Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Comparison of different crystal forms of 3-dehydroquinase from *Salmonella typhi* and its implication for the enzyme activity

Wen-Hwa Lee,[a]† Luis Augusto Perles,[b] Ronaldo A. P. Nagem,[a,b] Annette K. Shrive,[c]‡ Alastair Hawkins,[d] Lindsay Sawyer[c] and Igor Polikarpov[a,e]*

[a]Laboratório Nacional de Luz Síncrotron/LNLS, Campinas, SP, Brazil, [b]Departamento de Física, Universidade Estadual de Campinas (UNICAMP), Campinas, SP, Brazil, [c]Structural Biochemistry Group, ICMB, University of Edinburgh, Michael Swann Building, King's Buildings, Mayfield Road, Edinburgh EH9 3JR, England, [d]Department of Biochemistry and Genetics, New Medical School, Framlington Place, University of Newcastle upon Tyne, Newcastle-upon-Tyne NE2 4HH, England, and [e]Grupo de Cristalografia, Instituto de Física em São Carlos, Universidade de São Paulo, Av. Trabalhador Sãocarlense, 400, CEP 13560-970, São Carlos, SP, Brazil

† Present address: Computational Biology and Bioinformatics, Department of Molecular Biology TPC-28, The Scripps Research Institute, La Jolla, CA, USA.
‡ Current address: School of Life Sciences, Keele University, Keele, Staffordshire ST5 5BG, England.

Correspondence e-mail: ipolikarpov@if.sc.usp.br

Received 27 November 2001
Accepted 28 February 2002

PDB Reference: type I DHQase, 1l9w, r1l9wsf.

The type I 3-dehydroquinate dehydratase (DHQase) which catalyses the reversible dehydration of 3-dehydroquinic acid to 3-dehydroshikimic acid is involved in the shikimate pathway for the biosynthesis of aromatic compounds. The shikimate pathway is absent in mammals, which makes structural information about DHQase vital for the rational design of antimicrobial drugs and herbicides. The crystallographic structure of the type I DHQase from *Salmonella typhi* has now been determined for the native form at 1.78 Å resolution (R = 19.9%; $R_{\rm free}$ = 24.7%). The structure of the modified enzyme to which the product has been covalently bound has also been determined but in a different crystal form (2.1 Å resolution; R = 17.7%; $R_{\rm free}$ = 24.5%). An analysis of the three available crystal forms has provided information about the physiological dimer interface. The enzyme relies upon the closure of a lid-like loop to complete its active site. As the lid-loop tends to stay in the closed position, dimerization appears to play a role in biasing the arrangement of the loop towards its open position, thus facilitating substrate access.

1. Introduction

The enzyme 3-dehydroquinase (3-dehydroquinate dehydratase; DHQase; EC 4.2.1.10) catalyses the reversible dehydration of 3-dehydroquinic acid to 3-dehydroshikimic acid. This reaction occurs in two distinct metabolic pathways: (i) the biosynthetic shikimate pathway of aromatic compounds in microorganisms and plants (Bentley, 1990; Haslam, 1993) and (ii) the catabolic quinate pathway, a carbon-scavenging pathway common to many microbial species (Giles *et al.*, 1985).

The DHQases can be divided into two different classes according to the mechanism of action, stereochemistry, overall structure and sequence homology (White *et al.*, 1990; Servos *et al.*, 1991; Kleanthous *et al.*, 1992; Harris *et al.*, 1996). The type I enzymes, which are involved only in biosynthesis, occur either as homodimers of subunit M_r = 27 000 or as components of multifunctional enzymes with other shikimate-pathway enzymes (Lumsden & Coggins, 1977; Charles *et al.*, 1986; Bentley, 1990; Deka *et al.*, 1994). They use a covalent imine intermediate to catalyze a *syn* elimination (Butler *et al.*, 1974; Chaudhuri *et al.*, 1991). The type II enzymes are dodecamers consisting of identical subunits of M_r 16 000. They work by an entirely different mechanism and catalyze a *trans* elimination that almost certainly involves an enolate-type transition-state mechanism (Kleanthous *et al.*, 1992; Gourley *et al.*, 1994; Bottomley *et al.*, 1996; Harris *et al.*, 1996).

Recently, the structures of both a type I (from *Salmonella typhi*) and a type II (from *Mycobacterium tuberculosis*) DHQase were solved (Gourley *et al.*, 1999). The overall structure of type I DHQase from *S. typhi* consists of a single domain that folds into the commonly observed eight-stranded α/β (or TIM) barrel.

A comparative analysis of the native type I enzyme with that of the borohydride-reduced product–enzyme complex has suggested that there are critical structural changes that affect the activity of the enzyme. One of the loops of the enzyme, that between strand *h* and helix *H* (see below), acts as a lid for the active-site cleft, shielding the active site from the solvent environment.

Small differences in the crystallization conditions lead to significantly different crystal forms. Here, we present the structure solution of two new crystal forms of the type I DHQase enzyme from *S. typhi* and compare them with the original crystal form of the same enzyme (Gourley *et al.*, 1999). We have used the following arbitrary nomenclature: crystal form I (native enzyme; new; $P2_12_12$, one chain per asymmetric unit), crystal form II ($P2_1$, two chains per symmetric unit; PDB code 1qfe) and crystal form III (borohydride-reduced enzyme–product complex; Chaudhuri *et al.*, 1991; new; $P2_1$, four chains per asymmetric unit).

2. Materials and methods

Samples of DHQase from *S. typhi* were purified to homogeneity from an overproducing strain of *Escherichia coli* and subjected to a sparse-matrix crystallization screen. Several crystallization trials were performed and the best conditions were found using PEG 4000 as precipitant and 100 m*M* citrate–phosphate as buffer in the pH range 5.0–6.5 (Boys *et al.*, 1992). Soaking with different heavy atoms yielded derivatives for MIR phase determination (Gourley *et al.*, 1999).

X-ray diffraction data sets were collected at room temperature using synchrotron radiation at stations X31 and X11 at the EMBL Outstation, Hamburg and an Enraf–Nonius FR571 X-ray generator in-house. The images were collected using a MAR image-plate detector. The data were processed, indexed and merged using the programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). Data-collection statistics are summarized in Table 1. The structure of crystal form II was determined by multiple isomorphous replacement as described in Gourley *et al.* (1999). The structures of crystal forms I and III were solved by molecular replacement (*AMoRe*; Navaza & Saludjian, 1997) using the structure of crystal form II as search model. The single main peaks in the rotational function had correlation coefficients of 0.31 and 0.19 for crystal forms I and III, respectively, with the next highest peaks in both forms having values of less than 0.1. These rotation solutions were used in the translation-function search and the best solutions were subjected to ten cycles of rigid-body refinement, yielding an *R* factor of 21.6% for crystal form I (88.2% correlation factor) and 27.1% for crystal form III (81.3% correlation factor). Crystallographic refinement was carried out using the maximum-likelihood method implemented in the program *REFMAC* (Murshudov *et al.*, 1997; Collaborative Computational Project, Number 4, 1994). Successive rounds of cycles of refinement interspersed with manual adjustment using the program *O* (Jones & Kjeldgaard, 1993) were employed to improve the quality of the original molecular-replacement solution. Non-crystallographic symmetry restraints were in place at this stage. The model was inspected using both $F_o - F_c$ and $2F_o - F_c$ electron-density maps. The progress of the refinement was monitored using both the conventional and free *R* factors (Brünger, 1992). The refinement rounds continued until both *R* factors reached a minimum value, with no further improvements as new rounds were requested. Water molecules were then added with the program *ARP* (Lamzin & Wilson, 1993) before continuing refinement with the program *REFMAC*.

The *R* factor of the final model of crystal form I was 17.6%, with an $R_{\rm free}$ of 22.6% for data in the resolution range 10–1.78 Å. Removal of the NCS restraints at the end of the

Table 1
Crystallographic data and statistics for DHQase.

	Crystal form I	Crystal form II	Crystal form III
Data and parameters			
Space group	$P2_12_12$	$P2_1$	$P2_1$
Unit-cell parameters			
a (Å)	48.78	60.49	42.61
b (Å)	112.33	45.39	158.56
c (Å)	42.94	85.47	85.89
β (°)	90.00	95.48	93.61
Z (No. of chains in a.u.)	1	2	4
$V_{\rm M}$ (Å^3 Da^{-1})	2.12	2.12	2.61
Solvent content† (%)	41.62	41.50	52.57
Completeness (%)	97.2	86.1	76.45
$R_{\rm merge}$ (%)	5.1	10.4	6.9
Maximum resolution (Å)	1.78	2.1	2.1
R factor (%)	19.9	17.6	17.7
$R_{\rm free}$ (%)	24.7	22.6	24.5
Redundancy	4.12	2.84	3.42
No. unique reflections	22395	23517	49700
Ramachandran plot			
Most favoured (%)	93.9	94.1	93.9
Allowed (%)	4.4	5.0	5.2
Generously allowed (%)	1.7	0.9	0.8
Disallowed (%)	0.0	0.0	0.1
RMS deviation			
Bond length (Å)	0.013	0.010	0.018
Bond angle (°)	1.7	1.5	2.4

† Assuming a protein density of 1.34 g cm^{-3}.

Table 2
Interactions occurring in the dimer interface.

		Distance (Å)			
				Crystal form III	
Residue atom	Residue atom	Crystal form I	Crystal form II	Subunits 1 and 2	Subunits 3 and 4
Lys178 NZ	Val218 O	2.89	2.85	3.03	2.93
Lys207 NZ	Ala252 O	2.32	2.54	2.99	2.61
Val218 O	Lys178 NZ	2.89	3.06	2.97	3.00
Ala252 O	Lys207 NZ	2.32	2.67	3.27	3.02
Gly235 N	Ala252 O	2.64	n.a.	n.a.	n.a.

refinement followed by a few further cycles led to the R_{free} rising slightly, thereby indicating that their removal was unjustified (Kleywegt & Jones, 1997). The model for crystal form III had an R factor of 17.7% in the resolution range 10–2.1 Å (R_{free} = 24.5%). Both new structures have acceptable quality statistics as reported by *PROCHECK* (Laskowski *et al.*, 1993) and Ramachandran plots; statistics for the final crystallographic models are shown in Table 1 together with those from form II for comparison.

3. Results and discussion

3.1. General description of the *S. typhi* type I DHQase structure

The geometry of the molecules in each of the crystal forms is good, with all but one of the residues falling at least within the generously allowed regions of the Ramachandran plots. The residue lying in a disallowed region is Lys7 from subunit 4 of crystal form III, the distortion of which is caused by the interaction of its carbonyl group with the charged side chain of Lys160 of subunit 1.

The DHQase molecule has a typical $(\alpha/\beta)_8$ (TIM barrel) structure (Fig. 1), with two short antiparallel β-strands located at the N-terminal end of the barrel that block it off. The opposite end of the barrel provides the means by which substrate can reach the active site of the enzyme. While the short loops connecting the β-strands and α-helices consist of six residues on average, the *hH* loop (residues 227–239) contains 13 residues, four of which, including Gln236 which makes direct contact with the substrate, are strictly conserved. In addition, this loop is located at the C-terminal end of the barrel, adjacent to the entrance of the active site. When substrate is in the active site, the loop is closed; in the absence of substrate, the loop swings open and appears to adopt several conformations, as suggested by the poorly defined electron density and larger average *B* factors.

A particular feature of the family of TIM-barrel enzymes is the packing of side chains in the core. Normally, the core of a TIM barrel is arranged in three layers, where each layer contains four side chains from alternate β-strands (Branden & Tooze, 1999). In the structure of DHQase, Ile44, Met112, Ile168 and Ala223 of strands *b*, *d*, *f* and *h*, respectively, form the first layer, closest to the N-terminal end of the barrel. The second, or middle, layer comprises Ile19, Leu78, Val139 and Ile201 of strands *a*, *c*, *e* and *g*, respectively. Finally, the third layer, nearest the C-terminal end of the barrel, contains the active site but is made up of the polar residues Ser21, Glu46, Arg48, Thr80, Arg82, Asp114 and Lys170 on strands *a*, *b*, *b*, *c*, *c*, *d* and *f*, respectively.

In DHQase, closure of the *hH* loop forms an additional hydrophobic layer on top of the third hydrophilic one, composed of residues Phe145, Ala172, Met205 on strands *e*, *f* and *g*, respectively, and residues Ala233 and Pro234 of the *hH* loop.

3.2. The physiological dimer and crystal form I

The main interface of the molecule responsible for specific dimer formation is present in solution as the physiological unit (Reilly *et al.*, 1994) and is also found in all three lattices. It is formed by the side chains of residues located on helices *F*, *G* and *H* of one monomer that pack against the same helices of the other monomer. The barrels are arranged side-by-side,

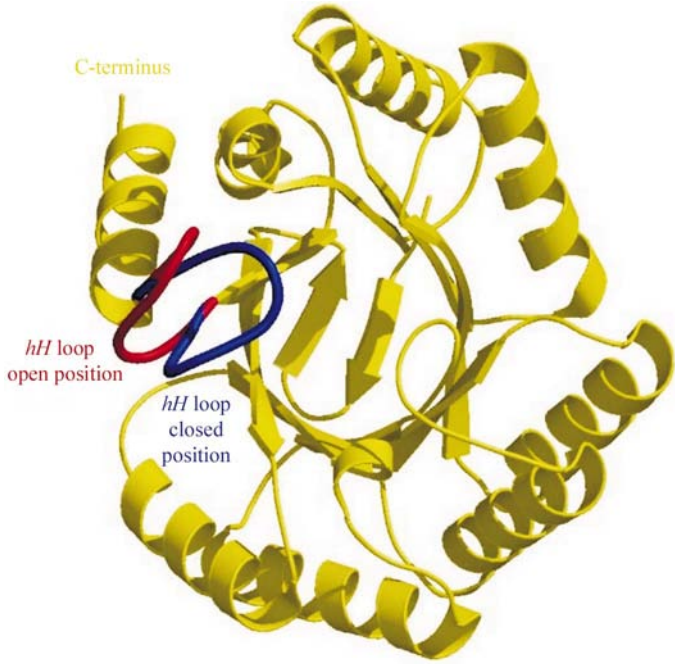

Figure 1
The overall fold of the DHQase polypeptide chain (yellow) with the lid loop depicted in the open position (red, crystal form I) and closed position (blue, crystal form II). The diagram was produced using *MOLSCRIPT* (Kraulis, 1991) and *Raster3D* (Merritt & Bacon, 1997).

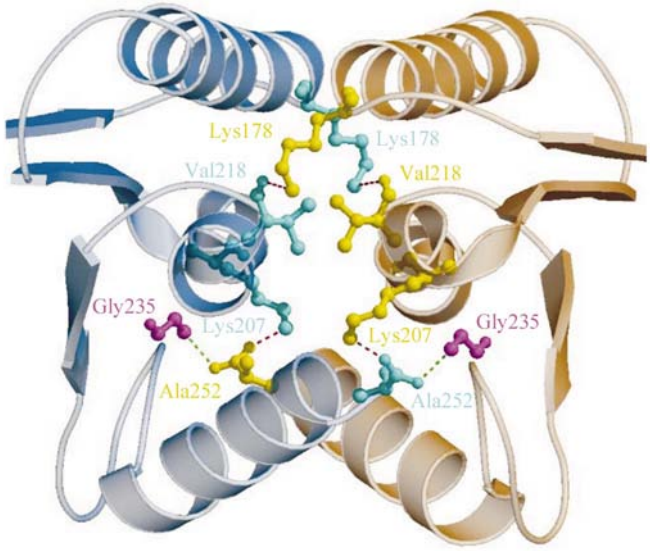

Figure 2
The physiological dimer interface. Blue and yellow shades represent elements belonging to different subunits. Magenta depicts the loop residue Gly235 interacting with Ala252.

with the active sites facing in opposite directions (Fig. 2). There are four residues involved in this packing, making only four interactions (Table 2). It is interesting to note that although the enzyme is a dimer, there are relatively few hydrogen bonds in the dimer interface. While generic dimeric interfaces have on average 0.88 ± 0.40 hydrogen bonds per 100 Å^2 buried surface area per subunit (Jones & Thornton, 1995), the dimer interface of the DHQase has only 0.34 hydrogen bonds per 100 Å^2 buried surface area per subunit. However, this is the highest value found of all the interfaces present in the three crystal forms. Moreover, the dimer-dissociation constant can be estimated from the data in Kleanthous *et al.* (1992) to be about 18 μ*M* (or $\Delta G^o \simeq 25$ kJ mol^{-1}), which is not an atypical value for a dimer-dissociation constant. The ΔG^o value is also consistent with the number of hydrogen bonds observed.

The native crystal, form I, has the subunit as the asymmetric unit, indicating a strict molecular as well as crystallographic twofold symmetry. However, in crystal forms II and III there is no such restriction. Superimposing the dimers from the different lattices showed no significant RMS deviation in the C$^\alpha$-atom positions. This suggests that the dimer interface is relatively stable, without significant movements between the monomers (such as a hinge movement), despite the few formal interactions that hold the two chains together. In addition, the RMS deviation of the side-chain atoms involved in the dimer interface shows there is no significant difference in their positions. Further supporting evidence of the high stability of this interface can be inferred from the solvent-accessible area buried on dimerization. The dimer interface buries ∼1150 Å^2 of the accessible surface area per dimer and similar values are found in all three crystal forms. The percentage of buried surface at this interface is thus constant at ∼10.6% for each lattice and lies well within the range typically found for strongly associated dimers (Jones & Thornton, 1995; Tables 2 and 3).

Table 2 shows that in crystal form I, the native enzyme, the C-terminal carboxyl group makes an additional hydrogen bond to the NH group of Gly235 in the open *hH* loop (Fig. 2). In the structures of the enzyme with labelled Lys170 in the active site (crystal forms II and III, loop closed), this Gly235–Ala252* interaction is absent. The presence of the *hH* loop in the interface in crystal form I increases the buried surface area by a small amount, reflecting the presence of the single hydrogen bond from Gly235.

3.3. The active site

Binding of the substrate to the native enzyme induces closure of the *hH* loop, breaking the Gly235–Ala252* hydrogen bond. The energy required is offset to some extent by the interactions between Ser234, Gln236 and the substrate. Several water molecules are also displaced during binding, from both the cavity and from the loop. For example, there are waters equivalent to the substrate 4- and 5-hydroxyl groups and a line of waters between Ser21 and Lys170, all of which occupy the space required by the C_3–C_5 part of the substrate ring. However, there is no clearly defined solvent observed in the volume occupied by the C_6, C_1 and C_2 C atoms or the carboxylate of the substrate. This may be because of the largely hydrophobic nature of the side chains of Phe225 and Met203 which dominate this part of the active site. Further, the side chain of Arg213, which forms a hydrogen bond with the main-chain carbonyl group of Phe225 in the free enzyme structure, moves to form an ion pair with the carboxyl group when the substrate binds. As observed elsewhere in crystal form II (Gourley *et al.*, 1999), the substrate is exquisitely located in the active site and the final hydrogen bonds with Gln236 and Ser234 are presumably formed as the lid loop *hH* swings over to close off the active site from the solvent.

In the absence of substrate, this loop is mobile and the hydrophobic nature of the fourth layer of residues should favour the closed conformation, but access to the active site requires it to be open. Thus, it seems possible that the Gly235–Ala252* hydrogen bond in crystal form I will enhance the activity of the enzyme by counteracting the slight bias towards the closed position of the loop. This explains why in the absence of substrate we see the loop, albeit sketchily, in the open substrate-receptive position. In fact, the average *B* factors for the *hH* loop in crystal form I are significantly higher than those of crystal forms II and III (Table 4).

In the active site of crystal form III, the dehydroshikimate (product) molecule has the same environment in each of the four independent chains of the asymmetric unit. Each subunit contains one borohydride-reduced dehydroshikimate, covalently linked to residue Lys170. As in crystal form II, the substrate is coordinated by Ser21 OG, Glu46 OE2, Arg48 NH1, Arg213 NH1 and NH2, Ser232 OG and Gln236 NE2. The active sites in all crystal forms are alike, with the exception of the Met203 and Arg213 side chains in the native crystal form I. In the native enzyme, both of these side

Table 3
Accessible areas of the functional dimer in three crystal forms, arranged in pairs.

The last pair (crystal form III, subunits 2 and 3) illustrates an ordinary crystal-packing interaction. Accessible areas were calculated using the program *AREAIMOL* (Collaborative Computational Project, Number 4). Structural waters were ignored in the calculations.

Crystal form	Area per subunit† (Å^2)	Area per subunit (after dimerization)‡ (Å^2)	Average buried area§ (Å^2)	Average % buried area
I (subunit 1)	11274	10094	1181.5	10.5
I (subunit 2)	11302	10119		
II (subunit 1)	10881	9698	1161	10.8
II (subunit 2)	10666	9527		
III (subunit 1)	10634	9537	1484.5	13.6
III (subunit 2)	10452	9357		
III (subunit 3)	10500	9389	1108	10.6
III (subunit 4)	10463	9358		
III (subunit 2)	10452	10115	329	3.1
III (subunit 3)	10500	10183		

† Subunit accessible areas were calculated separately from their respective pairs in the crystal forms. ‡ Accessible area calculated with the assembled dimer. § Calculated from the difference between 'area per subunit' and 'area per subunit after dimerization', hence giving the hidden area for every subunit arising from dimerization.

Table 4
B factors (Å^2) of the lid-loop residues.

Standard deviations are given in parentheses after the average values.

		Crystal form II (Å^2)		Crystal form III (Å^2)			
	Crystal form I (Å^2)	Subunit 1	Subunit 2	Subunit 1	Subunit 2	Subunit 3	Subunit 4
Loop (residues 227–239)	46.9 (23.2)	15.9 (5.2)	16.2 (5.1)	38.4 (8.4)	38.5 (8.7)	39.9 (8.3)	38.2 (8.3)
Subunit overall	26.7	15.3	15.9	29.2	28.7	31.3	30.8

chains are displaced toward the exit of the active site. When the substrate is present (crystal forms II and III) both Met203 and Arg213 adopt a more 'introspective' position. This is related to the closure of the loop, which breaks the interactions of Arg213 in order to form new ones with the substrate carboxyl group. Met203 must then move back in order to accommodate the substrate carboxylate in the active site. No other differences were noticed in the positions of the main and side chains.

The placement of the substrate in the active site of cystal form III was guided by the electron density. However, the electron density for the same regions was poorer than in crystal form II. Although the electron density was continuous, its volume was larger than in form II, suggesting less precise location. In fact, the electron density was slightly stretched midway between atom C3 of the substrate and the ring of Phe145. The program *ARP* (Collaborative Computational Project, Number 4, 1994) assigned two water molecules to this position. Interestingly, the same water molecules are present in the native form, which suggests that the active site may not have been fully occupied by covalently bound substrate in this crystal form. However, in subunits 1 and 3 of form III even the *B* factors are very similar to those found in the native form (subunit 1, form III, Wat169, 43.5 Å^2, Wat208, 32.5 Å^2; subunit 3, Wat204, 40.5 Å^2, Wat211, 33.5 Å^2; form I, Wat103, 51.6 Å^2, Wat32, 32.4 Å^2), yet the product 3-dehydroshikimate is clearly visible.

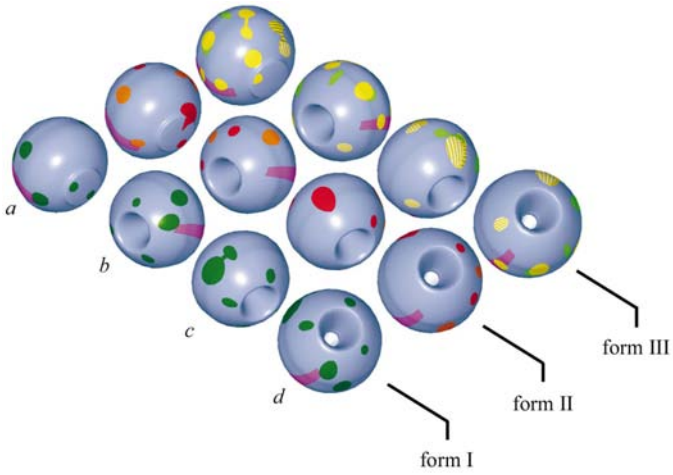

Figure 3
Schematic drawings of the different packing patterns of each crystal form. Contact areas are depicted as coloured patches colour coded for each crystal form (green for I, red for II and yellow for III). The active site is represented as a cavity and the bottom of the barrel as a flat raised disc on the sphere, clearly seen in row *a*. The magenta stripe represents the position of helix *H*. Comparison of the contact areas utilized by the three crystal forms is shown in four different orientations, *a*, *b*, *c* and *d*, but observed from the same viewpoint in each crystal form. In crystal forms II and III, where there is more than one subunit per asymmetric unit, the patches are colour coded according to the subunit to which the patch belongs. In crystal form II, subunit 1 has red contact patches, while those of subunit 2 are orange. In crystal form III the colour scheme is the following: subunit 1 (yellow with black stripes), 2 (green), 3 (yellow with white stripes) and 4 (plain yellow).

In addition to the several interactions with structural water molecules, a network of intramolecular hydrogen bonds also stabilizes the residues involved in binding the substrate. For example, Gln236 interacts with Ser232 in order to orient it towards the DHQ molecule. There are also hydrophobic interactions among the C atoms of the DHQ ring and several residues, two of which are implicated in the fourth hydrophobic layer (Met203 and Ala233, see above).

Finally, there is a significant increase in the stability of the enzyme when the substrate is bound (Kleanthous *et al.*, 1991): the concentration of GuHCl required to unfold the native enzyme is about 1.5 *M* and this increases to 4 *M* when the substrate or product is bound. Exclusion of water from the active site together with formation of the fourth hydrophobic layer locks the core of the structure, making it resistant both to chaotropic agents and to thermal denaturation.

3.4. Crystal form III

Two physiological dimers constitute the asymmetric unit of the crystal form III. These two dimers interact solely through subunits 2 and 3, making a structure similar to the Greek capital letter lambda (Λ). There are two hydrogen bonds between these two subunits, namely 2-Arg38 NH1 to 3-Asn135 OD1 (2.34 Å) and 2-Ala71 N to 3-Ala133 O (3.19 Å). In addition, there is an interaction mediated by a water molecule, bridging residues 2-Glu39 OE1–Wat88–3-Asp167 OD1 (2.78 and 3.28 Å). One face of subunit 2 of the 1–2 dimer interacts with the edge of a face of subunit 3 of the 3–4 dimer. In fact, the C-terminal ends of helices *A* (Arg38) and *B* (Ala71) in subunit 2 interact with the end of helix *D* (Ala133 and Asn135) in subunit 3. This interface buries an area of some 327 Å^2 which is well below the value expected for a physiologically relevant interface (Jones & Thornton, 1995).

Inspection of the 'lambda' interface and superposition of several chains from different crystal forms reveal that the side chains do not differ significantly in conformation, except for the side chain of Arg38 of subunit 2. In all other subunits, this side chain points straight out into the solvent. In crystal form III, residue 2-Arg38 assumes a relatively well defined conformation (*B* factor = 37.5 Å^2, compared with higher values for the residues *n* either side) that allows formation of the two hydrogen bonds that bind the subunits 2 and 3 together. None of the residues involved in the lambda interface is conserved except Asp167, suggesting that these interactions occur merely as a result of the crystallization conditions.

Table 5
Residues involved in crystallographic hydrogen-bonding contacts.

Residues making contact in the all three crystal forms are listed in the first column. The following columns specify the subunit and the residue(s) that are involved in the crystallographic contact(s) in each crystal form. The numbers in parentheses after residues identify the subunit to which they belong.

Residue	Form I (17)†		Form II (16)†		Form III (18)†	
	Subunit	Contacting	Subunit	Contacting	Subunit	Contacting
Arg38	1	Asp121	2	Ala252	2, 4	Asn135 (2), Asp167 (4)
Gln59	1	Asn8	1	Arg66	2	Ala133
Asp123	1	Lys2	1	Lys2	2	Thr93

† The values in parentheses are the total number of interacting residues in each crystal form.

3.5. Crystallographic contacts

Each crystal form presents several distinct crystallographic contacts, few of which are common to the three crystal forms. However, in each case different residues combine to form different patches that make the crystallographic contacts. There are few of these patches and in general they comprise at most three residues. The main patch is that forming the physiological dimer and is only present as a crystallographic contact in crystal form I. A schematic drawing of the areas of contact for each of the crystal forms is presented in Fig. 3, while Table 5 summarizes the crystallographic contacts made by residues Arg38, Gln59 and Asp123 in all three crystal forms. It is clear that these three residues do not form a patch common to all forms. The number of residues involved in crystallographic interactions of each of the three forms is similar (17 in form I, 16 in form II and 18 in form III) and often the contact is made through a single residue. Further, despite the similarity of the unit-cell lengths ($a^{\rm I} \simeq c^{\rm I} \simeq b^{\rm II} \simeq a^{\rm III}$, $2b^{\rm II} \simeq c^{\rm II}$; $2a^{\rm III} \simeq c^{\rm III}$, $c^{\rm II} \simeq c^{\rm III}$), which is presumably a reflection of the approximately cylindrical nature of the dimers, the arrangement of molecules is not related. Thus, for example, the *bc* face of form II bears little resemblance to the *ac* face of form III, a consequence of the widespread distribution of potential crystal contacts all over the molecular surface.

The lack of a consistent pattern of interacting patches or residues could explain the significant polymorphism that has hindered crystallographic studies. The borohydride-reduced Schiff-base intermediate of the enzyme that was eventually solved was expected to lock the structure and produce a unique crystal form, but it is in fact only the native form that makes some use of an active site-related crystal contact and involves the base of the lid loop *hH*: the side chains of residues at the base of the loop (Lys229 and Asn240) interact with a crystallographic neighbour.

4. Conclusions

Careful analysis of the different crystal forms shows that there are no very specific patches on the molecule surface that lead to a unique crystal packing. The effect of this is that there are several possible packing arrangements which can form from essentially the same conditions.

The environment of the substrate–product molecule is one that provides high specificity and is unaffected by the different crystal forms. Despite the large number of specific interactions between substrate and enzyme, slight variations in position do appear to be possible. However, it is not clear from this study that these minor differences have any functional significance.

Careful comparison of the three different crystal forms suggests that the functional dimer interface is implicated in the enzyme mechanism. It appears that the open position of the *hH* loop is stabilized by the dimer interface, thus facilitating access to the active site by the substrate. The stabilization, in the manner of a weak catch, is achieved through a hydrogen bond formed between the C-terminal residue from one subunit, Ala252, and the conserved Gly235 residue in the loop. When substrate binds, new hydrogen bonds as well as several hydrophobic interactions (the fourth hydrophobic layer) are formed, completing a solvent-shielding layer that allows the reaction to proceed efficiently.

We are grateful to Professor John Coggins for helpful discussions. The EMBL Outstation at Hamburg is thanked for providing data-collection facilities on stations X11 and X31. The financial assistance of the Brazilian funding bodies CNPq, FAPESP (*via* project 99/03387-4) and CAPES and of the Biotechnology and Biological Science Research Council is gratefully acknowledged.

References

Bentley, R. (1990). *Crit. Rev. Biochem. Mol. Biol.* **25**, 307–384.
Bottomley, J. R., Hawkins, A. R. & Kleanthous, C. (1996). *Biochem. J.* **319**, 269–278.
Boys, C. W. G., Bury, W. M., Sawyer, L., Moore, J. D., Charles, I. G., Hawkins, A. R., Deka, R., Kleanthous, C. & Coggins, J. R. (1992). *J. Mol. Biol.* **227**, 352–355.
Branden, C. & Tooze, J. (1999). *Introduction to Protein Structure*, 2nd ed. New York/London: Garland Publishing Inc.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–474.
Butler, J. R., Alworth, W. L. & Nugent, M. J. (1974). *J. Am. Chem. Soc.* **96**, 1617–1618.
Charles, I. J., Keyte, J. W., Brammar, W. J., Smith, M. & Hawkins, A. R. (1986). *Nucleic Acids Res.* **14**, 2201–2213.
Chaudhuri, S., Duncan, K., Graham, L. D. & Coggins, J. R. (1991). *Biochem. J.* **275**, 1–6.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Deka, R. K., Anton, I. A., Dunbar, B. & Coggins, J. R. (1994). *FEBS Lett.* **349**, 397–402.
Giles, N. H., Case, M. E., Baum, J. A., Geever, R. F., Huiet, L., Patel, V. B. & Tyler, B. M. (1985). *Microbiol. Rev.* **49**, 338–358.
Gourley, D. G., Coggins, J. R., Isaacs, N. W., Moore, J. D., Charles, I. G. & Hawkins, A. R. (1994). *J. Mol. Biol.* **241**, 488–491.
Gourley, D. G., Shrive, A. K., Polikarpov, I., Krell, T., Coggins, J. R., Hawkins, A. R., Isaacs, N. W. & Sawyer, L. (1999). *Nature Struct. Biol.* **6**, 521–525.

Harris, J. M., Gonzalez-Bello, C., Kleanthous, C., Hawkins, A. R., Coggins, J. R. & Abell, C. (1996). *Biochem. J.* **319**, 333–336.
Haslam, E. (1993). *Shikimic Acid: Metabolism and Metabolites.* Chichester: J. Wiley & Sons.
Jones, S. & Thornton, J. M. (1995). *Prog. Biophys. Mol. Biol.* **63**, 31–65.
Jones, T. A. & Kjeldgaard, M. (1993). *O Version* 5.9, *The Manual.* Uppsala University, Sweden.
Kleanthous, C., Deka, R., Davis, K., Kelly, S. M., Cooper, A., Harding, S. E., Price, N. C., Hawkins, A. R. & Coggins, J. R. (1992). *Biochem. J.* **282**, 687–695.
Kleanthous, C., Reilly, M., Cooper, A., Kelly, S., Price, N. C. & Coggins, J. R. (1991). *J. Biol. Chem.* **266**, 10893–10898.
Kleywegt, G. J. & Jones, T. A. (1997). *Methods Enzymol.* **277**, 208–230.
Kraulis, J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lumsden, J. & Coggins, J. R. (1977). *Biochem. J.* **161**, 599–607.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Navaza, J. & Saludjian, P. (1997). *Methods Enzymol.* **276**, 581–594.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Reilly, A., Morgan, P., Davis, K., Kelly, S. M., Greene, J., Rowe, A. J., Harding, S. E., Price, N. C., Coggins, J. R. & Kleanthous, C. (1994). *J. Biol. Chem.* **269**, 5523–5526.
Servos, S., Chatfield, S., Hone, D., Levine, M., Dimitriadis, G., Pickard, D., Dougan, G., Fairweather, N. & Charles, I. (1991). *J. Gen. Microbiol.* **137**, 147–152.
White, P. J., Young, J., Hunter, I. S., Nimmo, H. G. & Coggins, J. R. (1990). *Biochem. J.* **265**, 735–738.