Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 6761–6764, August 1992
Biochemistry

The *crtE* gene in *Erwinia herbicola* encodes geranylgeranyl diphosphate synthase

(carotenoids/biosynthesis/expression/prenyltransferase)

SHIVANAND K. MATH*, JOHN E. HEARST†, AND C. DALE POULTER*‡

*Department of Chemistry, University of Utah, Salt Lake City, UT 84112; and †Department of Chemistry, University of California and Division of Chemical Dynamics, Lawrence Berkeley Laboratory, Berkeley, CA 94720

Communicated by John R. Roth, April 27, 1992 (received for review February 12, 1992)

ABSTRACT **A cluster of genes essential for the biosynthesis of carotenoids in *Erwinia herbicola* has been isolated and characterized [Armstrong, G. A., Alberti, M. & Hearst, J. E. (1990) *Proc. Natl. Acad. Sci. USA* 87, 9975–9979]. Related gene clusters are found in other carotenoid-producing bacteria. Two of these genes, *crtB* and *crtE*, have been assigned to enzymes responsible for conversion of geranylgeranyl diphosphate (GGPP) to prephytoene diphosphate and prephytoene diphosphate to phytoene, respectively. We isolated *crtE* from the *Er. herbicola* cluster by PCR amplification and cloned the coding region into the *Escherichia coli* expression vector pARC306N. *Es. coli* JM101 was transformed with the expression plasmid, and transformants were assayed for GGPP synthase and phytoene synthase activity. Extracts from JM101/pSM145 accumulated [^{14}C]GGPP when incubated with [^{14}C]isopentenyl diphosphate and farnesyl diphosphate, whereas similar incubations with [^{3}H]GGPP did not yield prephytoene diphosphate or phytoene. Thus, *crtE* encodes GGPP synthase.**

The C_{40} carotenoid carbon skeleton is assembled from two molecules of geranylgeranyl diphosphate (GGPP). This part of the carotenoid pathway proceeds in two steps, as shown in Fig. 1. The first is a c1′-2-3 condensation of two molecules of GGPP to give prephytoene diphosphate, followed by rearrangement of the cyclopropylcarbinyl intermediate to the 1′-1 hydrocarbon phytoene (1). Similar steps are used to synthesize squalene from farnesyl diphosphate (FPP) in the sterol pathway (2). The enzymes responsible for these reactions have been purified from eukaryotic sources. Phytoene synthase from *Capsicum annuum* (3) and squalene synthase from *Saccharomyces cerevisiae* (4) are both bifunctional monomers that catalyze the c1′-2-3 condensation and the 1′-1 rearrangement. Although phytoene synthase is the first pathway-specific enzyme in carotenogenesis, most organisms synthesize the C_{20} substrate for the enzyme in two distinct stages. The first is synthesis of FPP from the fundamental C_5 building blocks isopentenyl diphosphate (IPP) and dimethylallyl diphosphate by FPP synthase (5). FPP lies at a multiple branch point in the isoprene pathway where individual prenyltransferases produce precursors for a variety of metabolites, including sterols (2), ubiquinones (6), dolichols (7), and farnesylated proteins (8). GGPP synthase catalyzes an additional FPP branching reaction with the condensation of FPP and IPP to provide the C_{20} precursor needed for biosynthesis of carotenoids (9), diterpenes (10), and geranylgeranylated proteins (8).

Chromosomal gene clusters capable of directing syntheses of carotenoids when cloned into *Escherichia coli* have recently been isolated and characterized from the carotenogenic bacteria *Erwinia herbicola* (11), *Erwinia uredovora* (12), and *Rhodobacter capsulatus* (13). Two of the genes, *crtB* and *crtE*, were assigned to enzyme activities for the c1′-2-3 condensation and the 1′-1 rearrangement, respectively. This assignment segregates the bifunctional activity of the eukaryotic enzymes into two separate proteins in carotenogenic bacteria and suggests a substantially different structure for 1′-1 condensing enzymes from prokaryotic and eukaryotic sources. The putative amino acid sequences encoded by *crtB* and *crtE* contain regions of substantial similarity with other enzymes in the isoprene pathway. The most notable similarities are among selected segments encoded by *crtE* and a variety of prenyltransferases that catalyze 1′-4 condensations (6) and among the *crtB* gene product, squalene synthase from *S. cerevisiae* (14), and *tom5*, which encodes phytoene synthase in tomatoes (GenBank accession no. X60441). The 1′-4 coupling reaction is the basic chain elongation process in the isoprene pathway and led us to suspect that *crtE* encodes GGPP synthase, an enzyme that directs synthesis of the essential carotenoid precursor from FPP and IPP.

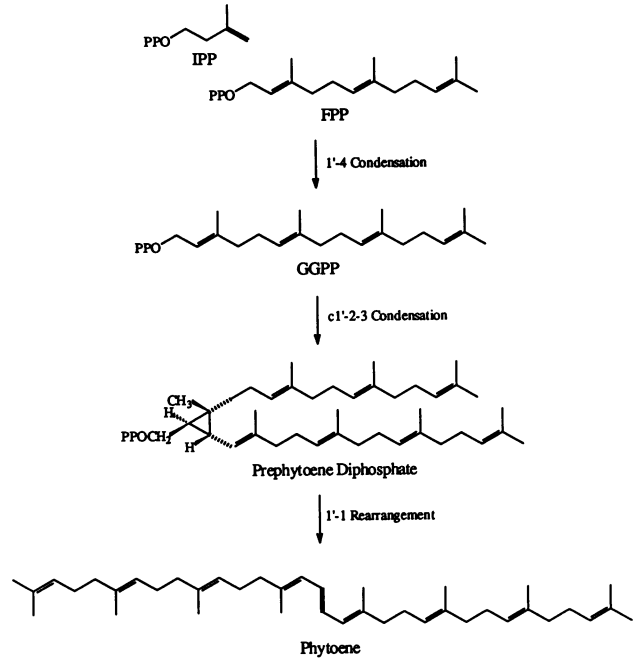

FIG. 1. Early steps in the carotenoid biosynthetic pathway.

MATERIALS AND METHODS

Materials and General Procedures. [^{14}C]IPP (37.6 mCi/mmol; 1 Ci = 37 GBq) was purchased from Amersham.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: FPP, farnesyl diphosphate; GGPP, geranylgeranyl diphosphate; IPP, isopentenyl diphosphate.

‡To whom reprint requests should be addressed.

[^{3}H]GGPP (10 mCi/mmol) was prepared as described (15). Phytoene was obtained from tomatoes. Radioactivity of individual samples was measured in Optifluor scintillation medium (Packard Instrument). HPLC was preformed on a μBondapak C_{18} reversed-phase column. Samples were loaded in methanol and eluted with H_2O/methanol, 15:85 (vol/vol). Radioactivity in the eluant was detected on a Flo-One/Beta Flow Detector (Radiomatic Instruments, Tampa, FL). The *Es. coli* expression vector pARC306N was provided by M. Bittner (Biotechnology Division, Amoco Research, Naperville, IL).

Recombinant DNA Procedures. Plasmid preparations were performed with the kit from Qiagen (Chatsworth, CA). Polymerase chain reactions (PCRs) were carried out with *Taq* polymerase from Stratagene. Restriction endonuclease digestions, DNA polymerase reactions, and ligations were conducted with enzymes from New England Biolabs as described by Sambrook *et al.* (16). DNA fragments were purified on 0.8% agarose gels (SeaKem, FMC), and purified DNA was extracted from the gel matrix by using Elutrap (Schleicher & Schuell). Sequencing was performed by the dideoxynucleotide chain termination procedure of Sanger *et al.* (17) on denatured double-stranded templates using the Sequenase kit (United States Biochemical). *Es. coli* DH5α was used for all plasmid manipulations. *Es. coli* JM101 was used for expression of *crtE*. Competent *Es. coli* cells were prepared, stored, and transformed.

Construction of an Expression Vector for and Expression of *crtE*. The strategy for construction of an expression vector for *crtE* is shown in Fig. 2. The *crtE* gene in pAPU211, which is a pUC19 derivative containing the 5.25- to 15.5-kilobase (kb) region from pPL376, was altered by PCR-mediated site-directed mutagenesis to introduce an *Eco*RI site near the 5′ terminus of the PCR fragment, a *Nde* I site within the translation initiation codon, and a *Hin*dIII site immediately downstream from the translation termination codon. The amplified DNA was ligated into pBluescript SK(+) and sequenced to verify the fidelity of the PCR. The open reading frame was excised and inserted into the *Es. coli* expression vector pARC306N to produce pSM145, a plasmid for expression of *crtE*.

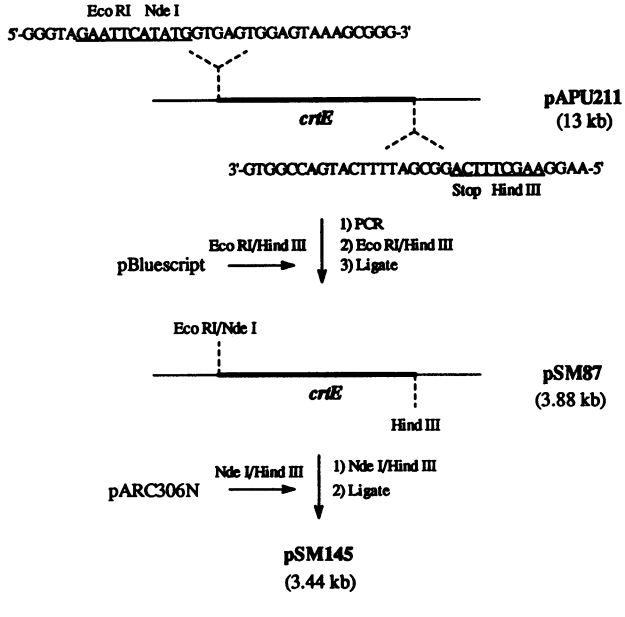

FIG. 2. Construction of an *Es. coli* expression vector for *crtE*. A cassette of the open reading frame of *crtE* flanked by *Eco*RI/*Nde* I and *Hin*dIII restriction sites was constructed by PCR amplification of the gene in pAPU211. The amplified DNA was treated with *Eco*RI/*Hin*dIII and ligated into the 2.95-kb *Eco*RI–*Hin*dIII fragment of pBluescript. After verifying that the PCR had not introduced unwanted mutations, the 0.93-kb *Nde* I–*Hin*dIII fragment from pSM87 was ligated into the 2.51-kb *Nde* I–*Hin*dIII fragment from pARC306N to create the *Es. coli* expression vector pSM145.

Expression of *crtE*. Cultures of JM101/pSM145 were grown on super broth containing ampicillin (100 μg/ml) at 37°C. The culture was allowed to grow for 12 h. The cells were harvested by centrifugation (7000 × *g*, 10 min); the cell paste was suspended in assay buffer consisting of 25 mM potassium phosphate (pH 7.2), 10 mM $MgCl_2$, 2 mM phenylmethyl sulfonyl fluoride, and 1 mM dithiothreitol; and the cells were disrupted by sonication. The extract was clarified by centrifugation (16,000 × *g*, 30 min) and used directly for assays and product studies.

GGPP Synthase Activity. GGPP synthase assays were performed by the acid-lability procedure (18). Assay samples contained 25 mM potassium phosphate (pH 7.2), 10 mM

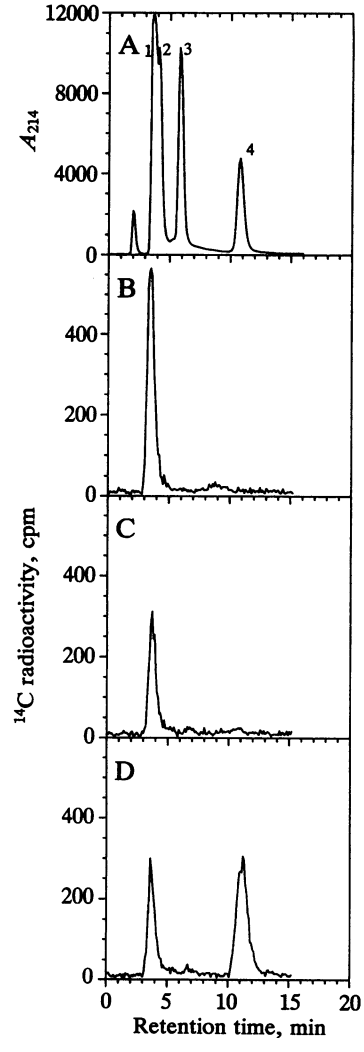

FIG. 3. HPLC traces of isoprene alcohols. Samples isolated from incubation mixtures containing cell-free extracts were coinjected with authentic samples of isopentenol (peak 1), geraniol (peak 2), farnesol (peak 3), and geranylgeraniol (peak 4) on a Waters μBondapak column and eluted with H_2O/methanol, 15:85 (vol/vol). Chromatograms were constructed from UV and radioactivity traces with simultaneous detection. (*A*) UV trace with detection at 215 nm. Identical patterns were seen for all samples. (*B*) Radioactivity trace of a control from incubation without addition of cell-free extract. (*C*) Radioactivity trace of a sample from incubation of [^{14}C]IPP and FPP with extract from JM101/pARC306N. (*D*) Radioactivity trace of a sample from incubation of [^{14}C]IPP and FPP with a cell-free extract from JM101/pSM145.

$MgCl_2$, 1 mM dithiothreitol, 1 mM KF, 100 μM FPP, and 50 μM [^{14}C]IPP. Product studies were conducted with identical samples. After incubation for 2 h at 37°C, the mixture was adjusted to pH 10.0 by addition of 10 μl of 2 M diethanolamine (pH 10.0) followed by addition of calf intestine alkaline phosphatase (2 units). Incubation was continued for 12 h. A 1-μl portion of a 1% solution of authentic samples of isopentenyl, farnesyl, and geranylgeraniol in methanol was added, and the mixture was extracted with hexane. A portion of the extract was analyzed by TLC on silica. Another portion was analyzed by HPLC.

RESULTS AND DISCUSSION

The *crtE* gene from *Er. herbicola* was cloned into the *Es. coli* expression vector pARC306N as described in Fig. 2. pSM145 contains the open reading frame of *crtE* behind the highly expressed *rec7* promoter, which is a hybrid based on the *Es. coli recA* promoter and the phage T7 gene *10* leader sequence (19). The translation initiation codon for *crtE* forms part of the *Nde* I (CATATG) restriction site. A strong transcription terminator is located at the end of the coding sequence.

Es. coli transformants containing pSM145 were grown to stationary phase and analyzed for GGPP synthase and phytoene synthase activities in cell-free extracts from fresh cultures. The assays were positive for conversion of [^{14}C]IPP and FPP to [^{14}C]GGPP (18) but were negative for synthesis of phytoene from [^{3}H]GGPP (1). Controls without addition of cell extracts or containing extracts from JM101/pARC306N were negative for both activities. Although *Es. coli* contains isoprenoid enzymes, including a higher prenyltransferase necessary for ubiquinone biosynthesis, the levels of these activities were extremely low and did not interfere with our assays or subsequent product studies (20).

The product of the reaction catalyzed by *crtE* was characterized by HPLC and TLC. In a typical experiment, [^{14}C]IPP and FPP were incubated with cell-free extracts from JM101/pSM145 for 2 h at 37°C. The pH of the mixture was adjusted to 10.0 by addition of diethanolamine buffer, and the diphosphate moieties were hydrolyzed by incubation with alkaline phosphatase for 12 h. Authentic samples of isopentenol (R_f = 0.24), geranylgeraniol (R_f = 0.30), and phytoene (R_f = 0.92) were added to the incubation mixture, and a portion of hexane-extractable material was spotted on a silica TLC plate. The plate was developed with 1:5 (vol/vol) ethyl acetate/hexane, lightly stained with iodine, and divided into sections. Radioactivity in the sections was then measured. All of the radioactivity comigrated with the C_5 and C_{20} alcohols. In another experiment, isopentenol, geraniol, farnesol, and geranylgeraniol were added to the phosphatase-treated mixture. The hexane extractables were analyzed by HPLC, and the eluant was simultaneously monitored in the UV at 215 nM and for radioactivity. The results for incubations without addition of cell-free extract (Fig. 3 *A* and *B*), with the addition of cell-free extract from JM101/pARC306N (Fig. 3 *A* and *C*), and with the addition of cell-free extract from JM101/pSM145 (Fig. 3 *A* and *D*) are shown. It is clear from these results that the gene product from *crtE* synthesizes GGPP from IPP and FPP, as expected for GGPP synthase. No evidence was found for more hydrophobic products corresponding to prephytoene alcohol or phytoene.

Maximal GGPP activity was seen in cultures from colonies of fresh transformants of pSM145. Activity decreased noticeably in cultures from week-old colonies. We have also transformed JM101 with expression plasmids from pARC306N containing genes for *S. cerevisiae* IPP isomerase (21) and FPP synthase (M. J. Yang and C.D.P., unpublished results), the two enzymes that immediately precede GGPP synthase in the isoprene pathway. Although cultures of these transformants slowly lost activity upon storage, the effect was much smaller than observed for GGPP synthase, and the levels of gene expression were higher than for *crtE*. The *rec7* promoter is not regulated in our system. Among the possible explanations for selection against high-level expression of the plasmid-encoded *crtE* gene are responses to cytotoxicity from increased levels of GGPP or from lowered levels of FPP required for biosynthesis of ubiquinone and bacterial dolichol.

Our biochemical studies provide direct evidence that *crtE* encodes GGPP synthase. This is further supported by comparisons of the amino acid sequences recently determined for several 1′-4 condensing prenyltransferases listed in Fig. 4. These include genes for FPP synthase from *Es. coli* (*ispA*) (22), *S. cerevisiae* (*fds1*) (23), *Rattus ratus* (24), and *Homo*

	Domain 1	Domain 2
FPP synthase		
ispA[a]		
Es. coli	73 vECIHAYSLIHDDLPAMDDDDLRRG 97	215 IGLAFQVQDDILDVVGDTATL----------GK 237
fds1[b]		
S. cerevisiae	89 IELLQAYFLVADD--MMDKSITRRG 111	232 LGEYFQIQDDYLDCFGTPEQI----------GK 254
Ra. ratus[c]	92 vELLQAFFLVLDD--IMDSSYTRRG 114	235 MGEFFQIQDDYLDLFGDPSVT----------GK 257
H. sapien[d]	92 vELLQAFFLVADD--IMDSSLTRRG 114	235 MGEFFQIQDDYLDLFGDPSVT----------GK 257
HPP synthase		
coq1[e]		
S. cerevisiae	182 vEMIHTASLLHDD--IVDHSDTRRG 204	356 LGICFQLVDDMLDFTV------GK--DL--GK 378
GGPP synthase		
al-3[f]		
N. crassa	170 ISMLHTASLLVDD--VEDNSVLRRG 191	297 IGLIFQIADDYHNLWREYTANKGMCEDLTEGK 328
crtE[g]		
Er. herbicola	82 vELTHTASLMLDDMPCMDNAELRRG 106	221 FGQAFQLLDDLRDDHPET----GKDRNKDAGK 248
crtE[h]		
Er. uredovora	83 vEMVHAASLILDDMPCMDDAKLRRG 107	222 LGQAFQLLDDLTDGMTDT----GKDSNQDAGK 249
crtE[i]		
Rh. capsulatus	69 LELMHCASLVHDDLPAFDNADIRRG 93	200 IGSAFGIADDLKDALMSAEAM----------GK 222

FIG. 4. Comparison among putative coding sequences from FPP synthase and GGPP synthase with *crtE*. Conserved bases are in boldface type. a, Ref. 22; b, ref. 23; c, ref. 24; d, ref. 25; e, ref. 6; f, ref. 26; g, ref. 11; h, ref. 12; i, ref. 13. HPP, hexaprenyl diphosphate.

Squalene synthase

Erg9[a]

S. cerevisiae [214] MGLFLQKTMIIRDYNEDLVDGRSFWPKEIWSQYA [247]

Phytoene synthase

tom5[b] [274] LGIANQLTNILRDVGEDARRGRVYLPQDELAQAG [307]

crtB[c]

Er. herbicola [164] LGLAFQLTNIARDIIDDAAIDRCYLPAEWLQDAG [197]

crtB[d]

Er. uredovora [151] LGLAFQLTNIARDIVDDAHAGRCYLPASWLEHEG [184]

crtB[e]

Er. capsulatus [156] LGLAMQMSNIARDVGEDARAGRLFLPTDWMVEEG [189]

FIG. 5. Comparison of putative coding sequences of *S. cerevisiae* Erg9 and tomato Tom5 with CrtB. Conserved bases are in boldface type. a, Ref. 14; b, ref. 24; c, ref. 11; d, ref. 12; e, ref. 13.

sapiens (25), and GGPP synthase from *Neurospora crassa* (al-3) (26), and hexaprenyl diphosphate synthase (*coq1*) from *S. cerevisiae* (6). All of the encoded 1′-4 prenyltransferases, including the GGPP synthase from *N. crassa*, contain two highly conserved aspartate-rich domains. Highly similar domains are also seen in all three *crtE* genes. Although fewer amino acid sequences are available for enzymes that catalyze 1′-1 couplings, DNA sequences for the yeast gene for squalene synthase (*erg9*) and the tomato gene for phytoene synthase (*tom5*) have been sequenced (14). As shown in Fig. 5, the putative amino acid sequences for *erg9* and *tom5* contain regions that are similar to highly homologous regions in *crtB* from *Er. herbicola*, *Er. uredovora*, and *Rh. capsulatus*. These sequence similarities are also consistent with related enzymatic activities and suggest that *crtB* in carotenogenic bacteria encodes a bifunction phytoene synthase. Thus, the early stages of carotenogenesis appears to be similar in both prokaryotes and eukaryotes.

We thank Dr. Matthias P. Meyer for helpful discussions. This work was supported by Grant GM 21328 from the National Institutes of Health and Grant DCB 8803825 from the National Science Foundation.

1. Gregonis, D. E. & Rilling, H. C. (1974) *Biochemistry* **13**, 1538–1542.
2. Poulter, C. D. (1990) *Acc. Chem. Res.* **23**, 70–77.
3. Dogbo, O., Laferrierc, A., d'Harlingue, A. & Camara, B. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7054–7058.
4. Sasiak, K. & Rilling, H. C. (1988) *Arch. Biochem. Biophys.* **260**, 622–627.
5. Poulter, C. D. & Rilling, H. C. (1978) *Acc. Chem. Res.* **11**, 307–313.
6. Ashby, M. N. & Edwards, P. A. (1990) *J. Biol. Chem.* **265**, 13157–13164.
7. Matsuoka, S., Sagami, H., Kurisaki, A. & Ogura, K. (1991) *J. Biol. Chem.* **266**, 3464–3468.
8. Seabra, M. C., Reiss, Y., Casey, P. J., Brown, M. S. & Goldstein, J. L. (1991) *Cell* **65**, 429–434.
9. Goodwin, T. W. (1971) *Biochem. J.* **123**, 293–329.
10. Coates, R. M. (1976) *Prog. Chem. Org. Nat. Prod.* **33**, 73–230.
11. Armstrong, G. A., Alberti, M. & Hearst, J. E. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9975–9979.
12. Misawa, N., Nakagawa, M., Kobayshi, K., Yamano, S., Izawa, Y., Nakamura, K. & Harashima, K. (1990) *J. Bacteriol.* **172**, 6704–6712.
13. Armstrong, G. A., Schmidt, A., Sandmann, G. & Hearst, J. E. (1990) *J. Biol. Chem.* **265**, 8329–8338.
14. Jennings, S. M., Tsay, Y. H., Fisch, T. M. & Robinson, G. W. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 6038–6042.
15. Davisson, V. J., Zabriskie, T. M. & Poulter, C. D. (1986) *Bioorg. Chem.* **14**, 46–54.
16. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
17. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
18. Laskovics, F. M., Krafcik, J. M. & Poulter, C. D. (1979) *J. Biol. Chem.* **254**, 9458–9463.
19. Olins, P. O., Devine, C. S., Ranwala, S. H. & Kavka, K. (1988) *Gene* **73**, 227–235.
20. Sherman, M. M., Petersen, L. & Poulter, C. D. (1989) *J. Bacteriol.* **171**, 3619–3628.
21. Street, I. P. & Poulter, C. D. (1990) *Biochemistry* **29**, 7531–7538.
22. Fujisaki, S., Hora, H., Nishimura, Y., Horiuchi, K. & Nishino, T. (1990) *J. Biochem. (Tokyo)* **108**, 995–1000.
23. Anderson, M. S., Yarger, J. G., Burck, C. L. & Poulter, C. D. (1989) *J. Biol. Chem.* **264**, 19176–19184.
24. Ashby, M. N. & Edwards, P. A. (1989) *J. Biol. Chem.* **264**, 635–640.
25. Sheares, B. T., White, S. S., Molowa, D. T., Chan, K., Ding, V. D. H., Kroon, P. A., Bostedor, R. G. & Karkas, J. D. (1989) *Biochemistry* **28**, 8129–8135.
26. Carattoli, A., Romano, H., Ballario, P., Morelli, G. & Macino, G. (1991) *J. Biol. Chem.* **266**, 5854–5859.