Crystal structure and mechanism of a carbon–carbon bond hydrolase

David E Timm[1,2]*, Heather A Mueller[1], P Bhanumoorthy[1], Joel M Harp[2] and Gerard J Bunick[2]

Background: Fumarylacetoacetate hydrolase (FAH) catalyzes the final step of tyrosine and phenylalanine catabolism, the hydrolytic cleavage of a carbon–carbon bond in fumarylacetoacetate, to yield fumarate and acetoacetate. FAH has no known sequence homologs and functions by an unknown mechanism. Carbon–carbon hydrolysis reactions are essential for the human metabolism of aromatic amino acids. FAH deficiency causes the fatal metabolic disease hereditary tyrosinemia type I. Carbon–carbon bond hydrolysis is also important in the microbial metabolism of aromatic compounds as part of the global carbon cycle.

Results: The FAH crystal structure has been determined by rapid, automated analysis of multiwavelength anomalous diffraction data. The FAH polypeptide folds into a 120-residue N-terminal domain and a 300-residue C-terminal domain. The C-terminal domain defines an unusual β-strand topology and a novel 'mixed β-sandwich roll' structure. The structure of FAH complexed with its physiological products was also determined. This structure reveals fumarate binding near the entrance to the active site and acetoacetate binding to an octahedrally coordinated calcium ion located in close proximity to a Glu–His dyad.

Conclusions: FAH represents the first structure of a hydrolase that acts specifically on carbon–carbon bonds. FAH also defines a new class of metalloenzymes characterized by a unique α/β fold. A mechanism involving a Glu–His–water catalytic triad is suggested based on structural observations, sequence conservation and mutational analysis. The histidine imidazole group is proposed to function as a general base. The Ca^{2+} is proposed to function in binding substrate, activating the nucleophile and stabilizing a carbanion leaving group. An oxyanion hole formed from sidechains is proposed to stabilize a tetrahedral alkoxide transition state. The proton transferred to the carbanion leaving group is proposed to originate from a lysine sidechain. The results also reveal the molecular basis for mutations causing the hereditary tyrosinemia type I.

Addresses: [1]Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, 635 Barnhill Drive, Indianapolis, Indiana 46202, USA and [2]University of Tennessee/Oak Ridge Graduate School of Biomedical Sciences, and Life Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA.

*Corresponding author.
E-mail: dtimm@iupui.edu

Introduction

Hydrolytic reactions are common metabolic processes. The structures and mechanisms of numerous enzymes that hydrolyze peptide, ester and anhydride linkages in biological molecules have been characterized. The hydrolytic cleavage of less reactive carbon–carbon bonds is also essential to normal metabolic functions and is particularly important to the metabolic degradation of the aromatic amino acids. Indeed, the loss of the fumarylacetoacetate hydrolase (FAH) activity required for phenylalanine and tyrosine degradation is lethal to humans [1,2]. Several compounds that accumulate upstream of FAH are toxic alkylating agents [3] that can induce apoptosis [4]. Thus, FAH deficiency causes the fatal metabolic disease hereditary tyrosinemia type I (HT1), an autosomal recessive disorder that leads to liver failure, liver cirrhosis, liver cancer, renal tubular damage and neurological crises [1,2].

FAH (EC 3.7.1.2) is a homodimer of 46 kDa subunits that shares no apparent sequence homology with any protein described to date [5]. In addition to its physiological substrate, FAH shows a range of specificity in catalyzing the hydrolytic cleavage of carbon–carbon bonds in a variety of diketoacids [5,6]. However, aside from chemical-modification studies suggesting the presence of an active-site cysteine [7], little is known about the enzymatic mechanism of FAH. Nine other enzymes of related catalytic activities have been classified by the Enzyme Commission. These enzymes hydrolyze carbon–carbon bonds in pathways that metabolize the aromatic compounds biphenyl, phloretin, phenylalanine, tyrosine, tryptophan and toluene. Carbon–carbon bond hydrolases also act on oxaloacetate [8] and related acylpyruvates in microbial metabolism. Thus, interest in these enzymes also arises from the involvement of this unique chemistry in the

global carbon cycle and from the potential use of these catalysts in bioremediation [9]. Despite the metabolic importance and novel chemistry of carbon–carbon bond hydrolysis, the structural basis for the hydrolytic cleavage of carbon–carbon bonds has not been studied in detail.

Results and discussion
Structure determination
In order to understand the molecular basis of HT1 and the enzymatic mechanism of FAH, the crystal structure of FAH was determined using the multiwavelength anomalous diffraction method [10]. The FAH structure was determined within 36 hours of data collection using automated Patterson and difference Fourier methods implemented by the program SOLVE (T Terwilliger), which successfully identified 30 of 32 selenium atoms present in the asymmetric unit. The experimental 1.9 Å electron-density map is of excellent quality (Figure 1) and has been interpreted with a model that includes 835 amino acid residues, 604 water molecules, 4 acetate molecules, 2 calcium ions and a nickel ion. A single round of refinement yielded an R factor of 0.208 (R_{free} = 0.233). This result highlights the current potential for rapid, high-quality macromolecular crystal structure determinations using synchrotron data, even when relatively large numbers of anomalous scatters are present in the asymmetric unit.

New protein fold
Consistent with a unique amino acid sequence and catalytic function, the FAH structure diverges within the general class of proteins containing α and β secondary
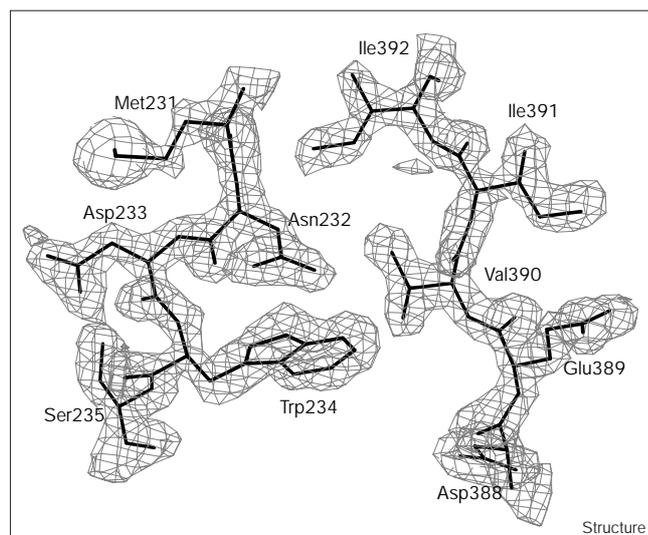
Figure 1

A section of the experimental solvent-flattened electron-density map calculated using data between 8 and 1.9 Å and contoured at 1.2σ. Figures 1, 4a and 6a were generated using the program O [38].

structures to define a ninth regular fold for α–β proteins. The FAH structure contains two α–β domains, a 120-residue N-terminal domain and a 300-residue C-terminal domain (Figure 2a). The C-terminal domain is described by a novel arrangement of β strands and a new three-dimensional domain structure. The polypeptide passes five times between sheet A and sheet B (Figure 2b), resulting in a spatial arrangement of β strands that is unique to FAH (i.e. strand 13 parallel to strand 1 and antiparallel to strand 7). The central features of the C-terminal domain involve the tightly curved mixed β sheet A, which rolls into a barrel-like structure that binds a calcium ion and makes many of the intermolecular interactions in the FAH dimer interface (Figures 2–4). This β roll lacks hydrogen bonds between its opposite edges, but is closed by a type IV β turn between residues 162 and 165. The mixed β roll is sandwiched between β sheet B and an adjacent helical region, resulting in a fold that contains three layers of β structure and a layer of α structure across the midsection of the roll (Figure 2). While two- (β–β, α–β), three- (β–β–α, α–β–α) and four-layered (α–β–β–α) sandwich structures represent the majority of observed α–β class protein folds [11,12], the FAH C-terminal domain represents the first instance in which a β sandwich and a β roll are combined to form a single β–β–β–α layered domain. Automated searches of the Protein Data Bank using DALI [13] and TOP [14] have failed to identify structural homologs of the FAH C-terminal domain.

The FAH active site
The newly described FAH mixed β-sandwich roll structure has functional roles in metal-ion binding, catalysis and dimerization. The FAH structure unexpectedly contained a Ca^{2+} that is octahedrally coordinated by carboxyl oxygens from Asp126, Glu199, Glu201, Asp233, an acetate molecule and a water bound to the sidechain oxygen of Thr350. The Ca^{2+} is located at one end of the mixed β-roll structure (Figure 2a) at the bottom of a 15 Å deep pocket. The carboxyl ligands to the Ca^{2+} originate from strands 1, 7 and 8 of the β roll. The pocket is formed primarily by helices between strands 1 and 2, and 8 and 9 of the β roll. A second acetate ion bound to Arg142 is located at the mouth of the pocket at a distance from the Ca^{2+}-bound acetate that is similar to the distance between carboxyl groups in fumarylacetoacetate. Thus, the pocket is complementary in shape, volume and charge to the physiological substrate fumarylacetoacetate. This observation is the first indication that FAH functions as a metalloenzyme.

The location of the FAH active site in this deep, charged pocket has been confirmed by determining the structure of FAH complexed with the physiological products fumarate and acetoacetate (Figure 3) and by mutating a carboxyl ligand of the Ca^{2+}. The 21 residues present in the FAH active site are absolutely conserved between the mammalian and fungal forms of the enzyme, which share

Figure 2

FAH structure, topology and HT1-associated mutations. **(a)** A stereo ribbon diagram illustrating the FAH subunit structure and position of point mutations causing hereditary tyrosinemia type I is shown. The N-terminal domain is located at the bottom of the figure. The mixed β-sandwich roll structure is centrally located in the figure. Helices are colored red; β strands are colored in shades of blue corresponding to the β sheet they form; the positions of point mutations are represented by green spheres; a calcium ion is colored yellow; acetate carbon and oxygen atoms are respectively colored orange and red (top of figure). **(b)** A topology diagram of the novel FAH β-strand arrangement is shown. β Strands are numbered in red according to their sequential occurrence in the polypeptide chain; residue numbering is in black. Sheets A, B and C are respectively colored in dark, light and medium shades of blue, as in (a). α Helices are represented by red rectangles. Figures 2a, 3, 4b and 6b were generated using MOLSCRIPT [40].
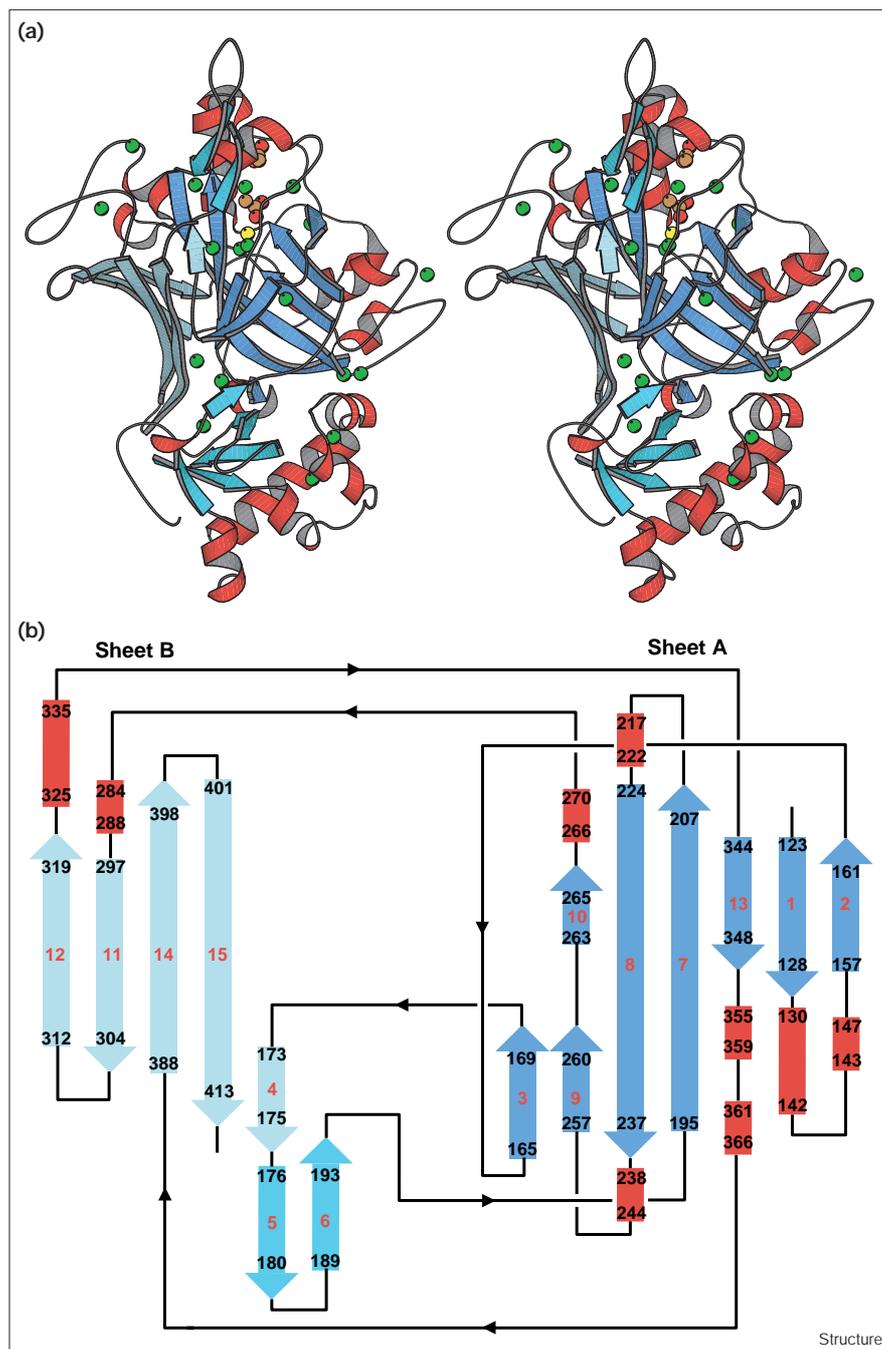
an overall amino acid sequence identity of 47% (data not shown). The replacement of the Glu201 ligand with a glycine residue causes a phenotype like HT1 in mice [15]. FAH containing this mutation has been expressed and lacks catalytic activity (data not shown). At least half of the point mutations associated with HT1 [16] are expected to directly affect the active site. Nine of 18 human HT1-associated point mutations also occur in the

vicinity of the β roll and active-site pocket (Figure 2a), including a D233V mutation that affects another Ca^{2+} ligand. The mutations A134D, G158D, V166G, C193R, G207D, W234G, P249T and G369V also occur near the active site and/or the β roll, and are expected to directly affect the active-site geometry. Five mutations, N16I, G158D, P261L, P342L and S405H occur between the N- and C-terminal domains and may disrupt activity by

Figure 3

A stereo ribbon diagram illustrating the subunit arrangement of the FAH dimer is viewed as the front of a horse-shoe. The domain arrangement and color scheme are as in Figure 2. A green asterisk marks the active site.

causing misfolding. The mutations G158D, P249T and G337S occur in the FAH dimer interface and may affect the active site by perturbing the FAH quaternary structure (Figure 3, see below). A number of mutations, including N16I, F62C, A134D, V166G, C193R, G207D, W234G, P249T, P261L, T294P, P342L and G369V may also cause general structural effects or misfolding by disrupting the hydrophobic core, the native mainchain conformation and/or important electrostatic interactions.

The FAH–product complex structure was determined by crystal-soaking experiments and has been refined at 1.9 Å

to an R factor of 0.153 (R_{free} = 0.195). The acetoacetate displaces the Ca^{2+}-bound acetate and water molecules to maintain a slightly perturbed octahedral coordination with Ca^{2+} through the carbonyl oxygen and one carboxyl oxygen (Figure 4). The second carboxyl oxygen serves as a hydrogen-bond acceptor for the mainchain N of Thr350 at a distance of 2.9 Å. The acetoacetate methyl group packs against the sidechains of Tyr159 and Pro249. Fumarate displaces the acetate ion bound to Arg142. Additional hydrogen bonds are also possible with the hydroxyl groups of Tyr128 and Tyr244 at the mouth of the active site and the His133 and Gln240 NE2 atoms located

Figure 4

Product binding to FAH. **(a)** An omit map illustrates the positioning of fumarate (top) and acetoacetate (bottom) in the FAH active site. The $2F_o-F_c$ (dark blue) and F_o-F_c (light blue) calculated at 1.90 Å are respectively contoured at 1.75 and 4.00σ. Carbon, nitrogen, oxygen and Ca^{2+} are colored orange, blue, red and yellow, respectively. The orientation is similar to the panel **(b)** stereo diagram of the fumarate/acetoacetate-bound active site. Numbers are given for the interatomic distances (in Å) indicated by dashed lines.

midway in the active site. Hydrophobic and/or pi interactions could also occur between fumarate and the sidechains of Tyr128, Val137, Phe141 and Tyr 244. The active site is also in close proximity to and is partially formed by the dimer interface (Figure 3).

A Glu–His–water catalytic triad

The crystal structures described above suggest a catalytic mechanism that involves substrate binding to Ca^{2+}. Other observations indicate that Ca^{2+} also functions in optimizing the orientation of a nucleophilic water molecule and in stabilizing an acetoacetate carbanion leaving group. Consistent with the pH dependence of FAH activity previously attributed to an imidazole group with an ionization constant of 6.6, the obvious general base present in the FAH active site is the His133 sidechain. The His133 ND1 atom is located at a distance of 2.8 Å from the Glu364 carboxyl group. The imidazole NE2 atom is located 2.7 Å from a water molecule bound to the OE2 atom of Glu199, making a 105° angle between the water and the two hydrogen-bond acceptors. Thus, the Ca^{2+} optimizes the Glu199 carboxyl group geometry and electron distribution to function as a hydrogen-bond acceptor that positions a nucleophilic water molecule. This water molecule should be oriented with its lone electron pairs facing the direction of the substrate's electrophilic carbonyl group and with the second hydrogen atom optimally oriented for abstraction by the His133 base stabilized by the Glu364 carboxyl (Figure 5a). No cysteine residues are present in the FAH active site, and the only cysteines conserved between fungal and mammalian forms of the enzyme are buried.

The formation of a tetrahedral alkoxide transition state (Figure 5b) could be stabilized by an oxyanion hole formed by the sidechains of Arg237, Gln240 and Lys253. Indeed, this arrangement is observed in a third crystal structure containing a tetrahedral cacodylate molecule bound midway between the His133 and Gln240 sidechains (Figure 6). Although cacodylate binding to FAH has not previously been reported, the enzyme is competitively inhibited by many monovalent anions [17]. An acetate ion bound to Arg142 also occurs in this structure in a position similar to that occupied by the fumarate carboxyl group in the FAH–product complex. One of the cacodylate oxygens is located in an almost identical position as the His133/Glu199-bound water; the second oxygen is hydrogen bonded to the Gln240 sidechain and a methyl group is located in an almost identical position to the methyl of acetoacetate (Figure 6b). Thus, this structure is consistent with the nucleophilic role of the His133/Glu199-bound water molecule and with the stabilization of the tetrahedral intermediate by Gln240. The formation of a Ca^{2+}-bound acetoacetate carbanion leaving group could be stabilized by charge neutralization of an enolate oxyanion in resonance with the carbanion (Figure 5c,e). The formation of a carbanion intermediate

also accompanies the carbon–carbon bond hydrolysis catalyzed by kynureninase during tryptophan degradation [18]. The final proton-transfer step is likely to involve abstraction from the Lys253 NZ atom (Figure 5c), as direct transfer from His133 would involve a distance of greater than 5 Å and crossing the path taken by the hydroxide ion. Water-mediated regeneration of the active-site protonation state is likely, given that the average number of nonprotein oxygens present in the active site is about 20 for the three structures described above.

In addition to defining a new protein fold, this work defines a novel active site and enzymatic mechanism. Based on sequence alignments and conserved serine, histidine and aspartate residues, two of the known carbon–carbon bond hydrolases are believed to be serine hydrolases of the α/β hydrolase fold family [19]. The FAH structure clearly differs from this fold, which is defined by an α–β–α layering with helices flanking an eight-stranded parallel β sheet [20]; however, FAH does share features in common with the Ser–His–Asp catalytic triad. The distances and angles made between the proposed FAH nucleophilic water, the His133 imidazole N atoms and the Glu364 carboxyl O atoms are very similar to those of the serine-protease catalytic triad. Thus, the active chemical groups in FAH are spatially arranged in common with a wide variety of hydrolase catalytic triads. The ability of catalytic triads to cleave carbon–carbon bonds is also suggested by a study demonstrating the chymotrypsin-catalyzed hydrolysis of an activated carbonyl–methyl bond [21].

The FAH mechanism is distinguished from typical Asp–His–Ser catalytic triads by the oxyanion hole constituents and by the involvement of a catalytic metal ion. Variations of the catalytic triad have received considerable attention and were recently reviewed [22]. The replacement of the aspartate hydrogen bond acceptor/acid component of the triad with a glutamate residue observed in FAH has also been reported for acetylcholinesterase and a fungal lipase [23]. Sidechain variations of the histidine base and serine nucleophile have also been described [22]. Direct activation of a nucleophilic water by a histidine base is also involved in the hydrolysis of serine protease acylenzyme intermediates, in nucleic acid hydrolysis by DNase I [24] and in phospholipid hydrolysis by phospholipase A_2 [25–28]. The formation of an oxyanion hole using a sidechain amide flanked by basic residues in the FAH structure deviates from the usual mode of oxyanion stabilization by mainchain amide groups. The use of positively charged residues in the FAH oxyanion hole may provide even greater stabilization of the charge acquired during nucleophilic attack than amide groups alone would provide. Finally, the involvement of an octahedrally coordinated Ca^{2+} in catalysis is another deviation from the prototypic catalytic triad. Ca^{2+} is commonly associated with a structural role,

Figure 5
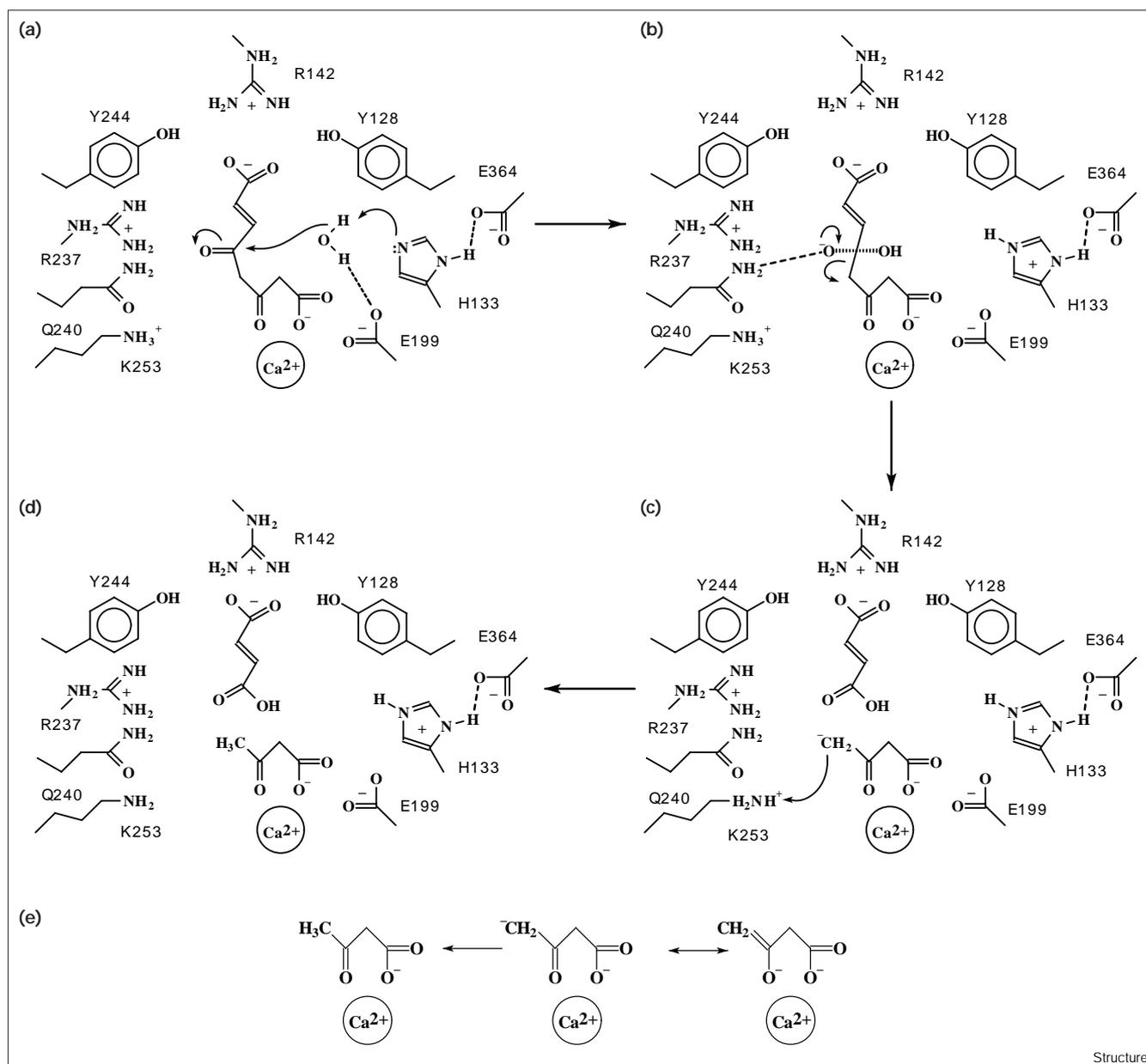
Schematic diagram illustrating the proposed catalytic mechanism for FAH. **(a)** The fumarylacetoacetate substrate is shown bound to Ca²⁺ and to Arg142. A nucleophilic water molecule is optimally positioned for attack on the substrate carbonyl carbon atom by hydrogen bonding to the His133 imidazole group and the carboxyl group of the Ca²⁺ ligand, Glu199. A proton is abstracted by the Glu364-stabilized His133 sidechain. Ca²⁺ participates indirectly in positioning the nucleophile. The flow of electrons is indicated by arrows. **(b)** The formation of a

tetrahedral alkoxide transition state is stabilized by the Gln240, Lys253 and Arg237 sidechains (see Figure 6). The flow of electrons leading to **(c)** the formation of the acetoacetyl carbanion leaving group is indicated by arrows. Proton transfer to the carbanion is proposed to originate from the Lys253 sidechain. **(d)** A schematic of the FAH enzyme–product complex is shown (see Figure 4). **(e)** Formation of the acetoacetate carbanion leaving group is stabilized by electron resonance with an enolate oxyanion, with charge neutralization by the Ca²⁺.

but is less frequently associated with catalytic functions. Ca²⁺ is proposed to participate in FAH catalysis directly, in substrate binding and in stabilizing a carbanion/enolate leaving group (Figure 5). Multiple catalytic roles, including substrate binding and stabilization of a negative

charge acquired by an oxygen atom, have also been attributed to the catalytic Ca²⁺ of phospholipase A₂ [27,28] and phospholipase C [29]. While a metal ion carboxylate ligand may act as a general base in inositol monophosphatase [30], the FAH His133/Glu364 dyad is the more

Figure 6
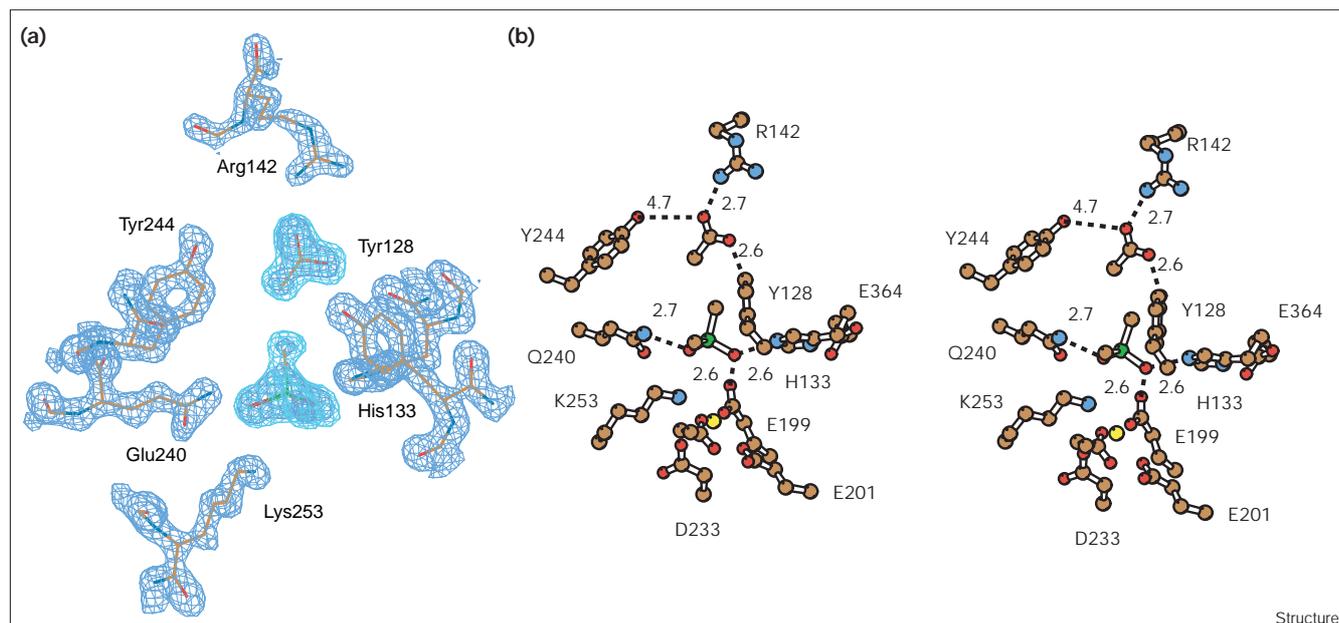
Cacodylate binding to FAH. **(a)** An omit map illustrates the positioning of cacodylate (top) and acetate (bottom) in the FAH active site. The $2F_o–F_c$ (dark blue) and $F_o–F_c$ (light blue) calculated at 1.55 Å are respectively contoured at 1.75 and 4.00σ. Carbon, nitrogen, oxygen, arsenic and Ca^{2+} are respectively colored orange, blue, red, green and yellow. The orientation is similar to the panel **(b)** stereo diagram of the cacodylate/acetate bound active site. Numbers are given for the interatomic distances (in Å) indicated by dashed lines.

likely general base, with the Ca^{2+} functioning indirectly in the orientation of the nucleophilic water molecule through the Glu199 carboxylate ligand. The bond cleaved by FAH is of significantly lower reactivity than the bonds cleaved by most other hydrolases. The combined effects of the metal ion and the positively charged oxyanion hole may account for the ability of FAH to hydrolyze the bond between two carbon atoms.

The structures and mechanism described here also provide explanations for the isomeric specificity of FAH and for the reduced catalytic efficiency of FAH for other diketoacid substrates, and suggest strategies for developing FAH inhibitors. The placement of Tyr128, Arg142 and Tyr244 in the fumarate-binding site only accomodates the extended *trans* fumarylacetoacetate isomer of the fumaryl carboxyl group. The *cis* carboxyl of the immediate precursor on the tyrosine catabolic pathway, malelylacetoacetate, could not sterically fit into the FAH active site. Mahuran *et al.* [5] studied the FAH-catalyzed hydrolysis of six 2,4-diketoacids and three 3,5-diketoacids. The K_M values were measured in the 1.2–8.9 mM range, which is about 1000-fold greater than the K_M of the physiological substrate. The V_{max} values for these substrates vary about 120-fold. The positioning of the fumarate-binding site and the His133 and Gln240 sidechains relative to the Ca^{2+} readily accounts for the reduced activity towards these substrates. Although imperfect octahedral coordination is

observed in all the FAH structures refined to date, it is anticipated that the distortion would be even more pronounced by the five-membered ring formed by 2,4-diketoacids and Ca^{2+}, as compared to the six-membered ring formed by acetoacetate and the Ca^{2+}. Thus, reduced binding affinity and misorientation of the scissile carbonyl carbon relative to His133 and Gln240 would be expected for 2,4-diketoacids. Less than optimal scissile-bond positioning and reduced binding affinity would also be expected for substrates that are not complementary to the electrostatic features of the fumarate-binding site. Finally, it has been suggested, based on studies of a mouse model of HT1, that FAH inhibitors might provide an effective selectable marker to overcome the low efficiency and permanence of gene-therapy approaches for treating liver diseases in general [31]. The results described here suggest clear strategies for developing substrate and transition-state analogs that rationally target the FAH active site.

The FAH dimer and N-terminal domain

The mixed β-sandwich roll is also a central feature of the FAH quaternary structure. The FAH subunits associate as a dimer with a horse-shoe shape that is approximately 86 Å wide, 74 Å tall and 56 Å thick (Figure 3). Consistent with the formation of a tightly associated dimer with a dissociation constant of less than 10 nM (data not shown), over 4000 of the 5700 Å^2 solvent-accessible surface area buried in the FAH dimer interface is hydrophobic.

Residues 149–190 represent the largest contiguous sequence of contacts present in the dimer interface. The first three β strands of β sheet A and the residues closing the mixed β roll between strands 2 and 3 are part of this sequence (Figure 2b), as are residues in strand 4 of sheet B and the two strands of sheet C. However, extended β sheets are not formed between subunits, and intersubunit mainchain hydrogen bonds do not occur in the dimer interface. Additional contacts in the dimer interface are made by residues located between strands 1 and 2, 2 and 3, 7 and 8, 8 and 9, 10 and 11, and 12 and 13. It should also be noted that the sidechains of Pro246 and Leu247 originating from one subunit as part of the dimer interface form one side of the fumarate-binding site in the opposite subunit. The acetoacetate methyl group also packs against the Leu247 sidechain. These residues are likely to be influenced by the presence of substrate or product in the active site, as they shift by about 0.8 Å in the structures described here. Thus, the mixed β-sandwich roll structure functions in the association of the FAH dimer and may allow regulation across the dimer interface and cross-talk between the active sites.

Although multiple functions have been attributed to the FAH C-terminal domain, a function for the N-terminal domain is less clear. The N-terminal domain forms a two-layered α–β-sandwich structure containing a curved five-stranded antiparallel β sheet (Figure 2a). Residues 46–106 between strands 3 and 4 form a helical region that cups the convex surface of the β sheet with four short helices protruding as fingers from the 'knuckles' of a longer four-turn helix. The N-terminal domain and the interdomain residues 117–120 close off the bottom end of the C-terminal β roll, with interdomain contacts primarily mediated by strands 1 and 5 and the 'index finger' of the helical region viewed as a left hand. The occurrence of HT1-associated mutations within the N-terminal domain (Figure 2a) and in the interdomain interface suggests a minimal structural requirement, although a regulatory function is conceivable given the contacts to the C-terminal β roll.

Biological implications

Hydrolytic reactions are essential features of biological systems. Indeed, numerous biological processes rely on water to break covalent bonds. Enzymes catalyzing hydrolytic reactions are involved in the metabolism of most biological compounds. Thus, a wide variety of hydrolases has evolved to metabolize proteins, lipids, nucleic acids and carbohydrates. The hydrolytic cleavage of carbon–carbon bonds by fumarylacetoacetate hydrolase (FAH) is essential for human protein metabolism. FAH deficiency is lethal in humans. FAH hydrolyzes a chemical bond that is of significantly lower reactivity than the bonds cleaved by most hydrolases. Carbon–carbon bond hydrolysis is also essential for the

microbial degradation of aromatic compounds and is, therefore, important in the global carbon cycle. The FAH crystal structure and mechanism reported here represent the first analysis of a carbon–carbon bond hydrolase at the molecular level.

FAH is distinguished by an amino acid sequence and three-dimensional structure that are not apparently related to any protein described to date. Only a small fraction of the protein structures required for life are believed to have been identified [11,12]. The FAH structure contains a 300-residue C-terminal domain that binds a Ca^{2+} and forms the active site and dimer interface. This domain combines known structural motifs in a manner that has not previously been reported and, thus, defines a ninth structural class of α–β proteins defined by a β–β–β–α layering of secondary structural elements.

FAH represents a new class of hydrolytic enzymes utilizing a catalytic-triad active site. The catalytic-triad active-site geometry first described for the serine protease family of enzymes involves groups from a serine nucleophile, a histidine base and an aspartate acid. Variations of this active-site geometry with a conserved spatial arrangement of the acid, base and nucleophile groups were recently reviewed [22], highlighting the point that proteins defined by distinct amino acid sequences and three-dimensional structures can readily define a common active-site geometry and catalytic chemistry. FAH employs a glutamate sidechain carboxyl group to stabilize a histidine base that directly activates a nucleophilic water molecule. The direct activation of water by the FAH histidine base is analogous to the hydrolysis of the acylenyzme intermediate formed during serine-protease catalysis and to nucleic acid hydrolysis by DNase I [24] and phospholipid hydrolysis by phospholipase A_2 [25–28]. Thus, the FAH structure described here represents another example of convergent evolution by proteins of diverse amino acid sequences to a common catalytic chemistry.

The FAH-catalyzed reaction is distinguished from other catalytic triad containing hydrolases by the involvement of an enzyme-bound metal ion and an unusual oxyanion hole, which stabilizes a negatively charged tetrahedral transition state. A metal ion functions in FAH catalysis by binding substrate, activating the nucleophile and stabilizing a carbanion leaving group. In contrast to most catalytic-triad active sites that utilize mainchain amide groups for stabilizing the transition-state intermediate, FAH utilizes a glutamine sidechain amide group flanked by positively charged lysine and arginine sidechains. The metal ion and oxyanion hole may account for the ability of FAH to hydrolyze relatively unreactive chemical bonds. Thus, FAH defines a new variation of catalytic-triad chemistry. The distinguishing features of the FAH

structure and mechanism are likely to occur in other proteins yet to be identified.

The results also reveal the molecular basis for the fatal metabolic disease, hereditary tyrosinemia type I (HT1). HT1 is caused by FAH deficiency and results in death in infancy due to liver and kidney failure, liver cancer and neurological crises. Eighteen point mutations associated with HT1 were mapped onto the FAH structure. This analysis revealed prominent clusterings in the vicinity of the active site and the region between the N- and C-terminal domains. Over half of the point mutations are expected to affect the FAH active site, with the D233V mutation associated with HT1 and a E199G mutation causing an HT1-like phenotype in mice directly affecting ligands of the bound metal ion.

Materials and methods
Protein purification, mutagenesis and crystallization
A cDNA fragment encoding mouse FAH [32] was amplified and sub-cloned into the expression vectors pGEX-4T-1 (Pharmacia) and pRSETb (Invitrogen). The FAH was expressed and purified as a glutathione-*S* transferase (GST) fusion protein using a glutathione Sepharose affinity column (Pharmacia). The fusion protein was digested with thrombin, and the FAH was separated from the free GST using a Mono Q column. A polymerase chain reaction (PCR) mutagenesis protocol (Stratagene) was used to generate mouse FAH containing a substitution of Glu201 with a glycine residue (E201G). The E201G FAH was expressed with a poly-His tag using the pRSETb vector and purified using metal-chelate chromatography. The E201G FAH has a circular-dichroism (CD) spectrum that is very similar to the wild type (data not shown), but had no detectable hydrolysis activity during a 20 minute assay using a commercial substrate [6] at a concentration fivefold higher than the K_M determined for the wild-type His-tagged enzyme. Gel-filtration chromatography using a calibrated Superose 12 column (Pharmacia) and analytical ultracentrifugation using a Beckman XLA indicate that the recombinant mouse FAH is dimeric in solution and has an equilibrium dissociation constant of less than 10 nM (data not shown). Selenomethionine (SeMet) substituted FAH was expressed using the methionine auxotrophic strain B834(DE3)pLysS (Novagen) grown in a defined media containing selenomethionine and the other nineteen amino acids at 20 mg/l. Solutions used for preparing the SeMet-substituted FAH were purged with helium. Crystals of SeMet FAH were grown under a nitrogen atmosphere from hanging drops using equal volumes of a 6 mg/ml SeMet FAH solution in 20 mM Tris, 25 mM CaCl$_2$, pH 7.5 and a reservoir

solution containing 17% PEG 8000, 60 mM nickel acetate, 180 mM sodium acetate, 100 mM sodium cacodylate, pH 6.5. The FAH crystals, of space group P2$_1$, grow within four weeks, and contain one FAH dimer per asymmetric unit. The FAH crystals have unit-cell dimensions of $64.32 \times 110.34 \times 67.53$ Å^3, β = 102.38°.

Structure determination
The FAH structure was determined by the multiwavelength anomalous diffraction method [10] using a single SeMet-substituted crystal with data collected at the Brookhaven National Laboratory NSLS beamline X12C. The crystal was successfully flash-cooled in gaseous nitrogen at 100K using 30% PEG 400, 300 mM sodium acetate, 100 mM cacodylate, pH 6.5. The flash-cooled crystal initially diffracted beyond 2.0 Å resolution using a home source and detector but developed ice during cryogenic storage or transportation. The initial image taken at X12C showed diffraction to less than 6 Å resolution and strong ice rings. The crystal was rescued using the macromolecular crystal annealing method [33] by placing the crystal back into cryoprotectant solution at room temperature for 30 seconds and then flash cooling a second time. The annealed crystal showed no signs of ice and diffracted to 1.7 Å resolution. The experiment was completed by taking data at wavelengths corresponding to the inflection point and peak of the experimentally determined selenium K edge and at a remote high-energy wavelength (Table 1). Data at each wavelength were collected over a total 360° oscillation taken in 90° wedges using a Friedel flip strategy. The diffraction data were integrated using DENZO and SCALEPACK [34] and were local scaled using LOCAL (I Tickle, Birkbeck College) and SOLVE [35] (T Terwilliger, Los Alamos National Laboratory).

A Patterson function calculated using the anomalous differences at the peak wavelength showed 20 peaks above 5 rms on the Harker section (V = 0.5). The partial structure of 30 of the 32 Se atoms present in the asymmetric unit was determined, and an interpretable 2.1 Å experimental electron-density map was calculated in 36 hours using SOLVE [35]. The positions and occupancies of the 30 Se sites were further refined, and phases were calculated to 1.9 Å using MLPHARE [36], treating the data as a special case of the multiple isomorphous replacement method [37]. The mean figure of merit was 0.6773 for 52,858 reflections phased between 16.0 and 1.9 Å resolution, with a mean phasing power of 1.75. Further improvements in the phase estimates and map quality were achieved by solvent flattening and histogram matching using DM [36]. The FAH model, consisting of 835 residues, 604 water molecules, 4 acetate molecules, 2 calcium ions and a nickel ion, was built into a clear 1.9 Å map using O [38]. The nickel ion makes a lattice contact at the N terminus of one subunit. The position of two N-terminal residues and the three C-terminal residues of one subunit and the two C-terminal residues of the second subunit are not apparent in the experimental electron-density map. The starting model yielded an initial R factor of 0.308 and an initial R$_{free}$ of 0.305. Refinement of the FAH model against data to 1.9 Å using X-PLOR v 3.1 [39] with a bulk-solvent correction

Table 1

Crystallographic data collection statistics.

Identifier	Energy (keV)	Observations (total/unique)	Completeness (%)	R$_{sym}$	I/SigI	Number of sites
SeMet	12.644	419,511/136,433	96.6 (89.9)	0.052 (0.157)	19.6 (5.7)	30
SeMet	12.648	404,834/138,458	96.2 (89.2)	0.048 (0.147)	20.2 (6.0)	30
SeMet	13.300	427,158/135,817	97.1 (90.9)	0.053 (0.166)	19.3 (5.6)	30
Product	NA	148,826/67,124	91.4 (60.9)	0.067 (0.264)	13.5 (3.5)	NA
Cacodylate	NA	471,670/118,189	92.5 (47.9)	0.048 (0.209)	27.0 (3.2)	NA

Energy in keV ≈ 12.398/λ in Å. The data were integrated and reduced using DENZO and SCALEPACK. Bijvoet pairs are kept separate in the SeMet scaling. $R_{sym} = \Sigma \,|\, I - < I > \,|\, / \,\Sigma\, I$, where I is the integrated intensity of a given reflection. Numbers in parentheses represent the values obtained in the highest-resolution bin. NA, not applicable.

Table 2

Model refinement statistics.

Identifier	R (R$_{free}$) (%)	Resolution (Å)	B factor overall (Å^2)	Rmsd from ideal values		Ramachandran plot	
				bonds (Å)	angles (º)	favored (%)	allowed (%)
SeMet	20.8 (23.3)	22.0–1.90	12.4	0.007	1.700	91.5	8.5
Product	15.3 (19.5)	27.9–1.90	23.3	0.014	2.412	92.3	7.7
Cacodylate	16.5 (20.8)	29.1–1.55	19.1	0.011	2.091	92.4	7.6

The refinement residual, $R = \Sigma \, | \, F_{obs} - F_{calc} \, | \, / \, \Sigma \, F_{obs}$. The refinements respectively used 67,417, 63,703 and 112,215 mean structure-factor amplitudes for the SeMet, product and cacodylate structures; R$_{free}$ values were respectively calculated from 3,582, 3,399 and 5,952 test reflections for the SeMet, product and cacodylate structures.

and consisting of an overall β-factor refinement, 160 cycles of positional refinement, restrained individual β-factor refinement and a final 76 cycles of positional refinement resulted in an R factor of 0.208 and an R$_{free}$ of 0.233. Stereochemical details are given in Table 2. The placement of Ca^{2+} and Ni^{2+} ions is based on refined temperature-factor gradients and difference Fourier calculations. The temperature factor for a Ni^{2+} included in the active site refines to 16.4 Å^2 with an associated negative F$_o$–F$_c$ peak. This difference Fourier peak is absent when Ca^{2+} is included in the refinement. The Ca^{2+} temperature factor refines to 9.5 Å^2, which is more consistent with the oxygen ligands having refined values of 4.2–10.5 Å^2. Although a partially occupied Ni^{2+} would refine to a lower temperature factor, the metal-ion ligands are well ordered, suggesting that the fully occupied Ca^{2+} is more likely.

The structure of the FAH–product complex was determined using data collected at room temperature from a crystal soaked in precipitant solution containing 100 mM fumarate and 100 mM acetoacetate for 20 hours. Native FAH crystals are grown using a precipitant solution containing 18% PEG 8000, 150 mM nickel acetate, 100 mM sodium cacodylate, pH 6.5. The positions of the two products were clearly defined in both the 2F$_o$–F$_c$ and F$_o$–F$_c$ maps calculated using data collected from the soaked crystals and phases from an FAH model containing only protein atoms. The FAH–product complex model containing 835 amino acid residues, 539 waters, 2 fumarates, 2 acetoacetates, 2 Ca^{2+} and 1 Ni^{2+} was refined at 1.9 Å using REFMAC [36] and X-PLOR [39] to an R factor of 0.153 (R$_{free}$ = 0.195). The average B factors for the fumarate and acetoacetate molecules at full occupancy respectively refined to 25.3 Å^2 and 15.5 Å^2.

The structure of the FAH–cacodylate complex was determined using data collected from a crystal dialyzed for 24 hours against a 20 ml volume of 18% PEG 8000, 300 mM sodium acetate, 100 mM sodium cacodylate, pH 6.5. This experiment was intended to remove divalent cations from the crystal and resulted in the loss of the N-terminally bound Ni^{2+} and a shift in unit-cell parameters to 64.0 × 109.0 × 65.2 Å^3, β = 95.7°. Data were collected at 100K from the crystal on Brookhaven beamline X12C. The active-site cacodylate molecules were unexpectedly observed in both the 2F$_o$–F$_c$ and F$_o$–F$_c$ maps calculated using data collected from the dialyzed crystal and phases from an FAH model containing only protein atoms. A model including 836 amino acid residues, 1049 waters, 2 cacodylates, 2 acetates and 2 Ca^{2+} has been refined at 1.55 Å using REFMAC [36] and X-PLOR [39] to an R factor of 0.165 (R$_{free}$ = 0.208). The average B factor for the cacodylate molecules refined at full occupancy is 27.4 Å^2. The dialysis did not affect the Ca^{2+} occupancy in the active site. The placement of Ca^{2+} and Ni^{2+} ions in all three models is consistent with the refined temperature-factor gradients and difference Fourier calculations.

Accession numbers
Coordinates have been deposited in the Protein Data Bank under the accession codes 1QCN, 1QCO and 1QQJ.

References
1. Lindblad, B., Lindstedt, S., & Steen, G. (1977). On the enzymic defects in hereditary tyrosinemia. *Proc. Natl Acad. Sci. USA* **74**, 4641-4645.
2. Tanguay, R.M., Jorquera, R., Poudrier, J., & St-Louis, M. (1996). Tyrosine and its catabolites: from disease to cancer. *Acta Biochim. Pol.* **43**, 209-216.
3. Jorquera, R. & Tanguay, R.M. (1997). The mutagenicity of the tyrosine metabolite, fumarylacetoacetate, is enhanced by glutathione depletion. *Biochem. Biophys. Res. Commun.* **232**, 42-48.
4. Endo, F., *et al.* & Matsuda, I. (1997). Complete rescue of lethal albino c14CoS mice by null mutation of 4-hydroxyphenylpyruvate dioxygenase and induction of apoptosis of hepatocytes in these mice by *in vivo* retrieval of the tyrosine catabolic pathway. *J. Biol. Chem.* **272**, 24426-24432.
5. Mahuran, D.J., Angus, R.H., Braun, C.V., Sim, S.S., & Schmidt, D.E. (1977). Characterization and substrate specificity of fumarylacetoacetate fumarylhydrolase. *Can. J. Biochem.* **55**, 1-8.
6. Hsiang, H.H., Sim, S.S., Mahuran, D.J. & Schmidt, D.E. (1972). Purification and properties of a diketo acid hydrolase from beef liver. *Biochemistry* **11**, 2098-2102.
7. Nagainis, M.P., Pu, W., Cheng, B., Taylor, K.E. & Schmidt, D.E. (1981). Effects of pH and sulfhydryl specific reagents on 4-fumarylacetoacetate fumarylhydrolase. *Biochim. Biophys. Acta* **657**, 203-211.
8. Lenz H., Wunderwald P., Eggerer H. (1976). Partial purification and some properties of oxalacetase from *Aspergillus niger. Eur. J. Biochem.* **65**, 225-236.
9. Timmis K.N., Steffan R.J., Unterman R. (1994). Designing microorganisms for the treatment of toxic wastes. *Annu. Rev. Mirobiol.* **48**, 525-557.
10. Yang, W., Hendrickson, W.A., Crouch, R.J., & Satow, Y. (1990). Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein. *Science* **249**, 1398-1405.
11. Orengo, C.A., Michie, A.D., Jones, S., Swindells, M.B. & Thornton, J.M. (1997). CATH-a hierarchic classification of protein domain structures. *Structure* **5**, 1093-1108.
12. Murzin, A.G., Brenner, S.E., Hubbard, T. & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536-540.
13. Holm, L. & Sander C. (1998). Touring protein fold space with Dali/FSSP. *Nucleic Acids Res.* **26**, 316-319.
14. Lu G. A WWW service system for automatic comparison of protein structures. (1996). *PDB Newsletter* **78**, 10-11.

15. Aponte, J.L., Sega, G.A., Withrow, C.M., Carpenter, D.A., Rinchik, E.M. & Johnson, D.K. (1999). Point mutations in the murine Fah gene: Animal models for the human genetic disorder hereditary tyrosinemia type 1. *Proc Natl Acad Sci USA* **95**, In press.

16. St-Louis, M. & Tanguay, R.M. (1997). Mutations in the fumarylacetoacetate hydrolase gene causing hereditary tyrosinemia type I: overview. *Hum. Mutat.* **9**, 291-299.

17. Braun C.V. & Schmidt D.E. (1973) Inhibition of fumarylacetoacetate fumarylhydrolase by monovalent anions. *Biochemistry* **12**, 4878-4881.

18. Bild G.S & Morris J.C. (1984). Detection of beta-carbanion formation during kynurenine hydrolysis catalyzed by *Pseudomonas marginalsi* kynureninase. *Arch. Biochem. Biophys.* **235**, 41-47.

19. Diaz E. & Timmis K.N. (1995). Identification of functional residues in a 2-hydroxymuconic semialdehyde hydrolase. *J. Biol. Chem.* **270**, 6403-6411.

20. Ollis, D.L., *et al.* & Goldman, A. (1992). The alpha/beta hydrolase fold. *Protein Eng.* **5**, 197-211.

21. Doherty, D.G. (1955). The hydrolysis of carbon-carbon bonds by alpha-chymotrypsin. *J. Biol. Chem.* **77**, 4887-4892.

22. Dodson, G. & Wlodawer, A. (1998). Catalytic triads and their relatives. *Trends Biochem. Sci.* **23**, 347-352.

23. Schrag, J.D., Li, Y.G., Wu, S. & Cygler, M. (1991). Ser-His-Glu triad forms the catalytic site of the lipase from Geotrichum candidum. *Nature* **351**, 761-764.

24. Suck, D. & Oefner, C. (1986). Structure of DNase I at 2.0 Å resolution suggests a mechanism for binding to and cutting DNA. *Science* **321**, 620-625.

25. Dijkstra, B.W., Drenth, J. & Kalk K.H. (1981). Active site and catalytic mechanism of phospholipase A$_2$. *Nature* **289**, 604-606.

26. Thunnissen, M.M.G.M., *et al.* & Verheij, H.M. (1986). X-ray structure of phospholipase A$_2$ complexed with a substrate-derived inhibitor. *Nature* **347**, 689-691.

27. White, S.P., Scott, D.L., Otwinowski, Z., Gelb, M.H. & Sigler, P.B. (1990). Crystal structure of cobra-venom phospholipase A$_2$ in a complex with a transition-state analogue. *Science* **250**, 1560-1563.

28. Scott, D.L., Otwinowski, Z., Gelb, M.H. & Sigler, P.B. (1990). Crystal structure of bee-venom phospholipase A$_2$ in a complex with a transition-state analogue. *Science* **250**, 1563-1566.

29. Essen, L.O., Perisic, O., Katan, M., Wu, Y., Roberts, M.F. & Williams, R.L. (1997). Structural mapping of the catalytic mechanism for a mammalian phosphoinositide-specific phospholipase C. *Biochemistry* **36**, 1704-1718.

30. Bone, R., Frank, L., Springer, J.P. & Atack, J.R. (1994). Structural studies of metal binding by inositol monphosphatase: evidence for two-metal ion catalysis. *Biochemistry* **33**, 9468-9476.

31. Overturf, K., al-Dhalimy, M., Tanguay, R., Brantly, M., Ou, C.N., Finegold, M., & Grompe M. (1996). Hepatocytes corrected by gene therapy are selected *in vivo* in a murine model of hereditary tyrosinaemia type I. *Nat. Genet.* **12**, 266-273.

32. Klebig, M.L., Russell, L.B. & Rinchik, E.M. (1992). Murine fumarylacetoacetate hydrolase (Fah) gene is disrupted by a neonatally lethal albino deletion that defines the hepatocyte-specific developmental regulation 1 (hsdr-1) locus. *Proc. Natl Acad. Sci. USA* **89**, 1363-1367.

33. Harp, J.M., Timm, D.E. & Bunick, G.J. (1998). Macromolecular crystal annealing: overcoming increased mosaicity associated with cryocrystallography. *Acta Crystallogr. D* **54**, 622-628.

34. Otwinowski Z. & Minor W. (1997). Processing of X-ray diffraction data collected in oscillation data reduction mode. *Methods Enzymol.* **276**, 307-326.

35. Terwilliger T.C. (1997). Multiwavelength anomalous diffraction phasing of macromolecular structures: analysis of MAD data as single isomorphous replacement with anomalous scattering data using the MADMRG Program. *Methods Enzymol.* **276**, 530-537.

36. CCP4. (1979). The SERC (UK) Collaborative Computing Project No. 4, A suite of programs for protein crystallography, distributed from Daresbury Laboratory, Warrington WA4 4AD, UK.

37. Ramakrishnan V. & Biou V. (1997). Treatment of multiwavelength anomalous diffraction data as a special case of multiple isomorphous replacement. *Methods Enzymol.* **276**, 538-557.

38. Jones TA, Zou JY, Cowan SW, & Kjeldgaard (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.

39. Brünger, A.T. (1992). *X-PLOR version 3.1. A system for X-ray crystallography and NMR*. New Haven: Yale University Press.

40. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946-950.