Comparative analysis of the QUTR transcription repressor protein and the three C-terminal domains of the pentafunctional AROM enzyme

Heather K. LAMB*||, Jonathan D. MOORE*‡||, Jeremy H. LAKEY*, Lisa J. LEVETT*, Kerry A. WHEELER*, Hugo LAGO*, John R. COGGINS† and Alastair R. HAWKINS*§

*Department of Biochemistry and Genetics, University of Newcastle upon Tyne, Framlington Place, Newcastle upon Tyne, NE2 4HH, U.K., and

†Division of Biochemistry and Molecular Biology, Davidson Building, University of Glasgow, Gilmorehill, Glasgow, G12 8QQ, U.K.

The AROM protein is a pentadomain protein catalysing steps two to six in the prechorismate section of the shikimate pathway in microbial eukaryotes. On the basis of amino acid sequence alignments and the properties of mutants unable to utilize quinic acid as a carbon source, the AROM protein has been proposed to be homologous throughout its length with the proteins regulating transcription of the genes necessary for quinate catabolism. The QUTR transcription repressor protein has been proposed to be homologous with the three C-terminal domains of the AROM protein and one-fifth of the penultimate N-terminal domain. We report here the results of experiments designed to overproduce the QUTR and AROM proteins and their constituent domains in *Escherichia coli*, the purpose being to facilitate domain purification and (in the case of AROM), complementation of *E. coli aro*[−] mutations in order to probe the degree to which individual domains are stable and functional. The 3-dehydroquinate dehydratase domain of the AROM protein and the 3-dehydroquinate dehydratase-like domain of the QUTR protein were purified in bulk and subjected to comparative CD spectroscopy and fluorescence emission spectroscopy. The CD spectra were found to be virtually superimposable. The fluorescence emission spectra of both domains had the signal from the tryptophan residues almost completely quenched, giving a tyrosine-dominated spectrum for both the AROM- and QUTR-derived domains. This unexpected observation was demonstrated to be due to a highly unusual environment provided by the tertiary structure, as addition of the denaturant guanidine

hydrochloride gave a typical tryptophan-dominated spectrum for both domains. The spectroscopy experiments had the potential to refute the biologically-based proposal for a common origin for the AROM and QUTR proteins; however, the combined biophysical data are consistent with the hypothesis. We have previously reported that the AROM dehydroquinate synthase and 3-dehydroquinate dehydratase are stable and functional as individual domains, but that the 5-*enol*-pyruvylshikimate-3-phosphate synthase is only active as part of the complete AROM protein or as a bi-domain fragment with dehydroquinate synthase. Here we report that the *aroA* gene (encoding the AROM protein) of *Aspergillus nidulans* contains a 53 nt intron in the extreme C-terminus of the shikimate dehydrogenase domain. This finding accounts for the previously reported observation that the AROM protein was unable to complement *aroE*[−] (lacking shikimate dehydrogenase) mutations in *E. coli*. When the intron is removed the correctly translated AROM protein is able to complement the *E. coli aroE*[−] mutation. An AROM-derived shikimate dehydrogenase domain is, however, non-functional, but function is restored in a bi-domain protein with 3-dehydroquinate dehydratase. This interaction is not entirely specific, as substitution of the 3-dehydroquinate dehydratase domain with the glutathione S-transferase protein partially restores enzyme activity. Similarly an AROM-derived shikimate kinase domain is non-functional, but is functional as part of the complete AROM protein, or as a bi-domain protein with 3-dehydroquinate dehydratase.

INTRODUCTION

In microbial eukaryotes, the two metabolites dehydroquinate (DHQ), and dehydroshikimate, are shared by the quinate utilization pathway and the shikimate pathway. In the quinate pathway, the production of the enzymes necessary for the utilization of quinate as a carbon source are produced in response to the presence of quinate exogenously provided in the growth medium [1,2]. In *Aspergillus nidulans*, this control over the production of the enzymes is regulated at the level of transcription initiation by the interaction of the products of the *qutA* and *qutR* genes. The *qutA* gene encodes an activator protein (QUTA), of the six-cysteine type, and the *qutR* gene encodes a repressor

protein (QUTR), that has been proposed to mediate its repressing effect by binding directly to the QUTA activator protein [2–4]. In the shikimate pathway, the two intermediates DHQ and dehydroshikimate are interconverted by the pentafunctional AROM enzyme, and the corresponding gene (*aroA*) encoding the AROM enzyme has been proposed to have evolved by multiple gene fusions. The yeast *Saccharomyces cerevisiae* does not possess the quinate pathway; however, it does have the equivalent pentafunctional AROM enzyme. In contrast, in *Escherichia coli* the equivalent five enzymes are not associated and are encoded by unlinked genes [5–8].

The sequences of the quinate utilization pathway genes from *A. nidulans* and *Neurospora crassa*, the genes encoding the

Abbreviations used: AROM, a pentadomain protein catalysing the conversion of 3-deoxy-D-arabinoheptulosonic acid-7-phosphate to 5-*enol*pyruvylshikimate 3-phosphate; DHQ, dehydroquinate; DHQS, 3-dehydroquinate synthase; EPSP, 5-*enol*pyruvylshikimate 3-phosphate; GST, glutathione S-transferase; IPTG, isopropylthio- β -D-galactoside; MC, metal-chelating; QDH, quinate 5-dehydrogenase (EC 1.1.1.24); QUTA, positively acting protein mediating the control of transcription of the quinic acid utilization gene cluster in *Aspergillus nidulans*; QUTR, negatively acting protein mediating the control of transcription of the quinic acid utilization gene cluster of *Aspergillus nidulans*; SDH, shikimate 5-dehydrogenase (EC 1.1.1.25); SK, shikimate kinase (EC 2.7.1.71).

‡ Present address: Department of Microbiology, Duke University Medical Center, Durham, NC 27710, U.S.A.

§ To whom correspondence should be addressed.

|| The first two authors contributed equally and are cited in alphabetical order.

Figure 1 Modular structure of the enzymes and regulatory proteins comprising the quinate and pre-chorismate shikimate pathways in *A. nidulans*

The boxes denoted *aro* designate genes from bacteria encoding monofunctional shikimate pathway enzymes: *aroA*, *E. coli* EPSP synthase (EPSPS); *aroB*, *E. coli* DHQ synthase (DHS); *aroD*, *E. coli* type-I 3-dehydroquinate dehydratase (DQ); *aroE*, *E. coli* shikimate dehydrogenase (SDH); *aroL*, *E. coli* shikimate kinase (SK); and *aroQ*, *Mycobacterium tuberculosis* type-II 3-dehydroquinate dehydratase (DQII). AROM designates the modular structure of the AROM protein of *A. nidulans* which is specified by the *aromA* gene that arose by the fusion of the bacterial *aroA*, *B*, *D*, *E* and *L* genes. The boxes denoting the genes of the *qut* pathway are: *qutA*, activator; *qutB*, quinate/shikimate dehydrogenase (QDH); *qutD*, permease; *qutE*, type-II 3-dehydroquinate dehydratase (DQII); *qutG*, possibly a phosphatase; and *qutR*, repressor. The genes encoding the quinate pathway activator and repressor proteins are proposed to have arisen by duplication of the *aromA* gene followed by cleavage in the DNA sequence specifying the C-terminus of the EPSP synthase domain (*aroA* equivalent). The *aromA* sequence encoding the DHQS domain in the duplicated copy is proposed to have undergone an event in which the DNA sequence encoding a pre-formed zinc binuclear cluster motif (designated Zn) has become inserted into its N-terminus. The box denoted *aroQ* designates the type-II 3-dehydroquinate dehydratase from *M. tuberculosis*. The arrows between boxes are there to indicate that genes denoted by the boxes are related to one another. *A. nidulans*, *E. coli* and *M. tuberculosis* have been used to give specific examples, as the homologies between the various genes are evident across a wide range of species. Reproduced with permission from Hawkins et al. [2].

S. cerevisiae and *A. nidulans* AROM proteins, and the genes encoding the equivalent monofunctional enzymes in *E. coli* have been determined (see [2,7] and references therein). Comparison of the deduced amino acid sequences has revealed a complex web of evolutionary relationships, which have recently been reviewed (see Figure 1 and [2]). In the context of this work, the most remarkable proposal was that the pentafunctional AROM enzyme is homologous throughout its length with the QUTA and QUTR proteins. The QUTA protein is homologous with the N-terminal two domains and the QUTR protein is homologous with the C-terminal three domains of the AROM protein [4,9,10]. It has been proposed that part of the function of the three repressor domains is to act as a molecular sensor that detects the presence of quinate pathway intermediates as a prerequisite for the presumed conformational changes necessary for the control of transcription regulation [10]. If the proposed evolutionary origins and role as a molecular sensor are correct, the QUTR repressor protein should retain a similar fold structure to the AROM enzyme domains with which it is proposed to be homologous.

We have previously reported the characterization of some of the constituent domains of the AROM protein, demonstrating in particular that the 3-dehydroquinate synthase (DHQS; EC 4.6.1.3) and 3-dehydroquinate dehydratase (EC 4.2.1.10) domains are soluble, stable and enzymically active as monofunctional proteins [11–13]. All previous attempts to characterize the shikimate 5-dehydrogenase (SDH; EC 1.1.1.25) and shikimate kinase (SK; EC 2.7.1.71) domains (which along with the 3-dehydroquinate dehydratase domain, constitute the three AROM domains that are homologous with the QUTR protein) in *E. coli* were unsuccessful, due to the lack of suitable auxotrophic strains and the inability of a chromosomal clone of the *aromA* gene to

direct overexpression of a full-length active AROM protein in *E. coli*.

In this paper we report the results of experiments primarily designed to test the prediction that the QUTR protein and the three C-terminal domains of the AROM protein have similar tertiary structures and physical properties.

MATERIALS AND METHODS

The origins of the *E. coli* and recombinant λ strains used in this work and the media for their propagation have been described previously [14–19]. Materials were obtained as described in [20] and oligonucleotides were purchased from the Central Facility for Molecular Biology at the University of Newcastle upon Tyne. The plasmid pET21a was obtained from AMS Biotechnology, Witney, Oxon., U.K. Recombinant DNA techniques followed standard methods [21], and DNA fragments were purified by the freeze–squeeze method [22]. PCR amplification was performed as in [11], and the use of reverse transcriptase (Pharmacia) followed manufacturer's recommended conditions. Electrophoresis in the presence of SDS was performed as described in [23] with a 5% (w/v) polyacrylamide stacking gel and an 8.75% (w/v) polyacrylamide separating gel. Gels were stained for protein with Coomassie Blue, or for SDH activity as described in [24]. The growth of *E. coli* cells, complementation tests, and induction with isopropylthio- β -D-galactoside (IPTG) were as previously described [18,20]. For the low-temperature-induction experiments, 2 ml cultures of *E. coli* in ampicillin-supplemented (50 μ g/ml) Luria broth were grown at 37 °C with vigorous agitation until an attenuation of $D = 0.5$ was attained. The cultures were then transferred to a precooled incubator at 27 °C or 16 °C for 1 h before IPTG was added to a final concentration

of 200 µg/ml and the production of recombinant proteins induced by incubation for 45–120 min. Cells were then harvested by centrifugation for 1 min in a precooled microcentrifuge. To scale up the low-temperature inductions, 200 ml cultures of *E. coli* (in 2 litre baffled flasks) were induced in a 16 °C incubator after rapidly cooling the culture to 10–15 °C from the 37 °C growth temperature in an ice/salt/water bath. DHQS, 3-dehydroquininate dehydratase and SDH assays were performed at 25 °C as described in [24] in a CARY 4 UV-spectrophotometer. Purification of fusion proteins containing the glutathione S-transferase (GST) affinity domain of *Schistosoma japonicum* was performed as previously described except that 4 ml bed volumes of GSH-Sepharose 4B were used in conjunction with cells recovered from 2 litres of starting medium [13]. Purification of proteins containing the 4.47 kDa metal-chelating (MC) domain specified by pRSETB were purified from crude cell sonicates on a 4 ml bed volume of chelating Sepharose that had been charged with zinc and washed with de-ionized water before equilibration with the appropriate buffer. The full purification protocol is given in the Results section. Site-directed mutagenesis of the QUTR 3-dehydroquininate dehydratase-like domain was carried out by the PCR method described in [25] with minor modifications as previously described [11].

Fluorescence emission spectroscopy

Fluorescence emission spectra were recorded at 20 °C in an SLM 8100 spectrofluorimeter operating in ratio mode with band widths of 4 nm for excitation (280 nm) and emission (range 290–430 nm); 5-mm-pathlength quartz glass cuvettes (from Hellma) containing 400 µl of sample were used.

CD measurements

CD spectra were recorded at 20 °C using a Jobin-Yvon CD6 spectrometer. Mean residue delta epsilon ($\Delta\epsilon$) values were calculated using the predicted protein sequences and concentrations were derived from absorption measurements at 280 nm on a Cary 4E absorption spectrometer. Theoretical molar absorption coefficients were calculated from the predicted sequence and were shown to equal the values obtained by the method of Gill and von Hippel [26]. Samples of the QUTR domain which were insoluble at low salt concentrations were diluted into 250 mM NaF to improve far-UV transparency.

Native molecular mass determinations

Solution native molecular mass determinations were carried out by FPLC using a Superdex 75 HR 10/30 column (Pharmacia) equilibrated with 20 mM Tris/HCl (pH 7.2)/0.25 M NaCl and calibrated with BSA (66 kDa), chicken ovalbumin (43 kDa), chymotrypsin (25 kDa), cytochrome *c* (12.4 kDa) and aprotinin (6.5 kDa). The column was eluted at 30 ml/h and the void volume determined as 7.7 ml using Blue Dextran 2000 as a marker.

Amino acid sequence determination

Amino acid sequence determination was performed on a Beckman LF3000 protein sequencer in the University of Newcastle-upon-Tyne Facility for Molecular Biology using protein recovered on Beckman Hyperbond membrane. Desired proteins were subject to SDS/PAGE and recovered for amino acid sequence analysis according to the following protocol: Hyperbond membrane was soaked for 5 min in methanol and transferred to 100 ml of blotting buffer consisting of 10 mM 3-(cyclohexamine)-1-propanesulphonic acid, pH 11.0 (CAPS

buffer) for a further 10 min. After electrophoresis, the 1.5-mm-thick acrylamide gel was rinsed twice in 100 ml of ultra-pure water containing 0.2 % dithiothreitol. The gel was then equilibrated in 100 ml of CAPS buffer containing 0.2 % dithiothreitol at 4 °C for 30 min with one buffer change after 15 min. The proteins in the gel were then electroblotted on to the Hyperbond membrane using a Bio-Rad Trans Blot Cell at a constant voltage of 105 V for 3 h. After blotting, the Hyperbond membrane was rinsed with 2 × 100 ml of 0.2 % dithiothreitol and stained with 0.4 % Coomassie Blue (in methanol/water/acetic acid; 5:4:1, by vol.) for 20 min at 55 °C. The stained membrane was destained with the same solvent lacking Coomassie Blue. The stained protein band was then excised from the membrane with a scalpel blade and used directly for amino acid sequence determination.

RESULTS

The *A. nidulans* *aromA* gene contains a 53 nt intron near the 3' end of the sequence encoding the SDH domain

The *A. nidulans* AROM protein had previously been reported to be encoded by a single 4812 nt open reading frame [5]. However, expression of the entire coding sequence of the AROM protein in *E. coli* using the commercially available expression vector pKK233-2 had not led to complementation of the *aroE*[−] mutation in *E. coli* strain GLW55 [18]. When the 3' sequence of the *aromA* gene is translated in all three reading frames and compared with the available sequences of other SDH and quinate/shikimate dehydrogenases, the amino acid motif WTG was apparent in one of the 'incorrect' reading frames. This amino acid motif is present in the yeast AROM protein and the *A. nidulans* quinate/shikimate dehydrogenase. A related variant (FTG) is present in the *A. nidulans* and *N. crassa* [27,28] quinate pathway transcription repressor proteins, thereby strongly suggesting that the sequence of the coding region of the *aromA* gene had been incorrectly determined or translated. Re-analysis of the original DNA sequencing autoradiographs [5] failed to detect a sequencing error in this region. This suggests the obvious possibility of the presence of an intron which was causing a frame shift when the sequence was translated in *E. coli*.

In order to look for the presence of an intron, cDNA was prepared from total RNA and the sequences encoding the SDH domain were amplified by means of the PCR. The sequence from the *Hind*III site at nucleotide 4400 in the *aromA* coding sequence

ANAROM.1521	GLEVLVGGGMYQFKYWTGISPLYESARACSSPLI
SCAROM.1556	GSQMLVHQGMVQFELKWTGFKGPFKAI FDAVTKEL
ANQUTR.870	GLDVLPEQGFAQFELFTGCRAPRRLMRTVILQEYK
NCQA1S.861	GLDLLPEQGFAQFELFTGRRAPRRLMRREVLRAYP
ANQUTB.283	GSEALIWQGLEQARVWTGKDVVSEPGVLKQVAFV
NCQA3.291	GTEALIWQGLEQVCI RPELTLYCGGQLN

Figure 2 Amino acid sequence alignments of the extreme C-termini of the *A. nidulans* and *S. cerevisiae* AROM proteins

The extreme C-terminal amino acid sequences of the *A. nidulans* (A.N. AROM) and *S. cerevisiae* (S.C. AROM). AROM proteins are shown aligned with the corresponding regions of *A. nidulans* (ANQUTR) and *N. crassa* (NCQA1S) quinate acid utilization gene cluster transcription-repressing proteins and the *A. nidulans* (ANQUTB) and *N. crassa* (NCQA3) quinate/shikimate dehydrogenase enzymes. Individual amino acids are boxed where one or both of the repressor sequences has an identical or similar residue to at least one of the four enzymes. For the purposes of this Figure the following amino acids (single-letter code) are taken to be similar: A and G; I, L, M and V; S and T; D and E; K and R; N and Q; F, Y and W. The number following each acronym is the position in the polypeptide chain at which each sequence starts, and all sequences end at their natural C-terminus.

Table 1 The growth responses of three *E. coli aro⁻* strains transformed with recombinant plasmids specifying different combinations of AROM protein domains

The symbol '—' indicates background growth and the symbol '+' indicates growth graded such that '+++' is wild-type growth. The symbol '—/+' indicates poor growth that is, however, clearly greater than the background. *E. coli aroE⁻* strains lack SDH, *aroD⁻* strains lack the type 1 3-dehydroquinate dehydratase and *aroL⁻ aroK⁻* lack both of the SK isoenzymes. All three strains require aromatic amino acid supplementation to grow on minimal medium unless the missing enzymic activities can be supplied by complementation using recombinant plasmids containing *aromA* DNA. The AROM protein specified by plasmid pEA1 translates the intron in the SDH domain and therefore produces a protein with the 42 C-terminal amino acids translated out of frame. Consequently the SDH is non-functional. The AROM protein specified by plasmid pEKA15 is wild type and has a functional SDH. In order to facilitate the testing of the ability of the constituent domains of the AROM protein to complement *E. coli aroE⁻* mutants, the coding sequences of these constructs were subcloned into pUC18 or pUC19 where they were placed under the control of the IPTG-inducible *lac* promoter. Plasmids pETA3, pETA7 and pETA13 were digested with *Xba*I (which cuts between the T7 promoter and *lac* operator in the vector sequence of pET21a) and *Eco*RI, the fragment containing the AROM-derived sequence was isolated and ligated into similarly digested pUC19 DNA to produce pUCA3, pUCA7 and pUCA13 respectively. Plasmid pUCA5 was constructed by ligating an *Xba*I/*Hind*III restriction fragment containing the SK and 3-dehydroquinate dehydratase (DHQase) domains of the AROM protein isolated from pETA5 into similarly digested pUC18. As it was found that pUCA13 was unable to fully complement the *E. coli aroE⁻* mutation, plasmid pTR65 was constructed to ensure that the start of the SDH domain had not been omitted from pUCA13. Plasmid pTR66 was constructed to ensure that we had not omitted the 3' end of the SK domain. Plasmids pFP69 and pFP70 were constructed by ligating *Bam*H1/*Eco*RI-digested PCR products (the template was pETA7) encoding the same SDH sequences included in pUCA13 and pTR65 respectively into similarly digested pGEX2T.

Plasmid	Amino acid co-ordinates	Domains encoded	<i>E. coli</i> strain for complementation	Growth on minimal media without aromatic acid supplementation	
				30 °C	37 °C
pTrc99A	None	None	<i>aroE⁻</i>	—	—
pEA1	1–1603	All 5 (but SDH contains intron)	<i>aroE⁻</i>	—	—
pEKA15	1–1585	All 5	<i>aroE⁻</i>	+++	+++
pUCA13	1285–1585	SDH	<i>aroE⁻</i>	—/+	—
pTR65	1256–1585	SDH	<i>aroE⁻</i>	—/+	—
pUCA7	1034–1585	DHQase–SDH	<i>aroE⁻</i>	++	++
pGEX-2T	None	(GST)	<i>aroE⁻</i>	—	—
pFP69	1285–1585	(GST)–SDH	<i>aroE⁻</i>	+++	—
pFP70	1256–1585	(GST)–SDH	<i>aroE⁻</i>	+++	—
pTrc99A	None	None	<i>aroD⁻</i>	—	—
pEA1	1–1603	All 5 (but SDH contains intron)	<i>aroD⁻</i>	+++	+++
pEKA15	1–1585	All 5	<i>aroD⁻</i>	+++	+++
pUCA5	862–1290	SK–DHQase	<i>aroD⁻</i>	+++	—
pUCA7	1034–1585	DHQase–SDH	<i>aroD⁻</i>	+++	—
pKK36	1033–1290	DHQase	<i>aroD⁻</i>	+++	+++
pGEX-2T	None	(GST)	<i>aroD⁻</i>	—	—
pXF4	1036–1290	(GST)–DHQase	<i>aroD⁻</i>	+++	—
pTrc99A	None	None	<i>aroL⁻; aroK⁻</i>	—	—
pEA1	1–1603	All 5 (but SDH contains intron)	<i>aroL⁻; aroK⁻</i>	+++	+++
pEKA15	1–1585	All 5	<i>aroL⁻; aroK⁻</i>	+++	+++
pUCA5	862–1290	SK–DHQase	<i>aroL⁻; aroK⁻</i>	++	++
pUCA3	862–1034	SK	<i>aroL⁻; aroK⁻</i>	—	—
pTR66	862–1053	SK	<i>aroL⁻; aroK⁻</i>	—	—

to an *Eco*RI sequence included in the oligonucleotide complementary to the 3' end of the *aromA* coding sequence, was subcloned into M13mp18 and 19 and sequenced in both directions. It was concluded that the *aromA* gene contained a 53 nt intron very near the end of the coding sequence (corresponding to nucleotide co-ordinates 4685–4737 in [5]). The correctly translated sequence of the extreme C-terminus of the AROM SDH is shown in Figure 2 compared with the equivalent domain from the *S. cerevisiae* AROM protein, the *A. nidulans* and *N. crassa* quinate/shikimate dehydrogenases [5,7,27,29,30]. Inspection of Figure 2 shows that removal of the intron does produce a new C-terminal sequence for the AROM protein that is much more similar to the other enzymes, and includes the WTG amino acid motif.

Complementation of *aro⁻* mutations and overproduction in *E. coli* of the complete AROM protein and its constituent domains

The discovery of the intron near the 3' end of the SDH-encoding domain and the availability of strain ALO807, an *E. coli aroL⁻ aroK⁻* double mutant strain (thereby lacking SK activity; see [19]) enabled us to address the question of whether or not the SK and SDH domains of the pentafunctional AROM protein were

capable of independent function by use of a sensitive *in vivo* assay. The basis of this assay is to produce the AROM protein or deletion forms thereof in *E. coli aro⁻* mutants that are unable to grow in minimal medium lacking aromatic amino acid supplementation. If the protein specified by a particular plasmid is able to produce any functional enzyme that can substitute for the missing *E. coli* enzyme, then the transformed *E. coli* strain will be able to grow in minimal medium in the absence of exogenously supplied aromatic amino acids. The *in vivo* assay is functional even when levels of enzyme activity which are undetectable by *in vitro* assay are produced, or when large amounts of insoluble protein are produced [11,18,31].

We also wished to overproduce the various AROM-derived proteins in *E. coli* as the first step towards bulk purification and characterization of their biochemical/biophysical properties. Additionally, because the AROM protein has five domains, we wished to look at the effect of substituting the GST domain of *Schistosoma japonicum*, contained within plasmid pGEX-2T, for the naturally occurring N-terminal domain of some individual AROM protein-derived domains. The domain structure and order are shown in bar diagram form in Figure 1.

In order to pursue these two parallel strategies, two series of recombinant *E. coli* plasmids were constructed. One series utilized

Figure 3 SDS/PAGE analysis of *E. coli* strain BL21 (DE3) harbouring recombinant plasmids encoding the complete AROM protein and subdomains thereof

(a) Each numbered lane was loaded with approximately 15 μ g of total protein from boiled extracts of BL21 (DE3) cells containing: (1) pET21a uninduced; (2) pET21a IPTG-induced; (3) pETA4 uninduced; (4) pETA4 IPTG-induced; (5) pETA5 uninduced; (6) pETA5 IPTG-induced; (7) pETA7 uninduced; (8) pETA7 IPTG-induced; (9) pETA9 uninduced; (10) pETA9 IPTG-induced; (11) pETA14 uninduced; (12) pETA14 IPTG-induced; (13) pGEX-2T uninduced; (14) pGEX-2T IPTG-induced; (15) pFP70 uninduced; (16) pFP70 IPTG-induced plasmids. Lane M contains the molecular mass markers detailed in the legend to Figure 4. The *aromA* DNA regions contained within each of the plasmids designed to overproduce the AROM protein or its subdomains are shown in the listing in (b). The template for the PCR reactions to produce pETA3, pETA4 and pETA5 was λ ARH8.1; for the PCR products that encoded the SDH domain cDNA produced from *A. nidulans* strain 1314 was used. Some of the restriction fragments used in the construction of these plasmids were derived from the products of a PCR which generated the 5' end of some of the domains. In these cases, the 5' PCR primer introduced an ATG initiation codon along with the *Nde*I site, thus converting the first residue listed into a methionine codon. All parts of the AROM expression vectors derived from PCR-amplified fragments were subcloned into M13mp18 and 19 and sequenced to verify that no adventitious changes had been introduced by the PCR process. The restriction sites introduced at the 5' and 3' ends as part of the subcloning strategy are shown. Abbreviations used: aa, amino acids; Skinase, shikimate kinase; DHase, 3-dehydroquinate dehydratase; SDHase, shikimate dehydrogenase.

the high-level expression vector pET21a, which makes use of the powerful IPTG-inducible T7 promoter. These plasmids, however, require the presence of the T7 RNA polymerase in the host *E. coli* strain for protein overproduction to occur, but none of the necessary *E. coli* strains needed for the sensitive *in vivo* complementation assay for enzyme activity carried the T7 RNA polymerase-encoding gene. A second series of recombinant plasmids using the *trc*, *tac* or *lac* IPTG-inducible promoters present in the *E. coli* expression vectors pTrc99a, pGEX-2T, pKK233/2, or pUC18/19 were constructed. The purpose of producing this second series of plasmids was to give added flexibility, as they had the potential to direct protein overproduction but could also be used in any strain for the complementation analysis, as the promoters utilize the host RNA polymerase. Previous experience has shown, however, that significant numbers of recombinant plasmids utilizing the *trc*, *tac* or *lac* promoters were unable to direct protein overproduction, whereas the identical sequences under the control of the T7 promoter directed protein overproduction far more reliably. The various plasmids constructed are summarized in the legend to Table 1 and Figure 3.

Table 1 summarizes the ability of the appropriate constructs to complement the *E. coli* SDH- or SK-deficient mutants. Figure 3 shows the ability of the appropriate plasmid constructs to direct IPTG-inducible protein overproduction in *E. coli*. Inspection of Table 1 reveals that SDH is enzymically inactive as a monofunctional domain. It is, however, active as part of the complete AROM protein or as a bifunctional protein with 3-dehydroquinate dehydratase. The GST domain is also apparently able to give a low-level non-specific stabilization of SDH activity. Similarly, the SK domain is not enzymically active as a monofunctional domain, but is active as part of the complete AROM protein or as a bifunctional protein with 3-dehydroquinate dehydratase.

Inspection of Figure 3 reveals that all of the constructs designed for *in vivo* expression led to the induction of high levels of heterologous protein synthesis; however, further analysis showed that the proteins overproduced were in a substantially insoluble

state. In addition, plasmids specifying the production of SK and SDH (pETA3 and pETA13 respectively) also directed the production of largely insoluble proteins of the predicted molecular mass (results not shown). The enzyme activity of the 3-dehydroquinate dehydratase and SDH domains in cell-free extracts of *E. coli* harbouring plasmids pETA4, pETA5, pETA7 and pETA9, was monitored and, although very low, was found to be detectable over the background *E. coli* activity (results not shown). This low-level activity was correlated with the ability to complement the various *E. coli* *aro*⁻ mutant strains.

All attempts to increase solubility by growing and inducing the cells at 27 °C, or preincubation with 1 % (v/v) ethanol (to induce the *E. coli* heat shock proteins) prior to IPTG induction failed to give significantly elevated levels of soluble protein. Using the method of Coggins et al. [24] (results not shown) we were able to stain for the SDH activity of the trifunctional and bifunctional polypeptides encoded by pETA9 and pETA7 respectively, and also GST-SDH encoded by pFP69 within a polyacrylamide gel, after refolding of the protein by removal of SDS. We therefore attempted to cross-link crude extracts of soluble protein with glutaraldehyde before electrophoresis and were indeed able to detect SDH activity even from the small amount of soluble protein which had been loaded on to the gel. However, although the cross-linked material changed in mobility, no SDH activity was detected in a band corresponding to the size of a dimer. It seems unlikely that the quaternary structure of the engineered fragments can be resolved without their purification.

Overproduction in *E. coli* of domains derived from the QUTR protein

DNA sequences specifying (a) the complete QUTR protein, and deleted forms encoding (b) the N-terminal two domains (SK-like/3-dehydroquinate dehydratase-like), (c) the C-terminal two domains (3-dehydroquinate dehydratase-like/SDH-like; see the Materials and methods section), and (d) each of the repressor's three domains individually, were subcloned into the *E. coli*

Figure 4 SDS/PAGE analysis of *E. coli* strain BL21 (DE3) harbouring recombinant plasmids encoding the two C-terminal domains of the QUTR repressor protein

Each numbered lane contains approximately 15 µg of total protein from boiled extracts of IPTG-induced sonicated BL21 (DE3) cells containing the plasmids pRF20 (lanes 1 and 2), pRF19 (lanes 3 and 4) or unmodified pRSETB (lanes 5 and 6). Plasmid pRF20 encodes the SDH-like domain and pRF19 encodes a bidomain protein comprising the 3-dehydroquinase dehydratase-like and SDH-like domains. Lanes 1, 3 and 5 contain protein comprising the insoluble fraction following sonication, and lanes 2, 4 and 6 comprise the soluble fraction following sonication. Lane M contains molecular mass markers of phosphorylase b (97 kDa), BSA (66 kDa), ovalbumin (45 kDa), carbonic anhydrase (29 kDa) and β lactoglobulin (18.4 kDa), co-electrophoresed to calibrate the gel.

expression vectors pKK233-2, pTrec99a and pRSETB. See Figure 1 for a bar diagram showing the domain structure and order within the QUTR protein.

The recombinant pKK233-2 and pTrec99a plasmids were used in the *in vivo* complementation assay to screen the complete QUTR protein and its various subdomains for enzyme activity. None of the recombinant plasmids was able to complement any of the mutations in the various *aro*⁻ mutant strains, implying that the QUTR protein had lost all of the activity associated with the various shikimate pathway enzymes with which it has been proposed to be homologous. This implication is likely to be sound for the 3-dehydroquinase dehydratase-like domain and for the SDH-like domain, but caution must be exercised when considering proteins specified by plasmids containing the sequence encoding the SK-like domain (see below).

Only recombinant pRSETB plasmids specifying the production of the 3-dehydroquinase dehydratase-like domain alone, the SDH domain alone, or a bi-domain protein specifying the 3-dehydroquinase dehydratase and SDH-like domains were able to direct heterologous protein overproduction in *E. coli*. These observations suggest that there may be an unidentified intron in the N-terminal SK-like encoding region of the *qutR* gene that is causing a shift in the reading frame when these sequences are translated in *E. coli*. Amino acid sequence alignments with the equivalent portion of the equivalent QA-1S protein of *N. crassa* demonstrate that the first 77 amino acids of the QUTR protein have no counterpart within the QA-1S protein, but that similarity downstream of this point is extensive, except for an additional

Figure 5 SDS/PAGE analysis of *E. coli* strain BL21 (DE3) harbouring recombinant plasmids specifying the QUTR 3-dehydroquinase dehydratase-like domain fusion protein

Boiled cells (15 µg) of *E. coli* strain BL21 (DE3) were analysed containing: lane 1, pGEX-2T induced with IPTG; lane 2, pRF12 induced with IPTG; and lane 3, 10 µg of purified MC-QUTR 3-dehydroquinase dehydratase-like fusion protein. Lane M contains the molecular mass markers detailed in the legend to Figure 4.

contiguous sequence of 55 additional amino acids present in the QA-1S protein [1,4,28]. These observations imply that any intron will be contained within the first 231 nt of the *qutR* gene. If this explanation is correct, it would account for the inability to detect any *qutR*-encoded protein that includes the sequences encoding the N-terminal SK-like domain. These observations also imply that there is a compensating error in the sequence reported by Hawkins et al. [4] that corrects the reading frame of the nucleotide sequence reported there. We are currently investigating this possibility.

Further analysis showed that only recombinant pGEX-2T and pRSETB plasmids containing the 3-dehydroquinase dehydratase-like domain alone were able to direct production of a substantially soluble protein (see Figures 4 and 5). All attempts to increase the solubility of the other overproduced QUTR-derived proteins by growing and inducing the cells at 27 °C, or pre-incubation with 1% ethanol (to induce the *E. coli* heat shock proteins) prior to IPTG induction, failed to give significantly elevated levels of soluble protein.

The 3-dehydroquinase dehydratase domains of the AROM protein and the QUTR protein have highly similar structures

The proposed evolutionary origins of the QUTR protein arising from the splitting of a duplicated copy of the *aromA* gene (or its

precursor) was deduced from alignments of inferred amino acid sequences [4,13]. This evolutionary model also proposed that the homologous domains should retain a strongly conserved fold structure. If this model is correct it can be unequivocally tested by crystallographic structure determination, and significant strides have been taken in this direction [32]; however, this is clearly a long-term project. If the fold structure of the two proteins is conserved then both the secondary structure content and the environment of key amino acid residues such as tryptophan should also be conserved. In order to test this more short-term tractable prediction, the secondary structure content and the tertiary structure (as measured by the environment of the tryptophan residues) of the AROM and QUTR 3-dehydroquinase dehydratase domains were determined by far-UV CD spectroscopy and fluorescence emission spectroscopy. These experiments therefore have the potential to refute the biological hypothesis on the common origins of these two proteins, if the results were to demonstrate that the two proteins were biophysically dissimilar.

Purification of the QUTR and AROM 3-dehydroquinase dehydratase domains as fusion proteins

The wild-type QUTR 3-dehydroquinase dehydratase-like domain was purified as a fusion protein with the GST domain specified by plasmid pGEX-2T or the MC domain specified by plasmid pRSETB. The GST fusion protein was purified from 100 ml cultures of plasmids pXF21–23 in *E. coli* strain BL21 (DE3) following sonication and elution from GSH–Sepharose 4B columns as described previously [13]. The MC fusion protein was purified as previously described for the AROM 3-dehydroquinase dehydratase–MC fusion [13] with the following modifications; (a) the column containing the immobilized MC fusion protein was washed with 5 column volumes of buffer 4 [13] containing 2.0 M glycine to remove nucleic acids and some contaminating proteins, and (b) the MC fusion protein was eluted from the column with a step gradient of 10 mM EDTA in buffer 4. The MC fusion protein was further purified to homogeneity by FPLC on a 4 ml NeoBar AQ4 column (Flowgen) using an 80 ml linear 0.0 → 0.5 M NaCl gradient in 20 mM Tris/HCl, pH 7.2, following overnight dialysis in 20 mM Tris/HCl, pH 7.2. Fractions containing the desired protein were pooled and stored at 4 °C. Figure 3 shows a typical induction and the state of purity following chromatography.

The MC–AROM 3-dehydroquinase dehydratase fusion protein was purified from *E. coli* strain BL21(DE3) harbouring plasmid pRF1 as described previously [13], except that an additional step of high-performance FPLC on a NeoBar AQ4 column was performed exactly as described for the MC–QUTR 3-dehydroquinase dehydratase-like domain. The MC–AROM 3-dehydroquinase dehydratase eluted at around 250 mM NaCl and was identified by *in vitro* enzyme assay and SDS/PAGE analysis. In each case, the yield of the purified protein was approximately 8 mg.

Characterization of the purified fusion proteins

(i) Comparative CD spectroscopy

The far-UV CD spectra of the MC–wild-type QUTR 3-dehydroquinase dehydratase-like domain and the MC–AROM 3-dehydroquinase dehydratase domain are shown in Figure 6. They reveal a significant similarity in secondary structure content between the two proteins. The QUTR-derived protein was fully soluble only in 250 mM salt, and storage at 4 °C or room

Figure 6 CD spectra of the QUTR and AROM 3-dehydroquinase dehydratase domains

The MC–AROM 3-dehydroquinase dehydratase fusion protein (solid line) was measured at a protein concentration of 0.43 mg/ml in a 0.2-mm-pathlength cell and 50 scans were accumulated between 185 and 250 nm. The MC–QUTR 3-dehydroquinase dehydratase-like fusion protein (broken line) was measured at a protein concentration of 0.61 mg/ml in a 0.1-mm-pathlength cell. The buffer was 62.5 mM NaCl, 187.5 mM NaF, 100 mM phosphate buffer, pH 7.2, and 99 scans were accumulated between 190 and 250 nm. Each spectrum was corrected by subtraction of a comparable blank. The abscissa is represented in units of $\Delta\epsilon$ ($M^{-1} \cdot cm^{-1}$) using the molar concentration of amino acid residues present.

temperature led to increasing protein aggregation and a gradual degradation of the CD signal. Consequently the recording of CD spectra was begun within 30 min of the pure protein being eluted from the NeoBar column; however, the presence of the salt limited acquisition of data to above 190 nm. Near-UV spectra, which can give indications of tertiary structure [33,34], were not recorded since the aromatic side chains which give rise to this signal are poorly conserved between the two domains. CONTIN analysis [35] gives estimates of secondary structure contents (α -helix/ β -sheet/other structures) as 30/10/60 % for QUTR and 32/22/46 % for AROM. Unresolved regions at the termini of these artificial domains, together with the histidine-rich tag, may account for the high content of disordered structure estimated from this analysis. The resolution between the weak β -sheet component and disordered structure is often a problem in CD analysis [35] and the difference in this analysis is not excessive. The consistency of the stronger α -helix signal is therefore the best statistical basis for the obvious visual similarity of the two spectra.

(ii) Fluorescence emission spectroscopy

In order to compare the tertiary structures, purified AROM 3-dehydroquinase dehydratase domain and the QUTR 3-dehydroquinase dehydratase-like domain were subject to fluorescence emission spectroscopy in the presence and absence of the denaturant guanidine hydrochloride. The four spectra obtained are shown in Figure 7 (upper and lower panels), inspection of which highlights the following points. (1) The native spectra are highly unusual, as they are dominated by the signal from tyrosine rather than the expected tryptophan signal. The contribution from the tryptophan residues (three for AROM and one for QUTR) is highly quenched and only measurable at the highest machine sensitivity. (2) When the secondary structure of the two domains is disrupted by the denaturant guanidine hydrochloride, the expected signal from the tryptophan residues appears and is in proportion to the tryptophan content predicted by the nucleotide sequence. Excitation of either protein at 295 nm, which only excites tryptophan residues, showed a low-yielding tryptophan fluorescence which increased on the addition of the denaturant guanidine hydrochloride (results not shown).

Figure 7 Fluorescence emission spectra of native and denatured QUTR and AROM 3-dehydroquinatase domains

Samples were dissolved in 100 mM sodium phosphate, 500 mM NaCl, pH 7, or (for denatured spectra) the same buffer including 6 M guanidine hydrochloride (Sigma). Spectra shown have been corrected for baseline effects by subtraction of a buffer blank and for instrument response using standards supplied by SLM. Upper panel: 70 μ g/ml of the AROM 3-dehydroquinatase domain; lower panel: 66 μ g/ml of the QUTR 3-dehydroquinatase-like domain.

These results demonstrate that the tryptophan residues in the two proteins are in very similar, but highly unusual, environments as a consequence of a highly conserved tertiary structure. The best-known example of such quenching is the case of subtilisin carlsberg [36,37] in which a single tryptophan is statically quenched, allowing dominance of the spectrum by tyrosine. AROM would appear particularly unusual in that all three tryptophan residues are quenched, whereas in subtilisin BPN' the additional tryptophan residues are fluorescent [36,37]. Taken together with the comparative CD studies, these results demonstrate that the structures of the two domains are highly similar and provides independent supportive biophysical evidence consistent with the proposed common origins of the AROM and QUTR proteins.

(iii) Native molecular mass determinations

The solution subunit structure of the MC-QUTR 3-dehydroquinatase dehydratase-like domain and the AROM 3-dehydroquinatase dehydratase domain were probed by FPLC as described in the Materials and methods section. The QUTR 3-dehydroquinatase dehydratase-like domain was chromatographed four times and in each case was resolved into two peaks eluting at 7.70 ml (the void volume) and 9.65 ml. This corresponds to highly aggregated protein and protein corresponding to a molecular mass of 68.3 kDa respectively. The predicted molecular mass (from the DNA sequence) of the monomeric MC-QUTR 3-dehydroquinatase dehydratase-like domain is 35.5 kDa, strongly suggesting that the native solution conformation of this domain is a dimer. When the peak corresponding to the dimer was collected and the protein rechromatographed, it eluted in an identical position, indicating that the dimeric form is stable and not an aggregation artefact. In a control experiment, the MC-AROM 3-dehydroquinatase dehydratase domain specified by plasmid pRF1 eluted at a volume of 10.2 ml, which corresponds to a molecular mass of 39.6 kDa. The predicted molecular mass of this protein (coordinates 3098–3870 in the *aromA* sequence) is 32.7 kDa, supporting the previously published observation that this domain is monomeric in solution [13].

(iv) Enzyme activity

Although it was not expected that the wild-type or mutant forms of the QUTR 3-dehydroquinatase dehydratase-like domain would have any enzyme activity, it nevertheless had to be determined if in fact this was the case. Approximately 1 μ g of each purified fusion protein (including the solubilized MC-mutant fusion) was assayed for 3-dehydroquinatase activity; however, no conversion of substrate by either QUTR domain could be detected. The plasmid pGEX-2T contains the *trc* promoter, therefore plasmids pXF21–27 were transformed into *E. coli* strain SK3430 in order to carry out the sensitive *in vivo* complementation assay. *E. coli* strain SK3430, harbouring plasmids pXF21–27, were unable to grow on minimal medium lacking aromatic amino acid supplementation, showing that the GST-wild-type fusion protein could not supply 3-dehydroquinatase dehydratase activity and complement the *aroD*[−] mutation. This test could not be repeated with MC fusion protein as plasmid pRSETB utilizes the T7 promoter and *E. coli* strain SK3430 does not produce the T7 polymerase. In all cases *E. coli* strain SK3430 harbouring plasmid pRF1, which encodes the AROM 3-dehydroquinatase dehydratase domain (positive), or plasmid pGEX-2T (negative), were used as controls [13].

(v) Amino acid sequence determination

As there is no *in vitro* assay for the QUTR 3-dehydroquinatase dehydratase-like domain, the identity of the proposed MC-QUTR 3-dehydroquinatase dehydratase-like domain was verified by amino acid sequence analysis as described in the Materials and methods section. The sequence of the first eight amino acids was determined as Met, Arg, Gly, Ser, His, His, His, His, which matches exactly with the N-terminal sequence of the pRSETB-encoded MC domain.

Conclusions

The combined data discussed above demonstrate that the two C-terminal domains of the QUTR transcription-repressing protein have similar physical characteristics to the equivalent domains of the AROM protein. The pattern of solubility of the

Table 2 Designations of recombinant *E. coli* expression vectors specifying QUTR domains or the AROM 3-dehydroquinase dehydratase domain

The AROM 3-dehydroquinase dehydratase domain was from nucleotide co-ordinates 3098 to 3870. The QUTR dehydroquinase dehydratase-like domain was from nucleotide co-ordinates 936 to 1716. The QUTR dehydroquinase dehydratase/SDH-like domains were from nucleotide co-ordinates 936 to 2790. The QUTR SDH-like domain was from nucleotide co-ordinates 1723 to 2790. The template for all PCR reactions involving *qutR* DNA sequences was plasmid pQ5 [38]. The nucleotide sequence of the *qutR* gene corresponding to the complete QUTR protein and all possible combinations of single and double domains were subcloned into the commercially available vectors pKK233-2, pTrc 99a, pRSETB, pGEX-2T and pUC18. For subcloning into pKK233-2 and pTrc99a the 5' oligonucleotide contained either an *NcoI* or *BspHI* site and the 3' oligonucleotide contained a *HindIII* site. For subcloning into pGEX-2T the 5' and 3' oligonucleotides contained an *EcoRI* site and for subcloning into pRSETB the 5' and 3' oligonucleotides contained *NcoI* sites. The arginine residue at position 501 in the QUTR protein is the equivalent residue to the active-site lysine residue conserved in all type-I 3-dehydroquinase dehydratase enzymes [4]. Arg-501 was altered to lysine by PCR-mediated site-directed mutagenesis using *qutR* DNA from nt 936 to 1764 cloned into the plasmid pUC18 (see the Materials and methods section). Subsequently the *qutR* DNA from three independently isolated recombinant plasmids was purified and subcloned into M13 sequencing vectors and plasmids pGEX-2T and pRSETB. The desired sequence changes were directly verified by DNA sequence analysis and recombinant pGEX-2T plasmids (designated pXF11 and pXF13) and pRSETB plasmids (designated pRF3-5) overproducing appropriate mutant fusion proteins were isolated and identified as described in the Materials and methods section. The purpose behind this site-directed mutagenesis was to attempt to produce a QUTR 3-dehydroquinase dehydratase-like domain that could be used to demonstrate the predicted binding of DHQ. The homologous domain in the AROM protein is a type-I 3-dehydroquinase dehydratase whose enzyme reaction mechanism proceeds by Schiff-base formation using a conserved lysine residue [39–43]. Treatment with borohydride traps the substrate DHQ into the catalytically active centre by covalent attachment to the conserved lysine residue [40]. The equivalent residue in the QUTR 3-dehydroquinase-like domain is Arg-501, therefore by changing Arg-501 → Lys we hoped to be able to demonstrate an equivalent covalent attachment of DHQ using this mutant domain. Abbreviations: GST, glutathione S-transferase affinity purification domain; MC, the six-histidine-containing metal-chelating domain facilitating purification by immobilized metal affinity chromatography; NA, not applicable.

Plasmid	Vector	Fusion domain	Promoter	Source of recombinant DNA	Strain for overexpression	Strain for <i>in vivo</i> complementation	Protein produced
pXF21–27	pGEX-2T	GST	<i>trc</i>	<i>qutR</i>	BL21 (DE3)	SK3430	DHQase
pRF9–12	pRSETB	MC	T7	<i>qutR</i>	BL21 (DE3)	NA	DHQase
pXF11 + pXF13	pGEX-2T	GST	<i>trc</i>	<i>qutR</i>	BL21 (DE3)	SK3430	Mutant (R → K) DHQase
pRF1	pRSETB	MC	T7	<i>aroma</i>	BL21 (DE3)	NA	DHQase
pRF3–5	pRSETB	MC	T7	<i>qutR</i>	BL21 (DE3)	NA	Mutant (R → K) DHQase
pRF19	pRSETB	MC	T7	<i>qutR</i>	BL21 DE3	NA	DHQase/SDH
pRF20	pRSETB	M C	T7	<i>qutR</i>	BL21 DE3	NA	SDH

individual and bi-domain fragments comprising the two proteins is the same, and the chromatographic behaviour of the 3-dehydroquinase dehydratase domains on ion-exchange chromatography is identical. The use of biophysical techniques such as CD spectroscopy and fluorescence emission spectroscopy to probe the secondary and tertiary structures of the 3-dehydroquinase dehydratase domains shows them to be virtually indistinguishable.

There are, however, notable differences between the two proteins, in that the AROM 3-dehydroquinase dehydratase domain is monomeric whereas the equivalent domain in QUTR is dimeric. This probably reflects the different evolutionary pressures acting on these two domains. Taken as a whole, however, the results presented here (which could have refuted the hypothesis), are consistent with the hypothesis that the AROM protein and the QUTR repressor protein have a common ancestor [2,4,9,10].

Attempts to demonstrate DHQ binding to the QUTR 3-dehydroquinase dehydratase-like domain using a mutant form of the protein which has the putative arginine residue binding site at position 501 changed to lysine (see the legend to Table 2) were frustrated as the mutant protein proved to be insoluble. Although the protein could be solubilized and purified using chaotropic solubilizing agents, the final purified protein, although soluble at low concentrations, was unable to fold properly, as judged by CD spectroscopy. Plasmids pXF11 and pXF13 specifying a GST–mutant QUTR fusion protein were also tested in the sensitive *in vivo* complementation assay, but no complementation was observed.

The use of the *in vivo* complementation test suggests that the QUTR 3-dehydroquinase dehydratase-like and SDH-like domains have no enzyme activity. However, the possibility of the presence of an as yet unidentified intron in the sequence which encodes the SK-like domain means that no conclusions can be drawn as to any possible enzyme activity associated with this

domain. The AROM SK and SDH domains are not active as monofunctional domains, but are active as bifunctional proteins in combination with the 3-dehydroquinase dehydratase domain. In the case of the SDH domain this interaction/stabilization may not be specific as substitution of the GST protein for the 3-dehydroquinase dehydratase domain does partially restore SDH function. The EPSP synthase domain has similarly been shown to be dependent on covalent attachment to the DHQS domain, but in this case substitution of the GST protein for DHQS does not restore EPSP synthase activity [11].

This work was supported by U.K. Research Council funding.

REFERENCES

- Giles, N. H., Case, M. E., Baum, J., Geevers, R., Huiet, L., Patel, V. and Tyler, B. (1985) *Microbiol. Rev.* **49**, 338–358
- Hawkins, A. R., Lamb, H. K., Moore, J. D., Charles, I. G. and Roberts, C. F. (1993) *J. Gen. Microbiol.* **139**, 2891–2899
- Beri, R. K., Whittington, H., Roberts, C. F. and Hawkins, A. R. (1987) *Nucleic Acids Res.* **15**, 7991–8001
- Hawkins, A. R., Lamb, H. K. and Roberts, C. F. (1992) *Gene* **110**, 109–114
- Charles, I. G., Keyte, J. W., Brammar, W. J., Smith, M. and Hawkins, A. R. (1986) *Nucleic Acids Res.* **14**, 2201–2213
- Hawkins, A. R. (1987) *Curr. Genet.* **11**, 491–498
- Duncan, K., Edwards, R. M. and Coggins, J. R. (1987) *Biochem. J.* **246**, 375–386
- Pittard, J. (1987) in *Escherichia coli and Salmonella typhimurium*; Cellular and Molecular Biology, (Neidhard, F. C., Ingraham, J., Low, K. B., Magasanik, B., Schaechter, M. and Umberger, H. E., eds.), pp. 368–394. Am. Soc. Microbiol., Washington D. C.
- Anton, I. A., Duncan, K. and Coggins, J. R. (1987) *J. Mol. Biol.* **197**, 367–371
- Hawkins, A. R., Lamb, H. K., Moore, J. D. and Roberts, C. F. (1993) *Gene* **136**, 49–54
- Moore, J. D. and Hawkins, A. R. (1993) *Mol. Gen. Genet.* **240**, 92–102
- Moore, J. D., Coggins, J. R., Virden, R. and Hawkins, A. R. (1994) *Biochem. J.* **301**, 297–304
- Hawkins, A. R., Moore, J. D. and Adekun, A. M. (1993) *Biochem. J.* **296**, 451–457

- 14 Kinghorn, J. R. and Hawkins, A. R. (1982) *Mol. Gen. Genet.* **186**, 145–152
- 15 Blattner, F. R., Williams, B. G., Blechl, A. E., Denniston-Thomson, K., Faber, H. E., Furlong, L., Grunwald, D. J., Keifer, D. O., Moore, D. D., Schumm, J. W., Sheldon, E. L. and Smithies, O. (1977) *Science* **196**, 161–169
- 16 Willetts, N. S., Clark, A. J. and Low, B. (1969) *J. Bacteriol.* **97**, 244–249
- 17 Hawkins, A. R., Da Silva, A. J. S. and Roberts, C. F. (1984) *J. Gen. Microbiol.* **130**, 567–574
- 18 Hawkins, A. R. and Smith, M. (1991) *Eur. J. Biochem.* **196**, 717–724
- 19 Lbner-Olesen, A. and Marinus, M. G. (1992) *J. Bacteriol.* **174**, 525–529
- 20 van den Hombergh, J. P. T. W., Moore, J. D., Charles, I. G. and Hawkins, A. R. (1992) *Biochem. J.* **284**, 861–867
- 21 Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
- 22 Tautz, D. and Renz, M. (1982) *Anal. Biochem.* **132**, 14–19
- 23 Laemmli, U.K. (1970) *Nature (London)* **227**, 680–685
- 24 Coggins, J. R., Boocock, M. R., Chaudhuri, S., Lambert, J. M., Lumsden, J., Nimmo, G. A. and Smith, D. D. S. (1987) *Methods Enzymol.* **142**, 325–341
- 25 Hemsley, A., Arnheim, N., Toney, M. D., Cortopassi, G. and Galas, D. J. (1989) *Nucleic Acids Res.* **17**, 6545–6551
- 26 Gill, S. C. and von Hippel, P. H. (1989) *Anal. Biochem.* **182**, 319–326
- 27 Geever, R. F., Huiet, L., Baum, J. A., Tyler, B. M., Patel, V. B., Rutledge, B. J., Case, M. E. and Giles, N. H. (1989) *J. Mol. Biol.* **207**, 15–34
- 28 Huiet, L. and Giles, N. H. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3381–3385
- 29 Hawkins, A. R., Lamb, H. K., Smith, M., Keyte, J. W. and Roberts, C. F. (1988) *Mol. Gen. Genet.* **214**, 224–231
- 30 Anton, I. A. and Coggins, J. R. (1988) *Biochem. J.* **249**, 319–326
- 31 Lamb, H. K., Bagshaw, C. R. and Hawkins, A. R. (1991) *Mol. Gen. Genet.* **227**, 187–196
- 32 Boys, C. W. G., Bury, S. M., Sawyer, L., Moore, J. D., Charles, I. G., Hawkins, A. R., Deka, R., Kleanthous, C. and Coggins, J. R. (1992) *J. Mol. Biol.* **227**, 352–355
- 33 Lakey, J. H., Masotte, D., Heitz, F., Dasseux, J. L., Faucon, J. F., Parker, M. W. and Pattus, F. (1991) *Eur. J. Biochem.* **196**, 599–607
- 34 van der Goot, F. G., González-Mañas, J. M., Lakey, J. H. and Pattus, F. (1991) *Nature (London)* **354**, 408–410
- 35 Provencher, S. W. and Glockner, J. (1981) *Biochemistry* **20**, 33–37
- 36 Longworth, J. W. (1971) in *Excited States of Proteins and Nucleic Acids* (Weinryb, I. and Steiner, R. F., eds.), pp. 312–494, Plenum Press, New York
- 37 Janot, J. M., Beeby, A., Bayley, P. M. and Phillips, D. (1991) *Biophys. Chem.* **41**, 277
- 38 Lamb, H. K., Hawkins, A. R., Smith, M., Harvey, I. J., Brown, J., Turner, G. and Roberts, C. F. (1990) *Mol. Gen. Genet.* **223**, 17–23
- 39 Chaudhuri, S., Duncan, D., Graham, L. D. and Coggins, J. R. (1991) *Biochem. J.* **275**, 1–6
- 40 Chaudhuri, S., Lambert, J. M., McColl, L. A. and Coggins, J. R. (1986) *Biochem. J.* **239**, 699–704
- 41 Kleanthous, C., Deka, R., Davis, K., Kelly, S. M., Cooper, A., Harding, S. E., Price, C., Hawkins, A. R. and Coggins, J. R. (1992) *Biochem. J.* **282**, 687–695
- 42 Harris, J., Kleanthous, C., Coggins, J. R., Hawkins, A. R. and Abell, C. O. (1993) *J. Chem. Soc. Chem. Commun.* **13**, 1080–1081
- 43 Shneider, A., Harris, J., Kleanthous, C., Coggins, J. R., Hawkins, A. R. and Abell, C. (1993) *Bio-organic Med. Chem. Lett.* **3**, 1399–1402