Cloning, sequencing, expression, purification and preliminary characterization of a type II dehydroquinase from *Helicobacter pylori*

Joanna R. BOTTOMLEY*, Christopher L. CLAYTON†, Peter A. CHALK† and Colin KLEANTHOUS*‡

*School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, U.K., and †Glaxo Wellcome Research and Development Ltd., Glaxo Wellcome Medicines Research Centre, Gunnels Wood Road, Stevenage, Herts. SG1 2NY, U.K.

A heat-stable dehydroquinase was purified to near homogeneity from a plate-grown suspension of the Gram-negative stomach pathogen *Helicobacter pylori*, and shown from both its subunit and native molecular masses to be a member of the type II family of dehydroquinases. This was confirmed by N-terminal amino acid sequence data. The gene encoding this activity was isolated following initial identification, by random sequencing of the *H. pylori* genome, of a 96 bp fragment, the translated sequence of which showed strong identity to a C-terminal region of other type II enzymes. Southern blot analysis of a cosmid library identified several potential clones, one of which complemented an *Escherichia coli* *aroD* point mutant strain deficient in host dehydroquinase. The gene encoding the *H. pylori* type II dehydro-

quinase (designated *aroQ*) was sequenced. The translated sequence was identical to the N-terminal sequence obtained directly from the purified protein, and showed strong identity to other members of the type II family of dehydroquinases. The enzyme was readily expressed in *E. coli* from a plasmid construct from which several milligrams of protein could be isolated, and the molecular mass of the protein was confirmed by electrospray MS. The *aroQ* gene in *H. pylori* may function in the central biosynthetic shikimate pathway of this bacterium, thus opening the way for the construction of attenuated strains as potential vaccines as well as offering a new target for selective enzyme inhibition.

INTRODUCTION

The enzyme dehydroquinase (DHQase; 3-dehydroquininate dehydratase; EC 4.2.1.10) catalyses the interconversion of 3-dehydroquininate and 3-dehydroshikimate, and is the central step in two distinct metabolic pathways: the biosynthetic shikimate pathway of prokaryotes and lower eukaryotes [1–3], and the catabolic quinate pathway identified in fungi such as *Neurospora crassa* and *Aspergillus nidulans* and the bacterium *Amycolatopsis methanolica* [4–6]. The end-product of the shikimate pathway is chorismic acid, which is a precursor for numerous aromatic compounds such as the aromatic amino acids, ubiquinone and vitamin E. The quinate pathway, on the other hand, is involved in the catabolism of quinic acid to acetyl-CoA via the β -ketoadipate pathway.

Although the shikimate pathway is ubiquitous in bacteria, fungi and plants, its organization varies substantially from one organism to another, especially with regard to the DHQase enzyme. *Escherichia coli* DHQase is monofunctional (encoded by the *aroD* gene) and has been cloned, sequenced and overexpressed [7]. *Salmonella typhi* [8,9], *Enterococcus faecalis* [10] and *Bacillus subtilis* [11] all possess monofunctional DHQases. In plants such as *Physcomitrella patens* [12], *Nicotiana tabacum* [13] and *Pisum sativum* [14], DHQase is part of a bifunctional polypeptide that also carries shikimate dehydrogenase activity, the next enzyme in the pathway. The DHQase in the shikimate pathway of fungi and yeast occurs as part of a pentafunctional polypeptide chain carrying the activities of the five consecutive steps leading to the production of 5-enoylpyruvylshikimate-3-phosphate [15,16]. In contrast to the varied organization of the shikimate pathway DHQase, its catabolic counterpart in the quinate pathway is

always monofunctional, although it is part of a transcriptionally coupled gene cluster [17,18].

In recent years it has become clear that the distinction of a DHQase as either a biosynthetic-like or a catabolic-like enzyme does not adequately describe the classification of this group of enzymes. The enzymes that have been isolated and characterized thus far fall into two groups: type I enzymes are generally, but not exclusively, found in a biosynthetic context, while type II enzymes have been found in both biosynthetic and catabolic pathways. The enzymological and biophysical properties of these two types of DHQase have been compared by Kleanthous et al. [19]. Type I enzymes have subunit molecular masses of ~25 kDa, form dimers in the case of the monofunctional enzymes and catalyse the elimination of water with *syn* stereochemistry [20]. The mechanism of catalysis by these enzymes is known to involve a Schiff base intermediate formed at the invariant residue Lys-170 [21], as well as a general acid/base catalysis involving His-143 [22,23]. Type II enzymes have subunit molecular masses of 16–18 kDa, oligomerize into dodecamers of ~200 kDa, are heat stable (in contrast to the type I enzymes) and catalyse the elimination reaction with *trans* stereochemistry [24]. Little is known of the mechanism of this class of enzyme. They do not require metals or cofactors for activity, even though they contain part of the core sequence of a nucleotide binding motif at their C-termini. The *Aspergillus nidulans* enzyme is known to undergo a ligand-induced conformational change at its N-terminus which exposes a reactive arginine residue (Arg-19) to modification by reagents such as 2,3-butanedione and phenylglyoxal [25,26].

A notable feature of the shikimate pathway is that it is absent from mammals, and so is a rich source of targets for anti-

Abbreviations used: DHQase, dehydroquinase; ESI-MS, electrospray ionization MS; ORF, open reading frame; PVDF, poly(vinylidene difluoride).

‡ To whom correspondence should be addressed.

The nucleotide sequence data reported have been deposited in the EMBL/GenBank/DBJ Nucleotide Sequence Databases under the accession no. X98878.

microbial agents and herbicides, as well as being useful in the production of live vaccines [27,28]. Examples include the herbicide glyphosate, which is a potent inhibitor of EPSP synthase [29], the last step in the pathway, and the development of oral vaccines against *Salmonella typhi* using attenuated strains created by deletion of the *aroD* (DHQase) and *aroC* (shikimate dehydrogenase) genes [30]. In the present paper, we focus on the DHQase from the Gram-negative bacterium *Helicobacter pylori* with a view to characterizing it as a target for anti-microbial therapy and as a candidate gene for the creation of attenuated strains of this organism. *H. pylori* is now recognized as one of the most common bacterial infections in humans. It has been connected with the onset of gastritis and peptic ulcer disease [31,32] and has been implicated as a contributing factor in gastric cancer [33]. We now report the isolation of the DHQase from *H. pylori*. The sequence and properties of this DHQase are consistent with it being a typical type II enzyme which may be part of the biosynthetic shikimate pathway of this organism.

MATERIALS AND METHODS

Bacterial strains and growth conditions

Helicobacter pylori strain NCTC 11637 was grown on chocolate blood colombia agar plates [prepared by adding 10% (v/v) horseblood to colombia agar base (Oxoid) and heating at 70 °C for 20 min] supplemented with amphotericin B, polymyxin B and vancomycin (each at 10 mg/ml) in an atmosphere of CO₂ (12%), O₂ (5%), H₂ (5%) and N₂ (78%) in a vain cabinet (Don Whitley). Plates were seeded with 100 µl of 24 h broth culture from 10 ml of Brain Heart Infusion (37 g/l; Oxoid) with horse spleen serum (5%, v/v; Oxoid) in a 25 cm² vertical tissue culture flask (Costar), and grown at 37 °C. Cells were harvested after 2 days of incubation when growth had become confluent, and extracts were prepared as described below. The *Escherichia coli* *aroD* point mutant AB2848, which is deficient in host DHQase, was used for cloning and expression of *H. pylori* DNA. In the absence of a functional DHQase, this strain requires aromatic amino acid supplements to grow on minimal media. For plasmid isolation and protein purification, *E. coli* AB2848 was grown in Luria-Bertani (LB) medium including ampicillin (200 µg/ml) at 37 °C with vigorous aeration.

Protein purification from *H. pylori*

Cells of *H. pylori* NCTC 11637 from 230 plates (~20 g of material) were resuspended in Tris/HCl (80 ml; 20 mM, pH 7.5) containing PMSF (0.5 mM) and sonicated (MSE Soniprep 150) on ice at maximum amplitude for 8 × 1 min pulses separated by 1 min intervals, and then centrifuged at 100 000 g for 40 min. The supernatant was dialysed against TEM buffer [50 mM Tris/HCl (pH 7.5), EDTA (1 mM), β-mercaptoethanol (1 mM)] containing PMSF (1.2 mM) at 4 °C and then incubated with gentle shaking for 30 min with DNase and RNase (0.1–0.2 mg per 30 ml of extract) at 37 °C. Cell debris was removed by centrifugation and the decanted supernatant was heated to 75 °C for 15 min. The resulting suspension was allowed to incubate on ice for 60 min. Precipitated proteins were removed by centrifugation and the retrieved supernatant was dialysed against TEM buffer containing KCl (40 mM) before further purification by anion-exchange (DEAE-15HR; Waters AP-1) FPLC (650E, Waters Advanced Protein Purification System, Millipore) at room temperature. Elution of the protein was performed with a 40 ml KCl gradient of 40 to 500 mM. Fractions displaying DHQase activity were pooled and dialysed against TEM buffer containing KCl (40 mM). The sample was concentrated by ultrafiltration

(Amicon) before being applied to a gel-filtration column (Waters; diol bonded silica; 10–300K) by FPLC in TEM buffer containing KCl (40 mM) at room temperature. The pool of appropriate fractions was concentrated by ultrafiltration (Amicon), dialysed against TEM buffer and concentrated further by lyophilization to a volume of ~100 µl.

SDS/PAGE analysis

Total cell protein and the progress of the enzyme purifications were analysed by standard SDS/PAGE or tricine SDS/PAGE (16% gels) [34]. All gels were run using a Hoefer system (Hoefer Scientific Instruments) according to the manufacturer's recommendations. Gels were stained for proteins with Coomassie Blue or silver stain.

N-terminal amino acid sequencing.

After electrophoresis, the purified protein (~1–3 µg) was transferred from a tricine SDS/polyacrylamide gel on to a poly(vinylidene difluoride) (PVDF) immobilizing membrane (Immobilon-P; Millipore) using the Pharmacia 2117 Multifor II electrophoresis system according to the manufacturer's recommendations. The membrane was stained for protein with Coomassie Blue. The N-terminal sequence was obtained using an Applied Biosystems model 470A/120A automated gas-phase sequencer equipped with on-line HPLC for detection of the phenylthiohydantoin derivatives.

Isolation of genomic DNA from *H. pylori* and screening of cosmid libraries

Genomic DNA was prepared from *H. pylori* 11637 by a standard method [35]. The genomic DNA was sonicated or partially restricted by various enzymes, size selected (1–2 kb) and shotgun cloned into pBluescript IKS⁺ or pCRScript SK⁺ vectors (Stratagene). Small-scale plasmid DNA was prepared from the random library clones and was used for cycle sequencing (using T3, T7, KS, SK and RP primers), and the sequence was determined using an ABI-373 sequencer (Applied Biosystems). In order to bias the random sequencing to reduce redundancy, PCR probes were generated from random clones and multi-probe hybridizations of a high-density *H. pylori* pUC18 plasmid library were performed. DNA sequences were analysed for protein identities by searching the databases using the BlastX1.4 program. One cloned sequence (designated Hp157T3) showed identity to the gene AROD-MYCTU in the Swiss Main database (X-blast probability score 4.4 × 10⁻⁵) corresponding to the type II DHQase of *Mycobacterium tuberculosis* [36]. This clone consisted of 96 bp of DNA: 5'-TCCACAAGCCGCTCCGGTGTAAGAATTTTTC-CTGAATTCTTCTCTAGCTTGAATGTTAGTGAGATGC-ACTTCAATAACGGGTTTGCCCGCTAGCAT-3'.

A 96 bp DHQase probe was prepared by ³²P random prime labelling (Prime it II; Stratagene) from NCTC 11637 genomic DNA to regenerate the Hp157T3 fragment. Cosmid DNA was restricted with *Hind*III and *Sal*I and subcloned into pUC18 restricted with the same enzymes. These subclones were then rescreened with the 96 bp DHQase probe. Bulk plasmid DNA was prepared from four positive subclones, restricted and run on an agarose gel for Southern blotting. Hybridizations were carried out under highly stringent conditions as described by Clayton et al. [37].

Preparation of plasmid DNA

Putative fragments of DNA encoding *H. pylori* type II DHQase were inserted into the plasmid pUC18 and transformed into *E.*

coli TG1. Plasmid DNA was prepared by modified alkaline lysis procedures (Qiagen Plasmid Kits). Plasmid DNA for sequencing was purified according to the Qiagen Plasmid Midi user manual designed for the preparation of up to 100 µg of plasmid DNA using a Qiagen-tip 100.

Transformations and complementation tests

E. coli AB2848 was transformed with pUC18 and the pUC18-derived recombinant plasmids containing the putative gene for *H. pylori* type II DHQase. Complementation was investigated by growth at 37 °C on minimal agar media containing ampicillin (200 µg/ml).

Screening for the expression of DHQase activity

The expression of DHQase activity by the pUC18-derived recombinant clones in AB2848 was determined by monitoring the activity of the transformed cells in LB broth containing ampicillin (50 µg/ml) with vigorous aeration at 37 °C over a period of 420 min. Samples (2 × 1 ml) of culture were removed at 30 min intervals. One aliquot was used to monitor cell density by attenuation at 550 nm (Philips 8700 series UV/Vis spectrophotometer) and the other was pelleted by centrifugation (8400 g) for 2 min. The supernatant was removed and the pellet stored at -20 °C until samples taken beyond the stationary phase had been collected.

Mini-preparations were carried out using the pelleted cells to obtain a crude cell extract for the analysis of DHQase activity and protein content. The cells were resuspended in sterile cell lysis buffer (Tris/HCl, 50 mM) at pH 8.0, containing EDTA (1 mM) and MgCl₂ (5 mM) to which freshly made PMSF (1 mM) and lysozyme (0.2 mg/ml) were added. The sample was mixed well and incubated at 37 °C for 30 min before being subjected to three freeze-thaw cycles using a solid CO₂/ethanol bath. DNase and RNase (0.04 mg/ml) were added and the sample was incubated at 37 °C for 30 min. After centrifuging (16400 g) for 5 min, the supernatant was retrieved and assayed for DHQase activity and the proteins were resolved by SDS/PAGE (16 % gels).

DNA sequencing and oligonucleotides

Double-stranded pUC18-derived recombinant plasmid DNA was sequenced on a Pharmacia LKB A.L.F. DNA Sequencer using the protocols supplied with the Pharmacia T7 DNA polymerase sequencing kit. Typical concentrations of plasmid DNA used were 0.01–0.02 A₂₆₀ unit/µl. The following synthetic fluorescently labelled oligonucleotide primers were synthesized on a Nucleic Acid Synthesis System-workstation (Millipore): CCA, 5' TCCACAAGCCGCTCCGGT 3'; CCB, 5' ATGCTA-GCGGGCAAACCC 3'; N33, 5' CCACAAGCCGCTCCGGTG-TAAG 3'; N34, 5' GTTATTGAAGTGCATCTCACTAAC 3'; JB3, 5' GCGTGAAATGCGATTCTCTCTC 3'; JB4, 5' CTTA-TGCTATAATGCGGTTTCATTCC 3'.

The oligonucleotide primers CCA and CCB were designed to prime from the beginning and end respectively of the Hp157T3 sequence obtained from the random sequencing of NCTC 11637 genomic DNA. The oligonucleotide sequencing primers N33 and N34 were designed to sequence outwards from each end of the Hp157T3 fragment. The sequencing primers JB3 and JB4 were designed from the nucleotide sequence derived from primers N33 and N34 in order to determine the nucleotide sequence of both DNA strands of the *H. pylori* type II DHQase gene. The primers were purified using an oligonucleotide purification device (Millipore).

Manipulation of nucleotide sequence data

Analysis of nucleotide sequence data obtained from the cloned *H. pylori* DNA, such as searching for open reading frames (ORFs) and computer-aided restriction digests, was carried out using the Apple Macintosh Gene Jockey package. Identification of similarities between the translated amino acid sequence and other proteins was carried out using the BLAST program.

Protein purification from *E. coli* AB2848

Cells carrying the pUC-derived recombinant plasmid containing the gene for *H. pylori* type II DHQase (*aroQ* in pCC1) were grown under the conditions described above. Type II DHQase was purified to homogeneity essentially as described by Bottomley et al. [26]. Cells (15.5 g from 2.5 litres of culture) were harvested by centrifugation (10400 g; 10 min), resuspended in a minimal amount of TEM buffer [50 mM Tris/HCl, EDTA (1 mM) and β-mercaptoethanol (1 mM), pH 7.5] containing PMSF (1.2 mM) and lysed using a French press [~ 62–100 kPa (9000–10000 psi)]. The lysate was incubated with RNase and DNase (0.1–0.2 mg/30 ml of extract) at 37 °C for 30 min and cell debris was removed by centrifugation (10400 g) for 30 min at 4 °C. The resulting supernatant was incubated at 70 °C for 15 min and then allowed to sit on ice for 60 min. Precipitated proteins were removed by centrifugation (10400 g) for 45 min at 4 °C and the decanted supernatant was dialysed against TEM buffer containing KCl (50 mM) for 12 h at 4 °C. The sample was then loaded on to a DE-52 anion-exchange column (50 ml) equilibrated with the same buffer at 1 ml/min. The column was washed extensively with this buffer until the absorbance of the eluant at 280 nm fell below 0.1. The remaining bound proteins were eluted with a 500 ml KCl gradient (50–500 mM) and the appropriate fractions containing DHQase activity were pooled and dialysed against TEM buffer containing KCl (50 mM) for 12 h at 4 °C. The sample was concentrated by ultrafiltration (Amicon) and then loaded on to a Sephacryl S-300 gel filtration column with TEM running buffer containing KCl (50 mM) at 1 ml/min. The appropriate fractions containing DHQase activity were pooled, concentrated by ultrafiltration and subjected to ammonium sulphate precipitation. Ammonium sulphate was added slowly to 30 % (w/v) while stirring vigorously at 4 °C. The precipitated proteins were removed by centrifugation (10400 g) for 30 min at 4 °C and resuspended in a minimal amount of TEM buffer. The redissolved pellet was dialysed against several changes of the same buffer at 4 °C and then stored at -20 °C.

Standard type II DHQase activity assay and protein determinations

Enzyme activity was determined by monitoring the formation of 3-dehydroshikimate at 234 nm ($\epsilon = 12 \times 10^3 \text{ M}^{-1} \cdot \text{cm}^{-1}$) at 25 °C. The standard assay mixture (1 ml) contained Tris/acetate buffer (50 mM, pH 7.0) and 3-dehydroquininate (1 mM) as substrate. Ammonium dehydroquininate was kindly supplied by Dr. A. Leech (UEA). K_m and k_{cat} values were determined by varying the substrate concentration over 20–800 µM, and the rate data were analysed using Enzfitter software (Biosoft, Cambridge, U.K.). Protein estimations during protein purifications were carried out using the Bradford assay and by amino acid analysis of purified enzyme from *E. coli*.

Electrospray MS (ESI-MS)

ESI-MS data were obtained with a Fisons VG Platform using a 1:1 (v/v) acetonitrile/water carrier solvent and horse heart

myoglobin as standard. The protein sample was dialysed against water to remove buffer salts and acidified with 0.1 % formic acid prior to analysis. A typical amount of protein analysed was 0.6 nmol per run. The data, averaged over 10 scans, were acquired in the m/z range 700–1300 and processed using the supplier's MassLynx software. Errors were calculated from mass determinations of three independent samples.

RESULTS AND DISCUSSION

Identification and purification of a type II DHQase from *H. pylori* extracts and determination of its N-terminal sequence

Extracts of *H. pylori* NCTC 11637 were prepared as described in the Materials and methods section and assayed for DHQase activity. Detectable levels of activity were observed, and the enzyme could be purified to apparent homogeneity in just three steps (Table 1 and Figure 1). The purification capitalized on the observation that this activity was heat stable ($> 70^\circ\text{C}$ for 10 min). Heat stability is a criterion by which type I and type II DHQases can be distinguished, since type II enzymes are resistant to heating whereas type I enzymes are thermally labile [19]. Hence our initial observation with the *H. pylori* extract immediately suggested that the majority of the activity detected was of a type II class. Over 90 % of this activity was recovered following the heating step (Table 1). Two chromatographic steps were then used in the purification: anion-exchange FPLC fol-

lowed by gel-filtration FPLC. The latter separation also gave the molecular mass of the native enzyme as ~ 230 kDa (results not shown). At this stage of the purification only the activity of the enzyme could be monitored since so little protein was present; nevertheless, it was clear from tricine SDS/PAGE (Figure 1) that a single silver-stained band was visible with an apparent molecular mass of ~ 16 – 18 kDa. The oligomeric structure of the enzyme is therefore suggested to comprise 12 subunits. This is entirely consistent with other type II enzymes, which are also dodecamers [19]. All of the protein from a single purification was loaded on to an SDS gel and blotted on to a PVDF membrane (see the Materials and methods section), and the Coomassie Blue-stained band was applied to a protein sequencer. A single clear sequence was obtained that ran for 10 cycles of the sequencer, i.e. MKILVIQGN. This matches the N-terminus of other type II DHQases (see Figure 3), thus confirming the identity of the *H. pylori* DHQase as a type II enzyme.

Identification of the *H. pylori* type II DHQase gene

During the course of random sequencing of the *H. pylori* NCTC 11637 genome (see the Materials and methods section) a clone was sequenced that was found to have low-level identity with the C-terminal end of the *aroQ* gene coding for the type II DHQase of *Mycobacterium tuberculosis* [36]. The translated amino acid sequence from the reverse strand of this 96 bp fragment shows a high degree of identity with the protein sequence of all type II DHQases (see Figure 3). PCR of *H. pylori* NCTC 11637 genomic DNA using oligonucleotide primers designed to the beginning and end of this sequence amplified the expected 96 bp fragment. This fragment was then ^{32}P -labelled and used to screen a *H. pylori* 11637 cosmid library. Two positive cosmids were detected, designated 2H4 and 4E8. Bulk cosmid DNA was isolated, and restriction digests of these two cosmids showed significantly different patterns. Bulk plasmid DNA was prepared from four positive subclones, restricted and run on an agarose gel for Southern blotting. The following fragments and hybridizations were observed: 2H4, hybridizing a 1.3 kb *HindIII* fragment (plus 1.4 kb and 300 bp fragments); 2H4, hybridizing an ~ 7.0 kb *SphI* fragment; 4E8, hybridizing a 500 bp *HindIII* fragment (plus 1.0 kb and 1.2 kb fragments); and 4E8, hybridizing a 2–3 kb *SphI* fragment.

Each of the four fragments that hybridized with the probe were cloned into pUC18, designated pCC1, pCC2, pCC3 and pCC4 respectively, and transformed into *E. coli* AB2848. This strain is unable to grow in the absence of aromatic amino acids, and so each of the constructs was tested for its ability to complement the mutant phenotype on minimal medium (see the Materials and methods section). Two of the plasmids (pCC1 and pCC2) complemented the *aroD* mutant strain, and so these were studied further.

Small-scale cultures of each AB2848 strain containing either pCC1 or pCC2 were grown on rich media and analysed for DHQase activity. AB2848/pCC1 expressed significantly more DHQase activity than AB2848/pCC2, while the control strain AB2848/pUC18 displayed no DHQase activity (results not shown). SDS/polyacrylamide gels (16 %) loaded with crude extract from some of these samples revealed an 18 kDa protein in the aliquots taken from the AB2848/pCC1 culture (results not shown). This clone was therefore characterized further.

DNA sequencing of the *H. pylori* DNA insert in the recombinant plasmid CC1

The total insert size of pCC1 was estimated from restriction analysis to be approx. 3 kb [38]. Sequencing primers were

Table 1 Purification scheme for type II DHQase from *Helicobacter pylori*

N.D., not determined; at this point all of the sample was applied to an SDS/polyacrylamide gel for blotting and sequencing. One unit of enzyme activity is 1 $\mu\text{mol}/\text{min}$.

	Volume (ml)	Protein (mg)	Activity (units)	Specific activity (units/mg)	Yield (%)	Purification (fold)
Crude extract	8.5	100	7.0	0.07	—	—
Heat treatment	8.1	36	6.4	0.18	91	2.6
Anion exchange	16	1.3	3.9	2.92	55	41.7
Gel filtration	3	N.D.	1.9	—	27	—

Figure 1 Purification of type II DHQase from extracts of *H. pylori*

Tricine SDS/PAGE (16 % gel) showing the purification of the enzyme from *H. pylori* extracts. Positions of molecular mass markers (M) are shown. Lane A, crude extract; lane B, heat-treated extract; lane C, following anion-exchange chromatography; lane D, following gel-filtration chromatography. An identical gel was run and blotted on to a PVDF membrane for protein sequencing.

1	ATATCAGTATGCGTTTTTTTAAAAAAGGATAA	
34	AAGCGCTTT[TTGACA]TCAAAGGTTTGTATGTTAGTCAAGCGGTAGTT	
81	TC[TTGACA]ATCCGTTCTTAGGGGA[TTTAAG]GGTTAAGATAAAACTT	
128	ATGC[TATAAT]CGGTTTCATTCCATATTCAACAAGGAGAAATAATA	
ORF1	Met Lys Ile Leu Val Ile Gln Gly Pro Asn Leu Asn	12
175	ATG AAA ATT TTA GTG ATT CAA GGG CCT AAT TTA AAC	
	Met Leu Gly His Arg Asp Pro Arg Leu Tyr Gly Met	24
211	ATG TTA GGA CAC AGA GAC CCA AGA CTT TAT GGC ATG	
	Val Thr Leu Asp Gln Ile His Glu Ile Met Gln Thr	36
247	GTA ACC TTA GAC CAA ATC CAT GAA ATC ATG CAA ACT	
	Phe Val Lys Gln Gly Asn Leu Asp Val Glu Leu Glu	48
283	TTC GTG AAA CAA GGC AAT TTA GAT GTG GAA TTA GAG	
	Phe Phe Gln Thr Asn Phe Glu Gly Glu Ile Ile Asp	60
319	TTT TTT CAA ACC AAT TTT GAG GGC GAA ATC ATT GAC	
	Lys Ile Gln Glu Ser Val Gly Ser Asp Tyr Glu Gly	72
355	AAG ATC CAA GAG AGC GTG GGC AGC GAT TAT GAA GGG	
	Ile Ile Ile Asn Pro Gly Ala Phe Ser His Thr Ser	84
391	ATT ATC ATT AAC CCT GGA GCG TTT TCG CAC ACT TCT	
	Ile Ala Ile Ala Asp Ala Ile Met Leu Ala Gly Lys	96
427	ATT GCG ATT GCG GAT GCG ATC ATG CTA GCG GGC AAA	
	Pro Val Ile Glu Val His Leu Thr Asn Ile Gln Ala	108
463	CCC GTT ATT GAA GTG CAT CTC ACT AAC ATT CAA GCT	
	Arg Glu Glu Phe Arg Lys Asn Ser Tyr Thr Gly Ala	120
499	AGA GAA GAA TTC AGG AAA AAT TCT TAC ACC GGA GCG	
	Ala Cys Gly Gly Val Ile Met Gly Phe Gly Pro Leu	132
535	GCT TGT GGA GGC GTG ATC ATG GGA TTT GGC CCG CTT	
	Gly Tyr Asn Met Ala Leu Met Ala Met Val Asn Ile	144
571	GGT TAC AAC ATG GCT TTA ATG GCG ATG GTC AAT ATT	
	Leu Ala Glu Met Lys Ala Phe Gln Glu Ala Gln Gln	156
607	TTA GCC GAG ATG AAA GGC TTC CAA GAA GCC CAA CAA	
	Asn Asn Pro Asn Asn Pro Ile Asn Asn Gln Lys Stop	167
643	AAC AAC CCT AAT AAC CCC ATT AAC AAT CAA AAA TAA	
	AAGGGGGTTATTG	
ORF2	Met Ser Gly Leu Glu Arg Glu Ser His Phe Thr Leu	12
692	ATG AGT GGA TTA GAG AGA GAA TCG CAT TTC ACG CTC	
	Asn Glu Asn Ala Met Phe Phe Glu Cys Ala Tyr Ser	24
728	AAT GAA AAC GCG ATG TTT TTT GAG TGC GCT TAT AGT	
	Cys Asp Asn Ala Leu Phe Leu Gln Leu Asp Asp Arg	36
764	TGC GAT AAC GCT TTG TTT TTG CAA TTA GAC GAT CGC	
	Ser Phe Phe Ile Thr Asp Ser Arg Tyr Thr Gln Glu	48
800	TCG TTT TTT ATC ACT GAT TCT CGC TAC ACT CAA GAA	
	Ala Lys Glu Ser Val Gln Pro Lys Asn Gly Val Leu	60
836	GCT AAA GAA AGC GTT CAG CCT AAA AAT GGC GTT TTA	
	Ala Glu Val Val Glu Ser Ser Asp Leu Ala Arg Ser	72
872	GCG GAA GTG GTA GAA TCT AGC GAT TTA GCG CGA AGT	
	Ala Gly Stop	74
908	GCG GGT TGA TTTGATCGCAAAAGTTCGGTTAAAAAACTCTTTT	
952	TTTGACCCCAATCAAGTGAATTACACACCTACAAGCGTTTAAATTC	
999	AGCGCTTGGGGATAAGGTTGCTTTAGAGGGCGTGCCTAGTTACCACC	
1046	GCCAAAACGCATCTTCGACGAGCATGAAATCCAACCTTCTCAAA	
1093	AAATCTCAAGCGCTGAATGTTGAAGCTTGCTCACTAAATGCCCTTA	

Figure 2 ORFs deduced from DNA sequencing of pCC1

ORF1 is type II DHQase, i.e. *aroQ*. The identity of ORF2 remains unknown. The boxed regions at the 5'-end of the *aroQ* gene are putative promoters, and the underlined region is the putative ribosome binding site.

designed based on the sequence of the original 96 bp fragment that had been identified from random sequencing of *H. pylori* NCTC 11637 genome DNA (see the Materials and methods section). Subsequent primers were made based on the sequence that was generated from these primers. Two ORFs were identified (ORF1 and ORF2; Figure 2) which were sequenced on both strands. The DNA sequence of ORF1 has a length of 504 bp and is observed to be A/T-rich (59%). Other *H. pylori* genes have also been found to be A/T-rich [39,40]. The reading frame

designated ORF2 in Figure 2 consists of 225 base residues and was also found to be A/T-rich (58%).

ORF1 translates to a polypeptide of 167 amino acid residues and shows a high degree of sequence identity with type II DHQases from other organisms (e.g. 41% with the *A. nidulans* type II DHQase; Figure 3). The sequence at the N-terminus is identical to that obtained from the purified enzyme by N-terminal amino acid sequencing. The gene encoding the type II DHQase of *H. pylori* has been designated *aroQ*, by analogy with the type II enzymes from *Mycobacterium tuberculosis* [36] and *Actinobacillus pleuropneumoniae* [41], which are thought to be shikimate pathway DHQases (see below).

ORF2 translates to a putative polypeptide of 74 amino acid residues, giving rise to a molecular mass of 8319Da. Searches of the sequence databases using the program BLAST failed to identify a related protein sequence, and hence the identity of this protein remains unknown.

Gene signals

E. coli nucleotide consensus sequences were identified for both the promoter and ribosome binding site in the *H. pylori* DNA upstream of ORF1 but not ORF2. No terminator sequence in the form of a complementary palindrome could be identified for either gene. There are two putative –35 boxes and two putative –10 boxes in the DNA sequence preceding the *aroQ* gene (Figure 2), such that there are four potential promoter motifs: A, T¹³TGACA(N₈₃)TATAAT; B, T⁴³TGACA(N₅₆)TTTAAG; C, T⁸³TGACA(N₄₃)TATAAT; D, T⁸³TGACA(N₁₆)TTTAAG.

Motifs A and C show perfect sequence identity with the consensus of sequences for the known promoter recognized by the RNA polymerase in *E. coli*. However, in *E. coli* the distance between the two motifs is usually 16–19 bp [42] and such a large gap between these promoter sequences, including that in B, would, therefore, be unusual. Motif D, on the other hand, fits the consensus that would be recognized by *E. coli*. Several promoter sequences that fit the *E. coli* consensus have been identified for other genes of *H. pylori* [40,43]. There are, however, examples of promoters that have been identified preceding *H. pylori* genes with a larger number of intervening base pairs, up to 32 bp {R. A. Moore, B. Beckthold, S. Wong, A. Kureishi and L. E. Bryan, unpublished work (GenBank accession no. 29481); [45]}. Thus motifs A to C identified upstream of the *H. pylori aroQ* gene may also be potential promoters.

The ATG codon at position 175 is presumed to be the start codon and, indeed, translation from here gives rise to the correct polypeptide mass observed by ESI-MS (see below) and an identical N-terminal sequence to the first 10 residues observed by N-terminal sequencing of the protein purified from the native host. Preceding this initiation codon by 9 bp is a putative ribosome binding sequence (AGGAG). The distance between the *aroQ* gene and the next ORF (ORF2) is 13 bp. Such a short distance does not allow ORF2 to have its own promoter but is sufficient to incorporate a ribosome binding site, although this does not seem to be present. This suggests that the two genes may be transcriptionally coupled, although at present it is not clear whether this gene is transcribed in *H. pylori*.

Expression in *E. coli*, purification and preliminary characterization of *H. pylori* type II DHQase

During the course of these studies it became clear that significant levels of *H. pylori* type II DHQase were expressed from pCC1 in *E. coli* AB2848 grown on rich medium (2 mg/l of culture). Since the levels of protein from *H. pylori* itself are very low, this

Figure 3 Sequence alignment of *aroQ* with all other type II DHQases

Sequences are shown of enzymes from *A. nidulans* [48], *N. crassa* [51], *M. tuberculosis* [36] and *A. pleuropneumoniae* [50]. The single underline indicates the 10 N-terminal amino acid residues that were sequenced from the protein purified from *H. pylori* (see the text), and the double underline is the translated sequence of the 96 bp probe that was originally identified by random sequencing of the *H. pylori* genome and which was used to identify the full-length clone from a cosmid library. The dashed line represents the partial nucleotide binding motif that has been identified at the C-termini of all type II DHQases [26].

permitted a more detailed analysis of the properties of the enzyme. The enzyme was purified to apparent homogeneity, deduced by SDS/PAGE, from AB2848 containing the plasmid pCC1 as described in the Materials and methods section. The molecular mass of the *E. coli*-expressed enzyme was determined by ESI-MS to be 18481.1 ± 2.0 Da, which is in good agreement with the calculated molecular mass of 18483.3 Da. The kinetic constants (determined from two independent purifications) k_{cat} , K_m and k_{cat}/K_m for the enzyme at pH 7.0 and 25 °C were $2.0 \pm 0.1 \text{ s}^{-1}$, $590 \pm 28 \text{ } \mu\text{M}$ and $3389 \text{ M}^{-1} \cdot \text{s}^{-1}$ respectively. The K_m values reported in the literature for type II DHQases fall broadly into two types. Enzymes with a low K_m (70–150 μM) have been identified in the quinate pathway of fungi and the bacterium *Amycolatopsis methanolica* [6], as well as in the shikimate pathway of *Mycobacterium tuberculosis* [46]. Previous to the present work, only a single high- K_m enzyme ($\sim 650 \text{ } \mu\text{M}$) had been reported, from the shikimate pathway of *Streptomyces coelicolor* [47]. It is interesting to note, therefore, that the K_m for the *H. pylori* enzyme falls into this latter group. In addition, the *H. pylori* enzyme has much lower k_{cat} and k_{cat}/K_m values than the *A. nidulans* (quininate) type II enzyme (1300 s^{-1} and $8.7 \times 10^6 \text{ M}^{-1} \cdot \text{s}^{-1}$ [19] respectively). Whether this is a consequence of, for example, different metabolic functions (see below) remains to be determined.

Distribution of DHQase enzymes in shikimate and quinate pathways

Type I and type II DHQases are distributed throughout the microbial kingdom. Prior to 1990, type I enzymes were regarded as the paradigm biosynthetic enzyme of the shikimate pathway of micro-organisms such as *E. coli* [7], whereas type II enzymes had only been defined in the quinate pathway of fungi [48]. Since that time, however, the characterization of DHQases from several sources has thrown new light on the distribution of these two classes of enzyme, especially in prokaryotes. Type I DHQases have been identified in the shikimate pathway of both the Gram-negative organism *Salmonella typhi* [9] and the Gram-positive organism *Enterococcus faecalis* [10], as well as in the inducible quinate pathway of the Gram-negative *Acinetobacter calcoaceticus* [49]. An even greater number of type II enzymes have been identified, primarily in the shikimate pathway of micro-organisms. These include the Gram-positive organisms *Streptomyces coelicolor* [47], *Mycobacterium tuberculosis* [36] and *Amycolatopsis methanolica* [6]. The latter type II DHQase is thought to serve a dual function in both the shikimate and quinate pathways. The enzyme has also been identified in the Gram-negative organisms *Actinobacillus pleuropneumoniae* [50] and *Acinetobacter calcoaceticus* [49].

The present report shows that the stomach pathogen *Helicobacter pylori* also possesses a type II DHQase enzyme which shows a high degree of sequence identity with other enzymes of this group (Figure 3), including the presence of a reactive arginine residue (equivalent to Arg-17) at its N-terminus [25,26] and a partial NAD⁺ binding motif at its C-terminus [26]. At the present time it is not possible from our data to unambiguously assign the enzyme from *H. pylori* to a role in the central biosynthetic shikimate pathway or a catabolic quinate pathway. While it is clear that the enzyme we have isolated is the major DHQase activity in extracts of *H. pylori*, we cannot exclude the possibility that a heat-labile type I enzyme may also be present in these extracts. In organisms such as *M. tuberculosis* the involvement of the type II DHQase in the shikimate pathway of the organism was deduced from the clustering of the *aroQ* gene with dehydroquinase synthase, another shikimate pathway enzyme [36]. The only other ORF identified in the present study failed to show any identity with other shikimate pathway enzymes. Preliminary evidence does, however, suggest that the type II DHQase of *H. pylori* functions in the shikimate pathway of this organism. The organism will not grow on quinic acid as a sole carbon source and, whereas the catabolic quinate pathway of fungi is inducible (by growth on quinic acid), we detect similar levels of type II DHQase activity in *H. pylori* extracts without induction (P. Chalk, unpublished work). In addition, deletion of the *aroQ* gene from *H. pylori* results in severely impaired growth even when the organism is cultured in rich media (C. L. Clayton and P. Chalk, unpublished work), implying a central role for this enzyme in metabolism.

In conclusion, we have cloned, sequenced, expressed and purified a type II DHQase from *H. pylori* which is likely to function in the shikimate pathway of this organism. Since the mechanism of type II enzymes is distinct from that of the type I class which predominates in other enterobacteriaceae, this raises the possibility that the enzyme from this stomach pathogen could be selectively inhibited while leaving other intestinal bacteria unaffected.

We gratefully acknowledge Dr. Andrew Leech (Norwich) for expert technical assistance with ESI-MS and for helpful discussions during this research. We also thank Kevin Bailey (Nottingham) for N-terminal amino acid sequencing, Barry Cauldwell (Dundee) for amino acid analysis and Dr. Richard James (Norwich) for helpful discussions. This work was funded by both the Science and Engineering Research Council and the Biotechnological and Biological Sciences Research Council. The Norwich Research Park ESI-MS facility is funded by the Biotechnological and Biological Sciences Research Council.

REFERENCES

- Haslam, E. (1974) The Shikimate Pathway, John Wiley and Sons, New York
- Bently, R. (1990) Biochem. Mol. Biol. Crit. Rev. **25**, 307–384
- Dewick, P. M. (1993) Nat. Prod. Rep. **10**, 233–263
- Hautala, J. A., Jacobson, J. W., Case, M. E. and Giles, N. H. (1975) J. Biol. Chem. **250**, 6008–6014
- Hawkins, A. R., Giles, H. H. and Kinghorn, J. R. (1982) Biochem. Genet. **20**, 271–286
- Euverink, G. J. W., Hessels, G. I., Vrijbloed, J. W., Coggins, J. R. and Dijkhuizen, L. (1992) J. Gen. Microbiol. **138**, 2449–2457
- Duncan, K., Chaudhuri, S., Campbell, M. S. and Coggins, J. R. (1986) Biochem. J. **238**, 475–483
- Servos, S., Chatfield, S., Hone, D., Levine, M., Dimitriadis, G., Pickard, D., Dougan, G., Fairweather, N. and Charles, I. G. (1991) J. Gen. Microbiol. **137**, 147–152
- Moore, J. D., Hawkins, A. R., Charles, I. G., Deka, R., Coggins, J. R., Cooper, A., Kelly, S. M. and Price, N. C. (1993) Biochem. J. **295**, 277–285
- Bensing, B. A. and Dunny, G. M. (1993) J. Bacteriol. **175**, 7421–7429
- Sorokin, A., Zumstein, E., Azevedo, V., Ehrlich, S. D. and Serron, P. (1993) Mol. Microbiol. **10**, 385–395
- Polley, L. D. (1978) Biochim. Biophys. Acta **526**, 259–266
- Bonner, C. A. and Jensen, R. A. (1994) Biochem. J. **302**, 11–14
- Deka, R., Anton, I. A., Dunbar, B. and Coggins, J. R. (1994) FEBS Lett. **349**, 397–402
- Charles, I. G., Keyte, J. W., Brammer, W. J., Smith, M. and Hawkins, A. R. (1986) Nucleic Acids Res. **14**, 2201–2213
- Duncan, K., Edwards, R. M. and Coggins, J. R. (1987) Biochem. J. **246**, 375–386
- Giles, N. H., Case, M. E., Baum, J., Geever, R., Huiet, L. and Patel, V. (1985) Microbiol. Rev. **49**, 328–358
- Hawkins, A. R., Lamb, H. K., Smith, M., Keyte, J. W. and Roberts, C. F. (1988) Mol. Gen. Genet. **214**, 224–231
- Kleanthous, C., Deka, R., Davies, K., Kelly, S. M., Cooper, A., Harding, S. E., Price, C. N., Hawkins, A. R. and Coggins, J. R. (1992) Biochem. J. **282**, 687–695
- Hanson, K. R. and Rose, I. A. (1963) Proc. Natl. Acad. Sci. U.S.A. **50**, 981–988
- Chaudhuri, S., Duncan, K., Graham, L. D. and Coggins, J. R. (1991) Biochem. J. **275**, 1–6
- Deka, R., Kleanthous, C. and Coggins, J. R. (1992) J. Biol. Chem. **267**, 22237–22242
- Leech, A. P., James, R., Coggins, J. R. and Kleanthous, C. (1995) J. Biol. Chem. **270**, 25827–25836
- Harris, J., Kleanthous, C., Coggins, J. R., Hawkins, A. R. and Abell, C. (1993) J. Chem. Soc. Chem. Commun. 1080–1081
- Krell, T., Pitt, A. R. and Coggins, J. R. (1995) FEBS Lett. **360**, 93–96
- Bottomley, J. R., Hawkins, A. R. and Kleanthous, C. (1996) Biochem. J. **319**, 269–278
- Kishore, G. M. and Shah, D. M. (1988) Annu. Rev. Biochem. **57**, 627–663
- O'Callaghan, D., Maskell, D., Liew, F. Y., Easmon, C. S. and Dougan, G. (1988) Infect. Immun. **56**, 419–423
- Steinruecken, H. C. and Amrhein, N. (1980) Biochem. Biophys. Res. Commun. **94**, 1207–1212
- Tacket, C. O., Hone, M. D., Curtiss, R., Kelly, S. M., Losonsky, G., Guers, L., Harris, A. M., Edelman, R. and Levine, M. M. (1992) Infect. Immun. **60**, 536–541
- Graham, D. Y. (1989) Gastroenterology **96**, 615–625
- Moss, S. and Calam, J. (1992) Gut **33**, 289–292
- Mitchell, H. M. (1993) in *Helicobacter pylori*, Biological and Clinical Practice (Goodwin, C. S. and Worsley, B. W., eds.), Chapter 6, pp. 95–114, CRC Press, Boca Raton
- Schagger, H. and Von Jagow, G. (1987) Anal. Biochem. **166**, 368–379
- Nowak, R. (1995) Science **267**, 172–174
- Garbe, T., Servos, S., Hawkins, A. R., Young, D., Dougan, G. and Charles, I. G. (1991) Mol. Gen. Genet. **228**, 385–392
- Clayton, C. L., Wren, B. W., Mullany, P., Topping, A. and Tabaqchali, S. (1989) Infect. Immun. **57**, 623–629
- Bottomley, J. R. (1995) Ph.D. Thesis, University of East Anglia, Norwich, U.K.
- Clayton, C. L. and Kleanthous, H. (1993) in *Helicobacter pylori*, Biological and Clinical Practice (Goodwin, C. S. and Worsley, B. W., eds.), Chapter 8, pp. 143–159, CRC Press, Boca Raton
- O'Toole, P. W., Logan, S. M., Kostrzynska, M., Wadstrom, T. and Trust, T. J. (1991) J. Bacteriol. **173**, 505–513
- Lalonde, G., Hanley, P. D. O., Stocker, B. A. D. and Denich, K. T. (1994) Mol. Microbiol. **11**, 273–280
- Hawley, D. K. and McClure, W. K. (1983) Nucleic Acids Res. **11**, 2237–2255
- Phadnis, S. H., Ilver, D. J., Janson, L., Normark, S. and Westblom, T. U. (1994) Infect. Immun. **62**, 1557–1565
- Reference deleted
- Evans, D. G., Karjalainen, T. K., Evans, D. J., Graham, D. Y. and Lee, C. H. (1993) J. Bacteriol. **175**, 674–683
- Moore, J. D. (1992) Ph.D. Thesis, University of Newcastle, U.K.
- White, P. J., Young, J., Hunter, I. S., Nimmo, H. G. and Coggins, J. R. (1990) Biochem. J. **265**, 735–738
- Da Silva, A. J. F., Whittington, H., Clements, J., Roberts, C. and Hawkins, A. R. (1986) Biochem. J. **240**, 481–488
- Elsmore, D. A. and Orntson, L. N. (1995) J. Bacteriol. **177**, 5971–5978
- Geever, R. F., Huiet, L., Baum, J. A., Tyler, B. M., Patal, V. B., Rutledge, B. J., Case, M. E. and Giles, N. H. (1989) J. Mol. Biol. **207**, 15–34