nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP) were found to be inhibitory (Fraser et al. 1992).

CrtB (ORF10)

The strain pAPU211B was generated by a frame shift mutation at the *Nco*I (9.468) site in ORF10. The resulting clone was white in color and accumulated geranylgeranyl pyrophosphate. Thus, ORF10 is required for conversion of GGPP to phytoene and this gene therefore codes for CrtB, by analogy to *R. capsulatus*. The identity of the *E. uredovora* ORF10 homology was further confirmed by its expression in *E. coli* and in vitro conversion of GGPP to phytoene (Sandmann and Misawa 1991). The predicted molecular weight of phytoene synthase is 34 kDa. It has a 64% sequence identity to *E. uredovora* CrtB and 31% identity to *R. capsulatus* CrtB (Armstrong et al. 1990). *E. herbicola* shares a conserved prenyl pyrophosphate binding domain with other phytoene synthases and S. cerevisiae squalene synthase: ^{164}(L/M)G(L/I)(A/F)XQX(T/S)NIXRDX_2(E/D)D(A/L)X_2(G/D)RX_2(L/W)PX_7(GA)^{197} (Armstrong et al. 1993; Math et al. 1992). Interestingly, clone pAPU211BZ, in which the *Eag*I (10.332)–*Eco*RI

(12.452) fragment is deleted, gave a white phenotype resembling pAPU211B. This clone expressed the *crtB* gene product truncated by 20 amino acid residues at the C-terminal end, but lacks phytoene synthase activity, indicating the important role of the CrtB C-terminus in enzymatic activity.

CrtZ (ORF11)

This ORF is encoded on the opposite DNA strand from *crtB*. The two overlap by 60 bp at their 3′-termini. The strain pAPU211Z was generated by a deletion in pAPU211 of the *Eag*I (10.447)–*Eco*RI (12.452) fragment. This clone is orange in color and accumulates β-carotene. The reaction catalyzed by this enzyme is thus the conversion of β-carotene to zeaxanthin. The in vitro activity of β-carotene hydroxylase was previously demonstrated by overexpression in *E. coli* in our laboratory (Hundle et al. 1993). This enzyme catalyzes two reaction steps converting β-carotene via a minor product, cryptoxanthin, to zeaxanthin. The protein encoded by this gene has a predicted molecular weight of 22 kDa and has 73% identity to *E. uredovora* β-carotene hydroxylase.

ORF12

This ORF is fused to the pBR vector sequence at its N-terminus and is missing the start codon. Amino acid sequence comparison of the gene product of ORF12 with the data banks showed extensive amino acid similarity to cystathionine γ-synthase. The conserved region of homology is shown (Fig. 4). This region is conserved between *E. coli* cystathionine γ-synthase and human, rat, and S. cerevisiae cystathionine γ-lyase. It is possible that this highly conserved large domain represents part of a larger protein involved in methionine biosynthesis in *Erwinia*.

Although the *E. herbicola* carotenoid biosynthesis pathway appears to be identical to that of *E. uredovora*, there are several differences between the two *crt* gene clusters that suggest considerable evolutionary distances between the two *Erwinia* species. Firstly, at the level of the overall organization of the clusters there is a nonessential gene, ORF6, inserted between *crtE* and the rest of the cluster in *E. herbicola*. A similar ORF is not present in the *E. uredovora* cluster. Secondly, the degree of amino acid identity between homologous pairs of gene products (CrtB = 64%, CrtE = 55%, CrtI = 58%, CrtX = 45%, CrtY = 58%, CrtZ = 73%) suggests that considerable divergence has taken place since their last common ancestor. Comparative biochemical analysis of the two sets of carotenoid biosynthetic enzymes may therefore provide valuable insight into the evolution of carotenoid synthesis.

The *E. herbicola* Eho10 carotenoid biosynthetic gene cluster has been demonstrated to support the biosynthesis of cyclized carotenoids (e.g. β-carotene and zeaxanthin) in both *E. coli* (Hundle et al. 1991) and *R. sphaeroides* (Hunter et al. 1994). By allowing manipulation of carotenoid biosynthesis in photosynthetic bacteria, these genes have tremendous implications for the understanding of the role of carotenoids in assembly of, and energy transfer in, light harvesting complexes. Carotenoid genes from *E. herbicola* Eho13 have been subcloned and used as genetic markers for gene cloning (Liu 1993). Sequence comparisons in various organisms should also assist in the characterization of the physical and catalytic properties of these carotenoid enzymes. Appropriate DNA probes should be useful in locating their homologues from higher organisms.

Acknowledgements Plasmid pPL376 carrying the *E. herbicola* carotenoid biosynthesis genes was kindly provided by the late Robert Tuveson. We thank David Koh for oligonucleotide synthesis and David O'Brien for critical reading of the manuscript. This work was supported by a National Institute of Environment and Health postdoctoral training grant no. 2 T32 ES07075-11 to B.H. and partly by the office of Basic Energy Sciences, Biological Energy Division, Department of Energy under contract no. DE-AC030-76SF00098 and the Deutsche Forschungsgemeinschaft.

References

Altschul SF, Warren G, Miller W, Myers E, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Armstrong GA, Alberti M, Leach F, Hearst JE (1989) Nucleotide sequence, organization, and nature of the protein products of the carotenoid biosynthesis gene cluster of *Rhodobacter capsulatus*. Mol Gen Genet 216:254–268

Armstrong GA, Alberti M, Hearst JE (1990) Conserved enzymes mediate the early reactions of carotenoid biosynthesis in nonphotosynthetic and photosynthetic prokaryotes. Proc Natl Acad Sci USA 87:9975–9979

Armstrong GA, Hundle BS, Hearst JE (1993) Evolutionary conservation and structural similarities of carotenoid biosynthesis gene products from photosynthetic and nonphotosynthetic organisms. Methods Enzymol 214:297–311

Barany F (1985) Single-stranded hexameric linkers: a system for in-phase insertion mutagenisis and protein engineering. Gene 37:111–123

Fraser PD, Misawa N, Linden H, Yamano S, Kobayashi K, Sandmann G (1992) Expression in *Escherichia coli*, purification, and reactivation of the recombinant *Erwinia uredovora* phytoene desaturase. J Biol Chem 267:19891–19895

Hundle BS, Beyer P, Kleinig H, Englert G, Hearst JE (1991) Carotenoids of *Erwinia herbicola* and an *Escherichia coli* HB101 strain carrying the *Erwinia herbicola* carotenoid gene cluster. Photochem Photobiol 54:89–93

Hundle BS, O'Brien DA, Alberti M, Beyer P, Hearst JE (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 BS, O'Brien DA, Beyer P, Kleinig H, Hearst JE (1993) *In vitro* expression and activity of lycopene cyclase and β-carotene hydoxylase from *Erwinia herbicola*. FEBS Lett 315:329–334

Hunter CN, Hundle BS, Hearst JE, Lang HP, Gardiner AT, Takaichi S, Cogdell RJ (1994) Introduction of new carotenoids into the bacterial photosynthetic apparatus by splicing the genes for carotenoid biosynthetic pathways of *Erwinia herbicola* and *Rhodobacter sphaeroides*. J Bacteriol 176:3692–3697

Lee L-Y, Liu S-T (1991) Characterization of the yellow-pigment genes of *Erwinia herbicola*. Mol Microbiol 5:217–224
Liu S-T (1993) Carotenoid-biosynthesis genes as a genetic marker for the purpose of gene cloning. Biochem Biophys Res Commun 195:259–263
Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
Marrs B (1981) Mobilization of the genes for photosynthesis from *Rhodobacter capsulata* by a promiscuous plasmid. J Bacteriol 146:1003–1012
Math SK, Hearst JE, Poulter DC (1992) The *crtE* gene in *Erwinia herbicola* encodes geranylgeranyl diphosphate synthase. Proc Natl Acad Sci USA 89:6761–6764
McClure WR (1985) Mechanism and control of transcription initiation in prokaryotes. Annu Rev Biochem 54:171–204
Misawa N, Nakagawa M, Kobayashi K, Yamano S, Izawa Y, Nakamura K, Harashima K (1990) Elucidation of the *Erwinia uredovora* carotenoid biosynthetic pathway by functional analysis of gene products expressed in *Escherichia coli*. J Bacteriol 172:6704–6712
Nagy M, Lacroute F, Dominique T (1992) Divergent evolution of pyrimidine biosynthesis between anaerobic and aerobic yeasts. Proc Natl Acad Sci USA 89:8966–8970
Ohnuma S, Suzuki M, Nishino T (1994) Archaebacterial ether-linked lipid biosynthetic gene. Expression cloning, sequencing, and characterization of geranylgeranyl-diphosphate synthase. J Biol Chem 269:14792–14797
Perry KC, Simonitch TA, Harrison-Lavoie KJ, Liu ST (1986) Cloning and regulation of *Erwinia herbicola* pigment genes. J Bacteriol 168:607–612
Sandmann G, Misawa N (1991) New functional assignment of the carotenogenic genes *crtB* and *crtE* with constructs of these genes from *Erwinia* species. FEMS Lett 90:253–258
Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467
Straub O (1987) In: Pfander H, Rychener M. Schwabe R (eds) Key to carotenoids. Birkhäuser Verlag; pp 32–90
Tuveson RW, Larson RA, 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
Weidemann M, Misawa N, Sandmann G (1993) Purification and enzymatic characterization of the geranylgeranyl pyrophosphate synthase from *Erwinia uredovora* after expression in *Escherichia coli*. Arch Biochem Biophys 306:152–157
Wierenga RK, Terpstra P, Hol WGJ (1986) Prediction of the occurrence of the ADP-binding βαβ-fold in proteins, using an amino acid sequence fingerprint. J Mol Biol 187:101–107

Note added in proof

ORF-6 shares 29.4% sequence identity to ORF-3 from an archaebacteria, *Sulfolobus acidocaldarius* (Ohnuma S et al. 1994). This unidentified ORF-3 also shares the putative fumarate binding domain shown in Fig. 5.