Structure, Vol. 10, 493–503, April, 2002, ©2002 Elsevier Science Ltd. All rights reserved. PII S0969-2126(02)00747-5

# The Structure and Mechanism of the Type II Dehydroquinase from *Streptomyces coelicolor*

**Aleksander W. Roszak,[1] David A. Robinson,[2] Tino Krell,[2] Iain S. Hunter,[3] Martyn Fredrickson,[4] Chris Abell,[4] John R. Coggins,[2] and Adrian J. Lapthorn[1,5]**
[1]Department of Chemistry and
[2]Division of Biochemistry and Molecular Biology
Institute of Biomedical and Life Sciences
University of Glasgow
Glasgow G12 8QQ
Scotland
United Kingdom
[3]Department of Bioscience and Biotechnology
University of Strathclyde
Glasgow G1 1XW
Scotland
United Kingdom
[4]University Chemical Laboratory
University of Cambridge
Lensfield Road
Cambridge CB2 1EW
United Kingdom

## Summary

**The structure of the type II DHQase from *Streptomyces coelicolor* has been solved and refined to high resolution in complexes with a number of ligands, including dehydroshikimate and a rationally designed transition state analogue, 2,3-anhydro-quinic acid. These structures define the active site of the enzyme and the role of key amino acid residues and provide snap shots of the catalytic cycle. The resolution of the flexible lid domain (residues 21–31) shows that the invariant residues Arg23 and Tyr28 close over the active site cleft. The tyrosine acts as the base in the initial proton abstraction, and evidence is provided that the reaction proceeds via an enol intermediate. The active site of the structure of DHQase in complex with the transition state analog also includes molecules of tartrate and glycerol, which provide a basis for further inhibitor design.**

## Introduction

The shikimate pathway is a central biosynthetic route using erythrose 4-phosphate and phosphoenol pyruvate to produce chorismate, the precursor for the synthesis of aromatic amino acids, folic acid, ubiquinone, and many other aromatic compounds. This pathway is found in bacteria, higher plants, and fungi [1, 2] and has been recently discovered in apicomplexan parasites [3]. Its absence from mammals has made this pathway an attractive target for the development of herbicides [4] and antimicrobial agents [5]. The enzyme 3-dehydroquinate dehydratase or dehydroquinase (DHQase; EC 4.2.1.10) catalyzes the third step of the shikimate pathway, dehydration of 3-dehydroquinate to 3-dehydroshikimate (Figure 1). This step is common to the catabolic quinate pathway, which converts quinic acid to p-hydroxybenzoic acid that can be further metabolised via the β-ketoadipate pathway to acetyl-CoA and succinyl-CoA [6, 7].

Two classes of DHQases have been described: the type I enzyme, which is associated exclusively with chorismate biosynthesis in fungi, plants, and some bacteria, and the type II enzyme, which is found in the quinate pathway of fungi and in the shikimate pathway of many bacteria. The two classes of enzymes operate by entirely different mechanisms, as exemplified by the contrasting stereochemistry of the reactions catalyzed [8]. Type I enzymes catalyze a *syn* elimination of water with the loss of the *pro-R* hydrogen from C2 [9, 10], and the mechanism involves an imine intermediate [11, 12]. Type II enzymes catalyze an *anti* elimination reaction with the loss of the more acidic axial *pro-S* hydrogen from C2, which is thought to proceed via an enolate intermediate [13, 14]. Recently, representative crystal structures of both types of enzyme have been determined [15], confirming that the enzymes are structurally unrelated and possess different folds. The *Salmonella typhi* type I enzyme has an α/β barrel structure and is dimeric, while the *Mycobacterium tuberculosis* type II enzyme has a flavodoxin-like fold and forms a dodecamer with tetrahedral symmetry [15].

In this paper we report the amino acid sequence, purification, crystallization, and the determination of a number of structures of the type II DHQase from *Streptomyces coelicolor*. These structures permit a detailed description of the active site, which includes the previously unobserved loop (residues 22–29) that closes over the active site pocket and allow us to assign functions to individual amino acid residues in the proposed mechanism. In addition, the structure of the enzyme in complex with a transition state analog includes a molecule of tartrate and glycerol in the active site, providing a structural basis for the development of inhibitors by rational design.

## Results and Discussion

### The Enzyme Fold

The type II DHQase from *S. coelicolor* is assembled from 12 identical monomers of 156 amino acid residues (molecular mass of 16,541 Da) to form a dodecamer. The monomer (Figure 2) is an α/β protein with a central five-stranded parallel β sheet with a strand order of 21345, which is a flavodoxin-like fold, as defined in the SCOP classification of protein folds [16]. There are four main α helices in the structure, with α1 and α4 on one face of the β sheet and α2 and the α3 on the other. The α3 helix is interrupted at Ser87, which leads to a change in its direction. The two parts of this helix are referred

[5]Correspondence: adrian@chem.gla.ac.uk

**Key words:** shikimate pathway; enzyme mechanism; transition state analog; dehydroquinase; X-ray structure; rational drug design

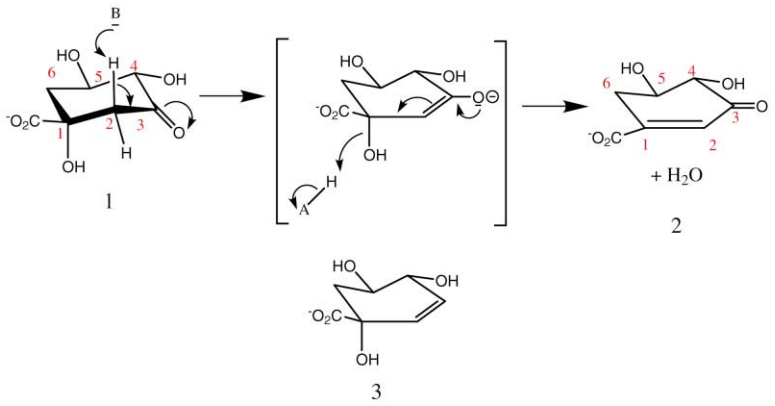


Figure 1. Dehydration of 3-Dehydroquinate (1) to Form 3-Dehydroshikimate (2) and Structure of the Transition State Analog 2,3-Anhydro-Quinic Acid (3)

to as α3′ and α3″. In addition to the α helices, there are four short stretches of $3_{10}$ helix, labeled H1–H4. H1 is located between β strand β1 and helix α1, and the other three segments of $3_{10}$ helix are found in a 23-amino acid sequence between β strands β4 and β5.

The primary structures of type II DHQase are strongly conserved, with 17% of residues invariant and a further 47% conserved over the 19 sequences shown in Figure 3. The enzyme shares no significant sequence similarity with any other group of proteins, and there is not sufficient structural similarity with other flavodoxin-like proteins to suggest common ancestry. It has been noted that the 20 C-terminal amino acids of type II DHQase contain three conserved glycines with the sequence Gly-X-Gly-X-X-Gly and a number of hydrophobic residues similar to the NAD binding motif [17]. From the crystal structure, it can be seen that this region of the *S. coelicolor* type II DHQase adopts a structure similar to the first part of an NAD binding motif, although this does not extend to the remainder of this region. NAD binding domains contain this Gly-X-Gly-X-X-Gly sequence at their N terminus (20–30 residues), but this often follows a substrate binding domain, as in alcohol dehydrogenase. It is tempting to speculate that the type II DHQase has arisen from the truncation of such a dehydrogenase, with the fragment of the NAD binding motif acquiring an alternative function.

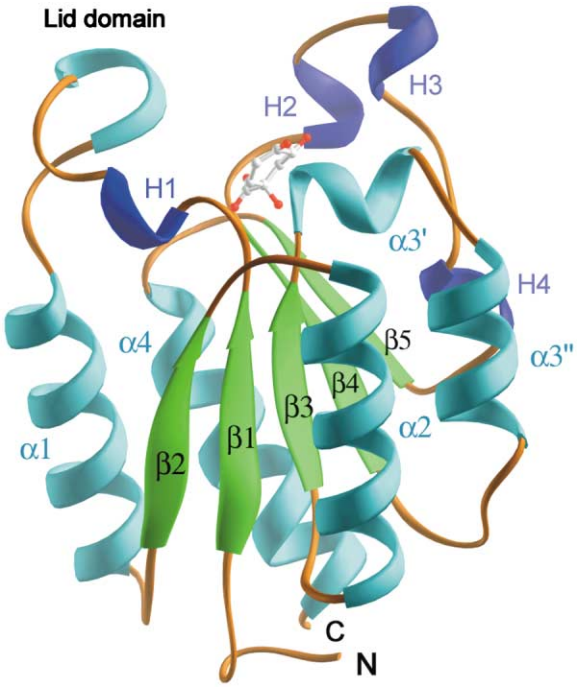


Figure 2. A Ribbon Representation of the Monomer of *S. coelicolor* DHQase

α helices (cyan) are labeled α1–α4, $3_{10}$ helices (purple) are labeled H1–H4, and β strands (green) are labeled β1–β5. The lid domain that is disordered in the *M. tuberculosis* structure [15] is also labeled. The ligand 2,3-anhydro-quinic acid is shown in ball and stick representation to highlight the position of the active site. The diagram was produced using the program RIBBONS [42].

## The Quaternary Structure

The *S. coelicolor* type II DHQase, like the *M. tuberculosis* enzyme [15], is dodecameric. On forming the dodecamer, each monomer buries 2041 Å$^2$ of solvent-accessible surface, which corresponds to 30% of the total monomer surface area. The dodecamer has tetrahedral 23-point group symmetry and is roughly spherical in shape, with a diameter of 100Å (Figure 4). This is the simplest form of cubic symmetry but has only previously been seen in the structures of ornithine carbamoyltransferase from *Pseudomonas aeruginosa* [18] and *Pyrococcus furiosus* [19]. In the dodecamer there are essentially two interfaces: a trimer interface, formed by two dissimilar surfaces lying parallel to the plane of the β sheet, and a symmetric dimer interface, formed by the continuation of the β sheets through two monomers (Figure 4B). Studies of the unfolding and refolding of type II DHQase [20] have shown that the trimer is the minimum catalytic unit of the enzyme. The dodecamer can thus be considered as a tetramer of trimers.

In forming the trimer, each monomer buries 1320 Å$^2$, or 20%, of its solvent-accessible area at two surfaces. If we consider a single subunit within the trimer, subunit A, this buries 673 Å$^2$ (surface 1), predominantly from helices α2 and α3, at the interface with subunit B. The loops and helices at the C-terminal end of the parallel β sheet of subunit A also bury 647 Å$^2$ of solvent-accessible area (surface 2) at the interface with subunit C (Figure 5B). There are seven hydrogen bonds and one salt bridge formed at each interface, and an equal distribu-

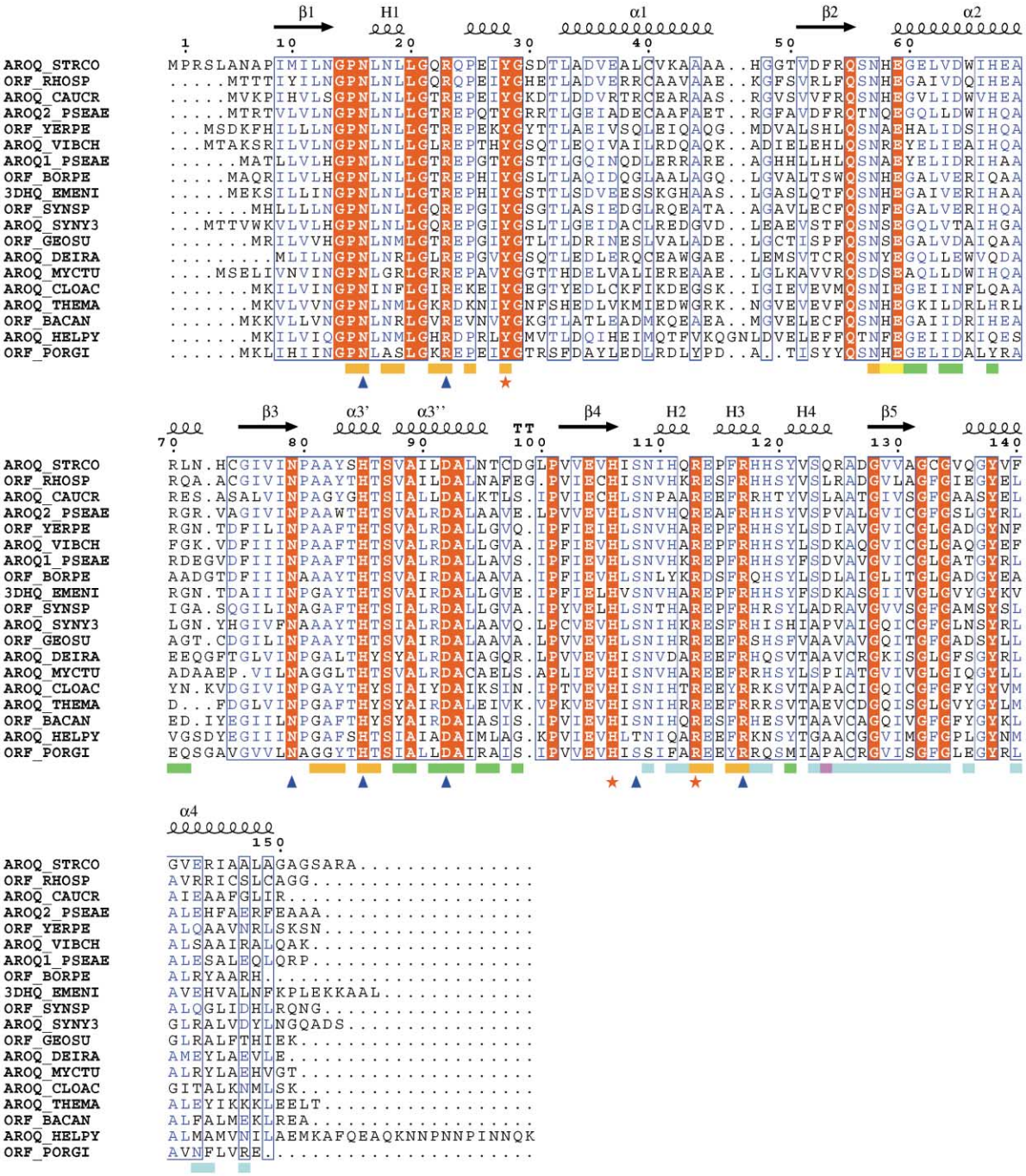


Figure 3. A Sequence Alignment of Type II DHQases from Representative Prokaryotic and Eukaryotic Organisms

α helices and β strands are represented as helices and arrows, respectively, and β turns are marked TT. The first row under the aligned sequences highlights those amino acid residues buried at the dimer interface (cyan) and the trimer interface, surface 1 (green) and surface 2 (orange). Some amino acid residues are buried at both the dimer and trimer interfaces (magenta) or at both trimer interfaces (yellow). Key residues in the active site are marked with blue triangles; those responsible for catalysis are marked with red asterisks. The sequence alignment was created using ESPript [43] and the following sequences (organism, Genebank/Swissprot accession in parenthesis): AROQ_STRCO (*S. coelicolor*, P15474), AROQ_CAUCR (*Caulobacter crescentus*, AAK23857), AROQ_SYNY3 (*Synechocystis sp*., P73367), 3DHQ_EMENI (*Emericella nidulans*, P05147), AROQ1_PSEAE (*Pseudomonas aeruginosa*, AF010322), AROQ2_PSEAE (*P. aeruginosa*, AAG03634), AROQ_VIBCH (*Vibrio cholerae*, AAF93471), AROQ_MYTUB (*M. tuberculosis*, P36918), AROQ_CLOAC (*Clostridium acetobutylicum*, AAK78875), AROQ_HELPY (*Helicobacter pylori*, Q48255), AROQ_THEMA (*Thermotoga maritima*, Q9WYI4), and AROQ_DEIRA (*Deinococcus radiodurans*, P54517). Open reading frames (ORF) from the incomplete genome sequences were obtained from The Institute for Genomic Research website at http://www.tigr.org for PORGI (*Porphyromonas gingivalis*), GEOSU (*Geobacter sulfurreducens*), and BACAN (*Bacillus anthracis*) and from the Sanger Centre Sequencing Project website at http://www.sanger.ac.uk/Projects/ for BORPE (*Bordetella pertussis*) and YERPE (*Yersinia pestis*), and preliminary sequence data was obtained from The DOE Joint Genome Institute (JGI) at http://www.jgi.doe.gov/JGI_microbial/html/index.html for RHOSP (*Rhodobacter sphaeroides*) and SYNSP (*Synechococcus sp*.).

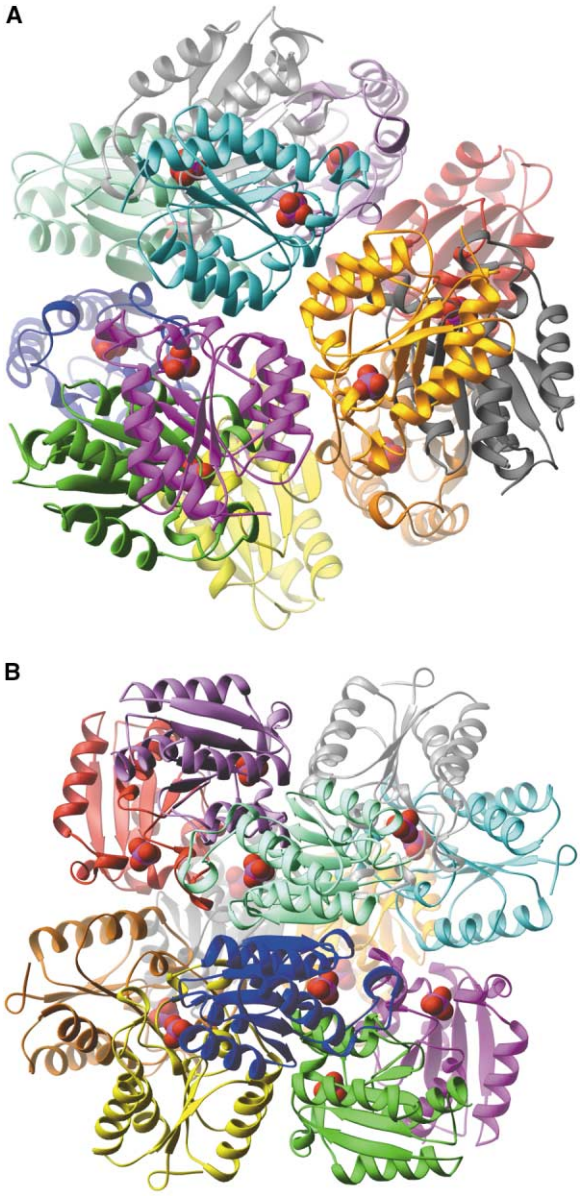


Figure 4. Two Views of the Dodecamer of DHQase in a Ribbon Representation

Each chain is colored uniquely and phosphate ions are in CPK representation to show the position of the active site.

(A) A view down the NCS 3-fold axis with the catalytic trimer in front clearly visible (chain A, cyan; chain B, gold; chain C, magenta). The other side of the NCS 3-fold axis is a large open cavity with no trimer interactions.

(B) The dodecamer viewed down the NCS 2-fold axis between chains E (light green) and L (blue). The other two NCS 2-folds are perpendicular and can clearly be seen to run diagonally through the dodecamer in this view.

tion of polar and nonpolar atoms at both of these two surfaces.

At the dimer interface, each monomer buries 795 Å$^2$ of solvent-accessible surface to form an interface that is flat, with a length of 30 Å and a breadth of 25 Å, and composed of approximately equal numbers of polar and nonpolar atoms [21]. The interface brings together equivalent regions of each subunit, which are separated by the central β sheet (Figure 4B). On one side of the interface, residues in the final turn and terminal helix (α4) of the two subunits pack against each other and against Gly128 and Val130 of strand β5 to bury approximately 50% of a predominantly hydrophobic surface (Figure 5A). The residues on the other side of the β sheet form a much more hydrophilic surface, which includes a significant solvent-accessible cavity formed between the three $3_{10}$ helices H2–H4. Histidine 111 makes the largest single contribution (83 Å$^2$) to the buried surface area at the dimer interface. The two imidazole rings of the dimer-related histidines point toward each other, effectively restricting the size of the solvent cavity and leaving a doughnut-shaped space, which is occupied by water.

## The Active Site

The active site in type II DHQase is located in the cleft formed at the carboxy edge of the β sheet between strands β1 and β3, which is common not only to enzymes with a flavodoxin-like fold but also to α/β structures in general [22]. This cleft is partially closed off by interaction with helix α3 of the neighboring subunit at the trimer interface (residues from this subunit will be marked with an asterisk), specifically by the formation of a salt bridge between Arg117 and Asp92*. A flexible loop (residues 21–31) forms a lid, which closes over the active site on binding substrate; this includes the residues Arg23 and Tyr28, previously identified as being catalytically important [23].

The initial crystallizations of *S. coelicolor* DHQase were carried out in the presence of 200 mM sodium phosphate, which has been shown to be a competitive inhibitor of the type II enzyme from *A. nidulans* [24]. This structure revealed a phosphate ion bound in the active site forming H bonds with ND1 of His106, ND2 of Asn79, OH of Tyr28, OG of Ser108, and the backbone nitrogens of Ile107 and Ser108; the remainder of the active site cavity is occupied by nine water molecules (Figure 6B). The position of the phosphate ion is surprising, as it was assumed that a negative ion would mimic binding of the carboxylate group of dehydroquinate and interact with an arginine (either Arg23 or Arg113), as has been seen in the type I DHQase [15]. Subsequent structures of both the wild-type enzyme in complex with a transition state analog, 2,3-anhydro-quinic acid [25], and of an inactive DHQase mutant R23A [23] complexed with product (DHS, dehydroshikimate) show conclusively that the phosphate binds in the carboxylate binding site (Figure 6). The carboxylate binding loop is formed from the main chain amide nitrogens of residues 107–108 and the side chain of OH of Ser108 and ND2 of Asn79 (Figure 6C) and forms a significant part of the substrate recognition.

Asn79 is invariant and is held in place in the active site by hydrogen bonds to its side chain from the amide N of Ala81 to OD1 and from the OH hydroxyl of Tyr138 to ND2. In addition to its role in binding the carboxylate, the side chain OD1 is positioned to act as the hydrogen bond acceptor for the C1 hydroxyl of the substrate, as seen in 2,3-anhydro-quinic acid structure (Figure 6D). Histidine 106 is therefore the hydrogen bond donor to the C1 hydroxyl, and this is consistent with it coordi-

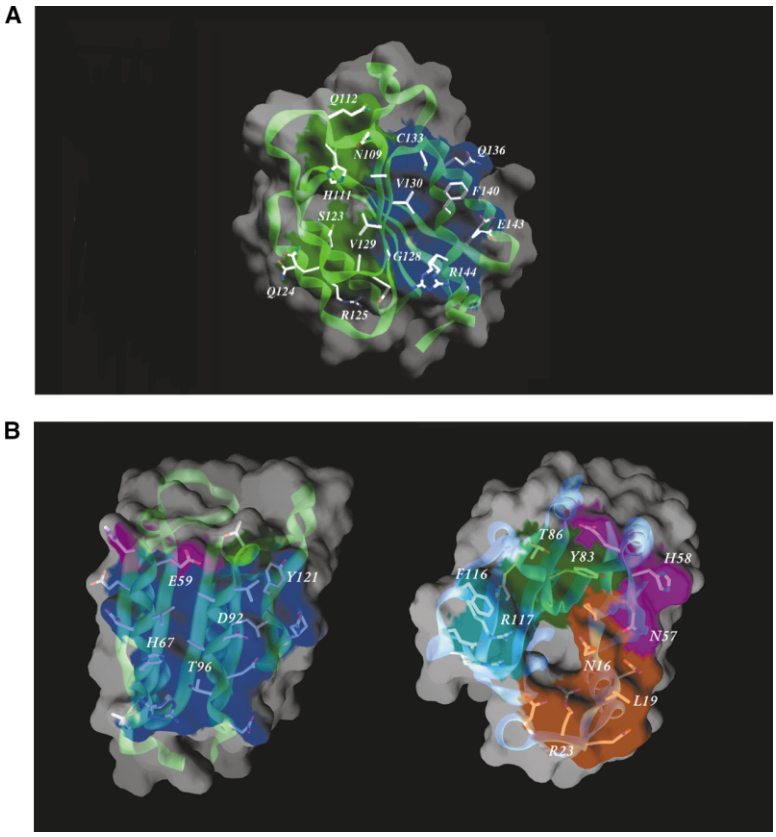


Figure 5. A Surface and Ribbons Representation of the Dimer and Trimer Interfaces

(A) The surface buried of chain A at the dimer interface, colored according to the description in the text, illustrating the hydrophobic surface on one side (dark blue) and the hydrophilic surface on the other (green).

(B) Chains A and B and associated surfaces have been rotated to expose the two different surfaces (surface 1 and 2, respectively) buried at the trimer. A ribbon representation of chain A (green) and chain B (blue), with residues (colored by atom type) contributing to the buried surface area shown in stick; key residues are labeled for clarity. Surface 1 is colored according to residue ranges: residues 15–24, orange; residues 57–59, magenta; residues 82–86, green; residues 113–117, cyan). Surface 2 is blue, except for residues that also contribute to surface 1, which are colored accordingly.

The figure was produced using the program GRASP [44].

nating to the phosphate ion. The proposed role for this invariant residue is somewhat counterintuitive, as His106 also hydrogen bonds with another invariant residue, Glu104, to form a histidine-glutamate pair similar to the residues (His143 and Glu86) implicated in the mechanism of the type I DHQase [15].

The six-membered ring of the substrate and product lie on the floor of the active site, which is formed from the carboxy end of strand β3 and helix α3′ (residues 79–85). The residues at positions 81 and 82 are highly conserved as either alanine or glycine, due, in part, to steric restrictions in the case of residue 82, which is almost totally buried at the trimer interface close to the Asp92*-Arg117 intrasubunit salt bridge. These residues form an extended conformation, with their planar peptide bonds giving rise to a flat surface with which the ligands interact.

The unambiguous identification of the C4 and C5 hydroxyls was possible in the R23A mutant DHS structure due to the high resolution of the X-ray data. Averaged omit maps show clearly the pucker in the ring system, the C4 hydroxyl oriented down, and the C5 hydroxyl oriented up, as found in the 2,3-anhydro-quinic acid structure. The C4 and C5 hydroxyl groups are important for substrate recognition, with the C5 hydroxyl of both 2,3-anhydro-quinic acid and DHS interacting with NE2 of His 85 and NH1 of Arg117. Arg117 forms a salt bridge with Asp92 from the neighboring subunit. This aspartate forms the major interaction with the C4 hydroxyl of the substrate through OD2. The carbonyl at C3 is only present in the mutant dehydroshikimate structure and forms a hydrogen bond to a water molecule that is found in all of the structures reported here and hydrogen bonded to ND2 of Asn16.

A key difference between the *S. coelicolor* DHQase structures reported here and the *M. tuberculosis* DHQase structure [15] is the presence of an ordered lid domain (residues 21–31) in which residues 25–28 adopt an α-helical structure. There is a significant hinged movement at residues Gly21 and Asp31, resulting in a movement of over 5 Å between complexed and apo structures (Figure 7A). Further evidence for flexibility is provided by the main chain conformational changes at Arg23 and Gly29. This structurally characterizes the ligand-induced conformational change proposed previously from proteolysis and chemical modification experiments [17]. When closed, the lid domain forms a loop that extends over the carboxy edge of strands β3 and β4 of the central sheet and interacts with helix α3 of a trimer-related subunit. In this orientation the side chain of Tyr28 is positioned into the active site with the hydroxyl, forming an H bond to the guanidino group of Arg113. The other invariant residue, Arg23, is not as well ordered in all structures but is not orientated toward the active site. In the complex with phosphate, Arg23 adopts an unusual conformation, packing against the ring of the tyrosine, while, in the apo structure and in the complex with 2,3-anhydro-quinic acid, it forms a salt bridge with Asp98 on the trimer-related subunit.

## Enzyme Mechanism

The first step of the proposed mechanism involves abstraction of the *pro-S* proton from the C2 carbon of dehydroquinate. It has previously been suggested that a

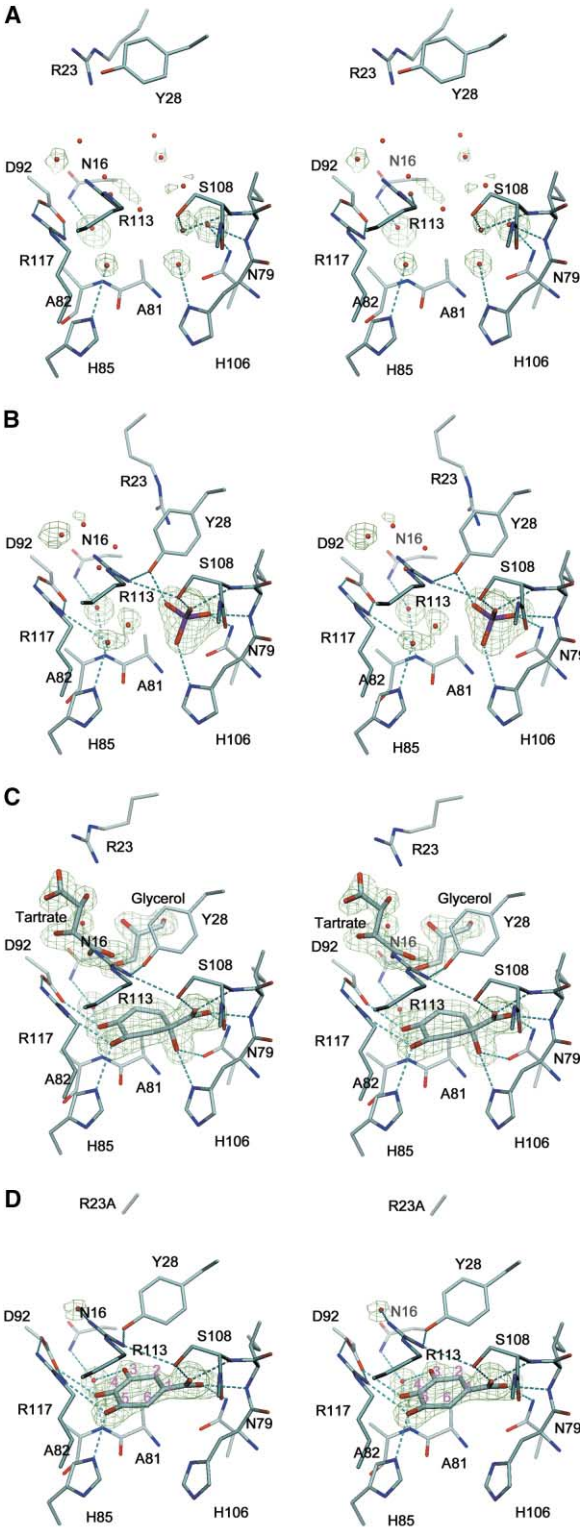


Figure 6. A Detailed View of the Active site, with Key Residues Labeled and Hydrogen Bonds ≤ 3.2 Å in Length Shown as Dashed Lines

The electron density shown is 12-fold averaged difference density, calculated after one round of refinement with all the active site waters and ligands removed. These omit maps show the active sites of the DHQase (A) in the absence of ligand, (B) complexed with phosphate, and (C) complexed with 2,3-anhydro-quinic acid (tartrate and glycerol labeled) and (D) the Arg23Ala mutant enzyme with dehydroshikimate (the carbon atoms numbered in magenta). In the 2,3-anhydro-quinic acid structure, glycerol and tartrate molecules present in the active site are shown (hydrogen bonds omitted for clarity). The glycerol is bound in a hydrophobic patch formed by residues Leu17, Leu20, and Tyr28 and forms hydrogen bonds to the carbonyl of Asn16, a carboxylate of tartrate, and several waters. The molecule of tartrate forms extensive hydrogen bond contacts with the protein, including electrostatic interactions between Arg23 and one of the carboxylates and Arg113 and the other carboxylate. The tartrate also forms hydrogen bonds from one of the hydroxyls to the side chains of Asp92 and Asn95 and a short contact from one of the carboxylate oxygens to the hydroxyl of Tyr28 (2.1 Å).

histidine (His106 in this structure) would have its basicity elevated as it forms a His-Glu pair with Glu104 and could act as the base toward the substrate [15]. The structure of the enzyme complexed with 2,3-anhydro-quinic acid unambiguously shows that His106 interacts with the C1 hydroxyl and does not have a role in proton abstraction.

Comparing the various structures of type II DHQase reported here, it is apparent that the only residue in the correct orientation for proton abstraction is the invariant residue Tyr28 (Figure 8). The hydroxyl of free tyrosine has a pKa of 10.0 and therefore would be expected to be fully protonated at pH 7.0; therefore, the pKa of Tyr 28 must be significantly shifted by its environment in the enzyme. The proximity of Arg113 and, to a lesser extent, Arg23 would be expected to have a significant effect on the pKa of Tyr218, as has been shown in other proteins, such as human aldose reductase [26].

Abstraction of the *pro-S* proton from C2 of the substrate would lead to a transition state with a double bond between C2 and C3 and the carbonyl at C3 presumably forming an enolate [14]. A basic residue would be needed to stabilize the enolate, but no residue is sufficiently close to perform this role. Analysis of all the structures reveals a conserved water molecule that is 2.8 Å away from the carbonyl of DHS and held in specific orientation by hydrogen bonds to an invariant residue, Asn16, the carbonyl of Pro15, and the main chain amide of Ala81. Whether or not the enolate obtains a solvent-derived proton to form an enol is debatable. On formation of an enol/enolate intermediate, the next step (removal of the hydroxyl at C1) is clearly brought about by His106 acting as the proton donor (Figure 8). The OD2 of Asn79 is a hydrogen bond acceptor for the C1 hydroxyl, thereby correctly orienting it to accept a proton. The role of the invariant Glu104 (data not shown) in the mechanism is unclear; from the structures, it can be seen that, in addition to hydrogen bonding to the imidazole of His106, the side chain of Glu104 hydrogen bonds to Ser120 and Ser123. It might be suggested, therefore, that Glu104 plays a structural role; however, in the product-bound form of the R23A mutant enzyme, Glu104 no longer forms a hydrogen bond with His106, resulting in the observed greater flexibility of this residue in the structure. It is possible that Glu104 has a role in the correct orientation of His106 for catalysis and must also move to permit the loss of water from the active site via a solvent-accessible cavity leading to the dimer interface (Figure 7B).

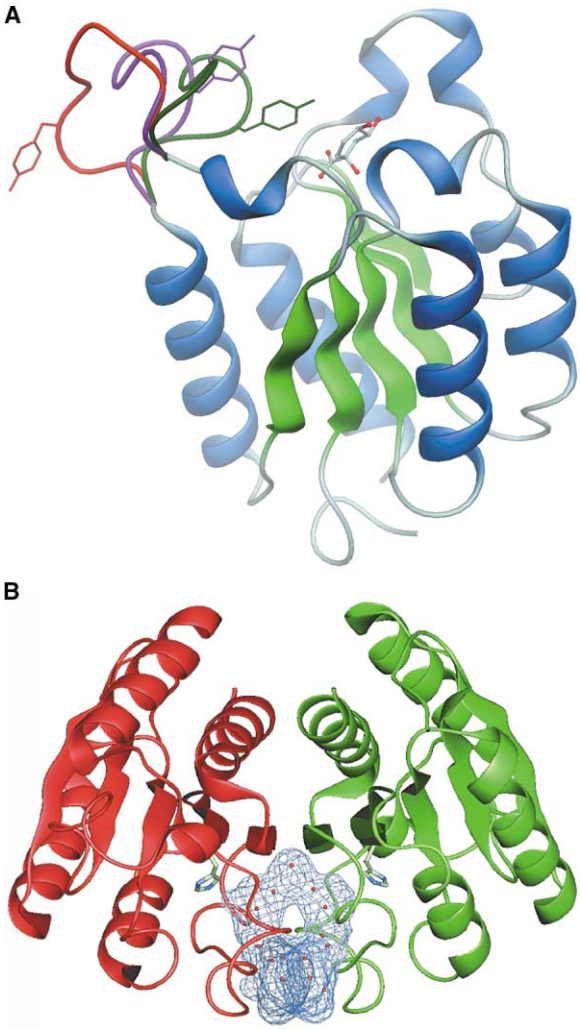


Figure 7. A Cartoon Representation of the Movement in the Flexible Lid Domain

The inhibitor structure 2,3-anhydro-quinic acid represented in ball and stick, and the corresponding lid domain in dark green. The two conformations of the lid domain seen in the apo structure are colored in red and mauve.

(B) A view along the central β sheet of the dimer illustrating a large solvent-accessible cavity formed in the hydrophilic face of the dimer, calculated using SURFNET [45], is shown in blue, and associated water molecules are shown as red spheres. The residue His106, the general acid involved in the removal of the hydroxyl at C1, is shown in both monomers to illustrate the proximity of this cavity to the active site of the protein.

## Discussion

The four structures reported give an insight into some of the changes that occur during enzyme catalysis. In the structure of the apo-enzyme, the lid loop is orientated in an "open" conformation, away from the active site in two quite distinct conformations, with the change in position of Tyr28 being especially notable (Figure 7A). In the active site there are five ordered water molecules, two occupying the carboxyl binding loop, two binding to the histidines that coordinate the C1 and C5 hydroxyls, and, finally, the conserved water that is bound to Asn16. In this apo-enzyme, Arg113 and His85 will give the active site a suitable net positive charge, which will be important for the binding of the substrate.

On binding to the substrate, the lid closes over the active site via a hinged movement between Gly21 and Asp31, as seen in the structures of the phosphate and inhibitor complexes. In addition to this hinged movement, there are significant changes in the main chain conformation involving Arg23 and Gly29, which are also a source of mobility in this loop. The result is effectively to close the active site and bring Tyr28 from over 10 Å away into a position to abstract the C2 axial proton. A catalytic role for this tyrosine had been indicated by the inactivation of the enzyme by tetranitromethane, which was shown to modify Tyr28 specifically [23].

The presumed orientation of Tyr28 for proton abstraction is not seen in any of the structures reported here. In the 2,3-anhydro-quinic acid complex, it is 3.44 Å away from C2, and, in the phosphate structure, it is 3.0 Å from C2. It is clear from these structures that, on ligand binding, Arg113 forms a hydrogen bond with Ser108, positioning the guanidinium group to interact with the hydroxyl of Tyr28. This is not sufficient in itself for the correct positioning of the hydroxyl, as illustrated by the R23A mutant structure. In this structure the side chain of Tyr28 is shifted over 4 Å away from C2 in comparison to the structures of the complexes with phosphate and the inhibitor 2,3-anhydro-quinic acid. Reassessment of site-directed mutagenesis results, obtained for R23 on the *S. coelicolor* enzyme [23], suggests that the replacement of Arg23 by smaller residues, regardless of charge, results in Tyr28 adopting an alternative conformation, more favorable than the one required for catalysis. The inactive conformation of the lid domain in the R23A mutant (the enzyme activity of this mutant is reduced by four orders of magnitude compared to that of the wild-type) is significantly more ordered on the basis of the temperature factors than the lid domains from the other 3 structures.

Harris et al. [14] have shown an unusual pH dependence of type II DHQases for the *A. nidulans* and *M. tuberculosis* enzymes, in which $k_{cat}$ increases sharply above pH 8 with a corresponding increase in the $K_m$. This led to the suggestion that the catalytic deprotonation was carried out by a histidine, with its basicity mediated by an arginine. As the pH increases, this leads to a reduced charge on arginine, thereby increasing the basicity of the histidine. From the structures reported here, it is clear that this explanation is broadly correct, albeit with a tyrosine, Tyr28, as the residue that performs the proton abstraction, its pKa mediated by the proximity of the side chain of Arg113 and possibly Arg23.

As a result of proton abstraction, it has been suggested that the substrate forms an enolate intermediate, with the double bond formed between C2 and C3 introducing an element of planarity into the six-membered ring. The 2,3-anhydro-quinic acid inhibitor mimics this proposed transition state, and it has been shown that inhibitors with a double bond between C2 and C3 bind significantly more tightly than those lacking this bond [25]. Although the carbonyl at C3 of the substrate is presumed to form an enolate intermediate, there is no amino acid residue close enough to stabilize the nega-

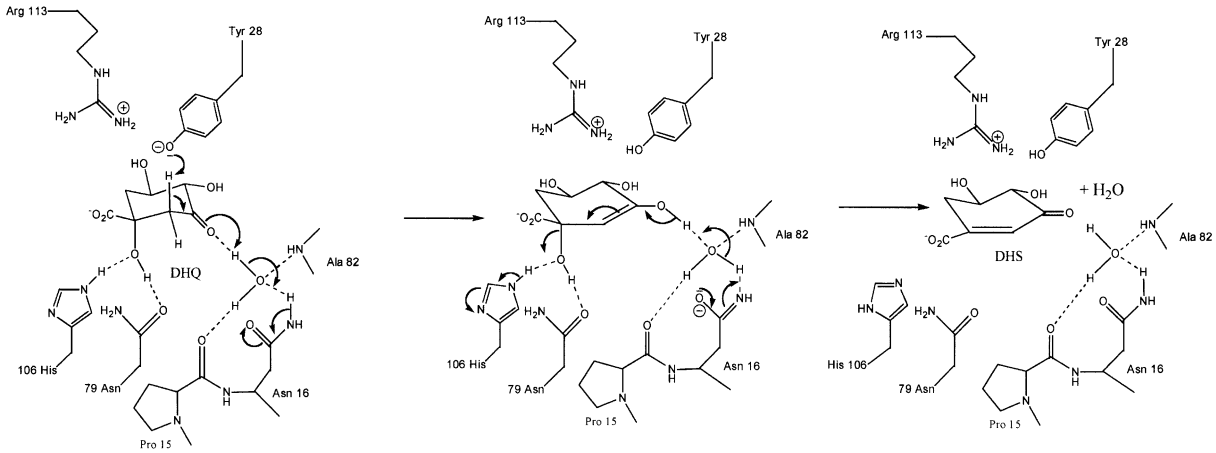


Figure 8. A Schematic Diagram of the Proposed Mechanism of Type II DHQases

tive charge. Instead, a conserved water molecule is correctly positioned 2.8 Å away from this group, which, in turn, coordinates to the ND2 of Asn15, the carbonyl of Pro14, and main chain amide of Ala82. From previous kinetic isotope studies [14] undertaken on the *A. nidulans* and *M. tuberculosis* enzymes in the absence of structural data, it was concluded that the data were consistent with an enolate rather than an enol formed during catalysis. In the light of the structures presented here, it is now possible to better interpret the kinetic data, particularly the large solvent isotope effect seen for the *A. nidulans* enzyme, which suggested that two protons contribute to the kinetically significant transition states. Only one water molecule is present in the active site, and, for it to contribute to an isotope effect, an enol intermediate rather than an enolate would need to be formed. The second proton would come from the subsequent elimination step, where the histidine acting as the general acid would have a proton readily exchangable with the solvent at the pH (7.0) at which the experiments were carried out.

The final step in catalysis is the acid-catalyzed elimination of the C1 hydroxyl. It is clear from the crystal structures of the enzyme complexed with 2,3-anhydroquinic acid and the R23A mutant complexed with DHS that His106 acts as the general acid, with Asn79 correctly positioning the C1 hydroxyl to accept a proton. On elimination, there is a conformational change in the product as the C1 becomes planar; however, there is insufficient space at the active site for a water molecule

Table 1. Crystollographic and Structure Determination Statistics

| Data Collection | | | | |
|---|---|---|---|---|
| Data Set | WT Apo | WT Plus Phosphate | R23A Plus DHS | WT Plus 2,3-Anhydro-Quinic Acid |
| Wavelength (Å) | 0.870 | 0.870 | 0.870 | 0.870 |
| Space group | $P2_12_12_1$ | $P2_12_12_1$ | $P2_1$ | $P2_12_12_1$ |
| Unit cell dimensions (a, b, and c in Å) | 116.2, 138.4, 141.2 | 113.2, 137.4, 140.2 | 86.1, 136.2, 96.7 | 113.6, 138.4, 141.5 |
| Resolution range (Å) | 30.0–1.8 | 45.0–1.8 | 45.0–1.6 | 30.0–1.8 |
| Observations | 429,534 | 798,417 | 426,136 | 648,516 |
| Unique reflections | 145,400 | 199,361 | 213,906 | 202,189 |
| Completeness (highest resolution shell) | 94.5% (93.5%) | 98.7% (93.5%) | 73.8% (21.5%[f]) | 98.1% (90.9%) |
| $R_{merge}$[a] (%) | 5.7 | 5.3 | 6.4 | 8.8 |
| **Refinement Statistics** | | | | |
| Resolution range (Å) | 24.0–2.0 | 45.0–1.8 | 45.0–1.6 | 22.0–1.8 |
| R factor[b] ($R_{work}/R_{free}$) | 19.9/24.2 | 14.9/20.3 | 17.8/21.2 | 14.9/19.1 |
| Number of atoms[c] | 13,327/32/1741 | 13,827/112/2023 | 13,572/176/1581 | 13,576/368/2635 |
| Rms bond length deviation (Å) | 0.018 | 0.018 | 0.019 | 0.019 |
| Rms bond angle deviation (°) | 2.0 | 1.8 | 1.9 | 1.8 |
| Mean B factor (Å)[d] | 18/19/—/32 | 20/22/26/37 | 11/13/17/29 | 8/10/7/28 |
| Rms backbone deviation (Å) | 1.1 | 0.9 | 0.7 | 0.7 |
| Coordinate error (Å)[e] | 0.250 | 0.252 | 0.124 | 0.120 |
| Intrasubunit average rms deviation (Å) | 0.114 | 0.222 | 0.185 | 0.118 |

[a] $R_{merge} = \Sigma|I - \langle I\rangle|/\Sigma\langle I\rangle$.
[b] R factor $= \Sigma|F_o - F_c|/\Sigma F_o$.
[c] Number of atoms of protein, heteroatoms, and water molecules, respectively.
[d] Mean B factor for main chain, side chain, inhibitor, and water atoms, respectively.
[e] Calculated using the method of Cruickshank [46].
[f] Data were processed to the edge (corners) of the CCD detector leading to a drop of in completeness above 1.8 Å resolution (76% complete).

and the product to be present at the same time. This may be a driving force for the release of product, but, additionally, there is a solvent cavity accessible to His106 through the dimer interface (Figure 7B). This cavity arises from the small size of residues such as Ser123, Val129, and Ala131, which are conserved over most type II DHQase sequences, suggesting that the release of the water molecule via this route may be important.

Finally, it should be noted that the structure of the type II DHQase was solved in complex with 2,3-anhydroquinic acid, which has a $K_i$ of 30 $\mu$M for this enzyme, 20-fold lower than the $K_m$ for substrate. This transition state analog has significant specificity for the type II enzymes, as the $K_i$ for the type I enzyme from *Salmonella typhimurium* is 3 mM [25]. The presence of a molecule of tartrate and glycerol in the active site of the structure (Figure 6C) was surprising and illustrates the available space and functional groups to be exploited in future inhibitor design.

## Biological Implications

The third reaction of the shikimate pathway is catalyzed by one of two structurally and mechanistically distinct enzymes: a type I or type II dehydroquinase (DHQase). The enzymes of this pathway have been identified as potential antimicrobial targets, with DHQases offering the possibility of targeting specific microbes.

The crystal structures of native type II DHQase from *S. coelicolor*, those in complex with the transition state analog 2,3-anhydro-quinic acid and the competitive inhibitor phosphate, and the catalytically compromised R23A mutant in complex with the product dehydroshikimate have been solved. These have allowed the definition of the active site residues and the mode of substrate binding. In addition, the lid domain (residues 21–31) not visible in the *M. tuberculosis* structure [15] has been resolved in both open and closed conformations.

From these structures we are able to propose a mechanism involving the 1,2-*anti* elimination of $H_2O$ from 3-dehydroquinate to form 3-dehydroshikimate via an $E_1CB$ mechanism with an enol intermediate. We propose that Tyr28 performs the deprotonation of 3-dehydroquinate as a result of a significantly reduced pKa arising from a basic environment formed by Arg113. The absence of any suitable residue to stabilize an enolate intermediate and the presence of a conserved active site water molecule instead supports the reaction proceeding via an enol intermediate. In the final step His106 acts as the general acid to catalyze the elimination of the C1 hydroxyl.

This proposed mechanism and the structures reported here offer the possibility for the rational design of selective mechanism-based inhibitors of type II DHQases based on 2,3-anhydro-quinic acid. These would block the shikimate pathway in pathogens such as *H. pylori* and *M. tuberculosis* without affecting important gut flora organisms, such as *E. coli*.

## Experimental Procedures

### Cloning, Sequencing, and Purification of DHQase from *S. coelicolor*

The N-terminal peptide of purified *S. coelicolor* DHQase [27] was used to design an oligonucleotide, taking into account the codon usage of *Streptomyces* genes [28]. This 38-mer (5′-GGCAACGCCC C(CG)ATCATGATCCTGAACGGC-CC(CG)AACCT) identified a 3.2 kb HindIII-SstI fragment in genomic Southern blots, which was cloned into pIBI25 [29]. The resulting plasmid was called pDHQ. The region around the site of hybridization by the oligonucleotide was sequenced manually by the dideoxy termination method. It has been deposited in the EMBL database with the accession number AJ001493. The construction of the R23A mutant enzyme has been described previously [23].

The protein was purified as previously described [27] with the following modifications: the initial ammonium sulfate fractionation step and the final heat treatment step were omitted. In particular the final heat treatment step proved to be detrimental for crystallization purposes. Protein was stored at −20°C in 50 mM Tris-HCl (pH 7.5) containing 30% glycerol.

### Crystallization and Data Collection

Protein was dialyzed overnight to remove glycerol and concentrated to 10 mg/ml using Centricon-30 centrifugal concentrators (Amicon, Stonehouse, Gloucestershire, UK). Initial crystallization attempts were focused around the conditions previously published for the *M. tuberculosis* enzyme [30], but these yielded crystals that were either twinned or dendritic in form. Successful crystals were obtained from linear screens using various PEG and salt combinations across a range of pHs. The crystals were morphologically of two habits: the first was cubic in form, and the other was orthorhombic. The cubic form crystals could only be indexed in a large cubic lattice (of F or P Bravais type), but the data would not merge in any space group higher than P1, so they were abandoned. Orthorhombic-shaped crystals grown from 20% PEG 8K and 0.2 M $Li_2SO_4$ in 0.1 M HEPES at pH 7.5 were found to diffract to better than 2.8 Å resolution at room temperature in house using a MacScience DIP2000 detector. Data were integrated and scaled using the HKL suite of programs [31]. The crystals were found to belong to the orthorhombic crystal system, with unit cell dimensions of a = 117.2 Å, b = 138.7 Å, and c = 141.4 Å. Analysis of the systematic absences in the data revealed absences at h = 2n, k = 2n, and l = 2n along the (h00), (0k0), and (00l) axes, respectively, which are consistent with the space group $P2_12_12_1$. From the Matthews calculation [32], assuming typical $V_m$ between 1.6 and 4.0, we estimated that there were between 9 and 22 monomers in the asymmetric unit. As the biologically active form of the enzyme is a dodecamer, we predicted that there were 12 monomers per asymmetric unit, giving a $V_m$ of 2.9 $Å^3\ Da^{-1}$, corresponding to a solvent content of 57.1%. This data was used to initially solve the structure. Crystallization conditions were optimized for each of the cocrystallization experiments, with the best crystals being obtained with 0.1M Tris (pH 8.5) and between 10% and 20% PEG 8K with 0.2 M of Na/K phosphate (phosphate), 0.5 M $MgCl_2$ (R23A mutant, apo-enzyme), and 0.2 M Na Tartrate (2,3-anhydro-quinic acid) added.

High-resolution datasets were collected from DHQase apo-enzyme and cocrystallised with phosphate, 2,3-anhydro-quinic acid, and an R23A mutant, DHQase with dehydroshikimate. The crystals were loop mounted in a cryoprotectant containing either 17.5% (v/v) glycerol or 20% MPD and cryocooled to 100 K using an Oxford Cryosystems cryostream. Data were collected in 0.5° oscillation frames on station 9.6 at the Daresbury SRS using the CCD (ADSC) Quantum 4 detector. The data were indexed and processed with the HKL suite [31]; the cell dimensions and space group are shown in Table 1. Further processing was carried out using programs from the CCP4 package [33].

### Molecular Replacement, Model Building, and Refinement

Molecular replacement was performed using the initial 2.8 Å data collected in house. A dodecamer constructed using the *M. tuberculosis* structure [15] was used as the search model. The level of sequence identity between the two enzymes is 40%. The crossrotation function and translation function were calculated using the data range of 10.0–4.0 Å in the standalone version of AMoRe [34]. Only six solutions were greater than 50% of the maximum peak height found in the rotation function, which were equivalent due to the inherent symmetry of the search model. Correct translation function solutions had correlation coefficients of 52.0 and R factors of 51.0.

The solutions were subjected to rigid body refinement in AMoRe to give a final correlation coefficient of 60% and an R-factor of 45.6%.

The resultant model was a full atomic model from the *M. tuberculosis* enzyme; therefore, structure factors were calculated from this model using SFALL, and then SIGMAA was used to calculate figures of merit to prepare input for phase refinement using the program DM [35]. Phase refinement included solvent flattening, histogram matching, and 12-fold noncrystallographic symmetry (NCS) averaging. The averaged electron density map calculated using DM phases was of excellent quality, allowing significant rebuilding of the structure using the program O [36]. Sulfate ions were found at the active site and on the NCS 3-fold axis. The measured reflections in the range 30–2.8 Å with no sigma cutoff were used, excluding 10.0% randomly distributed reflections to be used to calculate $R_{free}$ [37]. Rebuilding of the model was performed in the early stages using O and then subsequently using QUANTA (Molecular Simulations Inc.). Refinement of the model was performed using a maximum likelihood method as implemented in REFMAC [38] using tight NCS restraints. The model was refined to $R_{work} = 17.0\%$ and $R_{free} = 21.7\%$. This model was then used to phase the high-resolution data sets, extending the existing $R_{free}$ set of reflections to the highest resolution. Rigid body refinement or molecular replacement methods were used to obtain the starting model for each of the four structures.

Refinement was carried out using REFMAC with at first tight and then medium NCS restraints, and waters were added using ARP [39] and QUANTA (Molecular Simulations Inc.). In the later stages of refinement, the NCS restraints were removed, a fixed hydrogen atom contribution was added, and individual atom anisotropic temperature factors were applied where appropriate [40]. The stereochemical parameters, as determined using PROCHECK [41], were either inside or better than expected for the respective structures. There were two residues, Asn16 and Ala81, consistently in the generously allowed region; these were checked and found to be in convincing electron density in the structures. The final model statistics are shown in Table 1.

## Acknowledgments

This research was supported by Wellcome Trust grant number 043492, by BBSRC grant number SBD07628, and by the University of Glasgow. We would like to thank Mr. John Greene and Ms. Fiona Kellas for protein purification and Prof. Nicholas C. Price for useful discussions.

Received: October 3, 2001
Revised: December 18, 2001
Accepted: February 4, 2002

## References

1. Haslam, E. (1993). Shikimic Acid: Metabolism and Metabolites (Chichester, UK: J. Wiley and Sons).
2. Abell, C. (1999). Enzymology and molecular biology of the shikimate pathway. In Comprehensive Natural Products Chemistry, U. Sankawa, ed. (Amsterdam: Elsevier), pp. 573–607.
3. Roberts, F., Roberts, C.W., Johnson, J.J., Kyle, D.E., Krell, T., Coggins, J.R., Coombs, G.H., Milhous, W.K., Tzipori, S., Ferguson, D.J., et al. (1998). Evidence for the shikimate pathway in apicomplexan parasites. Nature *393*, 801–805.
4. Mousdale, D.M., and Coggins, J.R. (1991). Amino acid synthesis. In Target Sites for Herbicide Action, R.C. Kirkwood, ed. (New York: Plenum Press), pp. 29–56.
5. Davies, G.M., Barrett-Bee, K.J., Jude, D.A., Lehan, M., Nichols, W.W., Pinder, P.E., Thain, J.L., Watkins, W.J., and Wilson, R.G. (1994). (6S)-6-fluoroshikimic acid, an antibacterial agent acting on the aromatic biosynthetic pathway. Antimicrob. Agents Chemother. *38*, 403–406.
6. Giles, N.H., Case, M.E., Baum, J., Geever, R., Huiet, L., Patel, V., and Tyler, B. (1985). Gene organization and regulation in the qa (quinic acid) gene cluster of Neurospora Crassa. Microbiol. Rev. *49*, 338–358.
7. Hawkins, A.R., Lamb, H.K., Moore, J.D., Charles, I.G., and Roberts, C.F. (1993). The pre-chorismate (shikimate) and quinate pathways in filamentous fungi: theoretical and practical aspects. J. Gen. Microbiol. *139*, 2891–2899.
8. Shneier, A., Harris, J., Kleanthous, C., Coggins, J.R., Hawkins, A.R., and Abell, C. (1993). Evidence for opposite stereochemical courses for the reactions catalyzed by type-I and type-II dehydroquinases. Bioorg. Med. Chem. Lett. *3*, 1399–1402.
9. Hanson, K.R., and Rose, I.A. (1963). The absolute stereochemical course of citric acid biosynthesis. Proc. Natl. Acad. Sci. USA *50*, 981–988.
10. Harris, J., Kleanthous, C., Coggins, J.R., Hawkins, A.R., and Abell, C. (1993). Different mechanistic and stereochemical courses for the reactions catalyzed by type-I and type-II dehydroquinases. J. Chem. Soc. Chem. Commun. *13*, 1080–1081.
11. Butler, J.R., Alworth, W.L., and Nugent, M.J. (1974). Mechanism of dehydroquinase catalysed dehydration. I. Formation of a Schiff base intermediate. J. Am. Chem. Soc. *96*, 1617–1618.
12. Chaudhuri, S., Duncan, K., Graham, L.D., and Coggins, J.R. (1991). Identification of the active-site lysine residues of two biosynthetic 3-dehydroquinases. Biochem. J. *275*, 1–6.
13. Shneier, A., Kleanthous, C., Deka, R., Coggins, J.R., and Abell, C. (1991). Observation of an imine intermediate on dehydroquinase by electrospray mass spectrometry. J. Am. Chem. Soc. *113*, 9416–9418.
14. Harris, J.M., Gonzalez-Bello, C., Kleanthous, C., Hawkins, A.R., Coggins, J.R., and Abell, C. (1996). Evidence from kinetic isotope studies for an enolate intermediate in the mechanism of type II dehydroquinases. Biochem. J. *319*, 333–336.
15. Gourley, D.G., Shrive, A.K., Polikarpov, I., Krell, T., Coggins, J.R., Hawkins, A.R., Isaacs, N.W., and Sawyer, L. (1999). The two types of 3-dehydroquinase have distinct structures but catalyze the same overall reaction. Nat. Struct. Biol. *6*, 521–525.
16. Lo Conte, L., Ailey, B., Hubbard, T.J., Brenner, S.E., Murzin, A.G., and Chothia, C. (2000). SCOP: a structural classification of proteins database. Nucleic Acids Res. *28*, 257–259.
17. Bottomley, J.R., Hawkins, A.R., and Kleanthous, C. (1996). Conformational changes and the role of metals in the mechanism of type II dehydroquinase from Aspergillus nidulans. Biochem. J. *319*, 269–278.
18. Nguyen, V.T., Baker, D.P., Tricot, C., Baur, H., Villeret, V., Dideberg, O., Gigot, D., Stalon, V., and Haas, D. (1996). Catabolic ornithine carbamoyltransferase of Pseudomonas aeruginosa. Importance of the N-terminal region for dodecameric structure and homotropic carbomoylphosphate cooperativity (vol 236, pg 289, 1996). Eur. J. Biochem. *237*, pp. 884-884.
19. Villeret, V., Clantin, B., Tricot, C., Legrain, C., Roovers, M., Stalon, V., Glansdorff, N., and Van Beeumen, J. (1998). The crystal structure of Pyrococcus furiosus ornithine carbamoyltransferase reveals a key role for oligomerization in enzyme stability at extremely high temperatures. Proc. Natl. Acad. Sci. USA *95*, 2801–2806.
20. Price, N.C., Boam, D.J., Kelly, S.M., Duncan, D., Krell, T., Gourley, D.G., Coggins, J.R., Virden, R., and Hawkins, A.R. (1999). The folding and assembly of the dodecameric type II dehydroquinases. Biochem. J. *338*, 195–202.
21. Luscombe, N.M., Laskowski, R.A., Westhead, D.R., Milburn, D., Jones, S., Karmirantzou, M., and Thornton, J.M. (1998). New tools and resources for analysing protein structures and their interactions. Acta Crystallogr. D Biol. Crystallogr. *54*, 1132–1138.
22. Brandon, C., and Tooze, J. (1991) Enzymes that bind nucleotides. In Introduction to Protein Structure (New York: Garland Publishing, Inc.).
23. Krell, T., Horsburgh, M.J., Cooper, A., Kelly, S.M., and Coggins, J.R. (1996). Localization of the active site of type II dehydroquinases. Identification of a common arginine-containing motif in the two classes of dehydroquinases. J. Biol. Chem. *271*, 24492–24497.
24. Kleanthous, C., Deka, R., Davis, K., Kelly, S.M., Cooper, A., Harding, S.E., Price, N.C., Hawkins, A.R., and Coggins, J.R. (1992). A comparison of the enzymological and biophysical properties of two distinct classes of dehydroquinase enzymes. Biochem. J. *282*, 687–695.
25. Frederickson, M., Parker, E.J., Hawkins, A.R., Coggins, J.R.,

and Abell, C. (1999). Selective inhibition of type II dehydroquinases. J. Org. Chem. *64*, 2612–2613.
26. Tarle, I., Borhani, D.W., Wilson, D.K., Quiocho, F.A., and Petrash, J.M. (1993). Probing the active-site of human aldose reductase—site-directed mutagenesis of Asp-43, Tyr-48, Lys-77, and His-110. J. Biol. Chem. *268*, 25687–25693.
27. White, P.J., Young, J., Hunter, I.S., Nimmo, H.G., and Coggins, J.R. (1990). The purification and characterization of 3-dehydroquinase from Streptomyces coelicolor. Biochem. J. *265*, 735–738.
28. Wright, F., and Bibb, M.J. (1992). Codon usage in the G+C-rich Streptomyces genome. Gene *113*, 55–65.
29. Dente, L., and Cortese, R. (1987). Pembl—a new family of single-stranded plasmids for sequencing DNA. Methods Enzymol. *155*, 111–119.
30. Gourley, D.G., Coggins, J.R., Isaacs, N.W., Moore, J.D., Charles, I.G., and Hawkins, A.R. (1994). Crystallization of a type II dehydroquinase from Mycobacterium tuberculosis. J. Mol. Biol. *241*, 488–491.
31. Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.
32. Matthews, B.W. (1968). Solvent content of protein crystals. J. Mol. Biol. *33*, 491–497.
33. CCP4 (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D. Biol. Crystallogr. *50*, 760–763.
34. Navaza, J. (1994). AmoRe—an automated package for molecular replacement. Acta Crystallogr. A *50*, 157–163.
35. Cowtan, K. (1994) DM: an automated procedure for phase improvement by density modification. Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography *33*, 34–38.
36. Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron-density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.
37. Brunger, A.T. (1993). Assessment of phase accuracy by cross validation—the free R-value—methods and applications. Acta Crystallogr. D Biol. Crystallogr. *49*, 24–36.
38. Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. *53*, 240–255.
39. Lamzin, V.S., and Wilson, K.S. (1997). Automated refinement for protein crystallography. Methods Enzymol. *277*, 269–305.
40. Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S., and Dodson, E.J. (1999). Efficient anisotropic refinement of macromolecular structures using FFT. Acta Crystallogr. D Biol. Crystallogr. *55*, 247–255.
41. Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). Procheck: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291.
42. Carson, M. (1997). Ribbons. Methods Enzymol. *277*, 493–505.
43. Gouet, P., Courcelle, E., Stuart, D.I., and Metoz, F. (1999). ESPript: multiple sequence alignments in PostScript. Bioinformatics *15*, 305–308.
44. Nicholls, A., Bharadwaj, R., and Honig, B. (1993). Grasp—Graphical Representation and Analysis of Surface-Properties. Biophys. J. *64*, A166.
45. Laskowski, R.A. (1995). Surfnet—a program for visualizing molecular-surfaces, cavities, and intermolecular interactions. J. Mol. Graph. *13*, 323–330.
46. Cruickshank, D.W.J. (1999). Remarks about protein structure precision. Acta Crystallogr. D Biol. Crystallogr. *55*, 583–601.

## Accession Numbers

The coordinates for DHQase apo-enzyme, phosphate complex, 2,3-anhydro-quinic acid complex, and R23A mutant DHS complex have been deposited in the Protein Data Bank under accession codes 1GU0, 1D01, 1GU1, and 1GTZ, respectively.