Crystal structure of the copper-containing quercetin 2,3-dioxygenase from Aspergillus japonicus

Fusetti, Fabrizia; Schröter, Klaus H.; Steiner, Roberto A.; Noort, Paula I. van; Pijning, Tjaard; Rozeboom, Henriëtte J.; Kalk, Kor H.; Egmond, Maarten R.; Dijkstra, Bauke W.

Published in:
Structure

DOI:
10.1016/S0969-2126(02)00704-9

Publication date:
2002

Structure, Vol. 10, 259–268, February, 2002, ©2002 Elsevier Science Ltd. All rights reserved. PII S0969-2126(02)00704-9

Crystal Structure of the Copper-Containing Quercetin 2,3-Dioxygenase from *Aspergillus japonicus*

Fabrizia Fusetti,[1,4] Klaus H. Schröter,[1,4] Roberto A. Steiner,[1] Paula I. van Noort,[2] Tjaard Pijning,[1] Henriëtte J. Rozeboom,[1] Kor H. Kalk,[1] Maarten R. Egmond,[2] and Bauke W. Dijkstra[1,3]

[1]Laboratory of Biophysical Chemistry
Department of Chemistry
University of Groningen
Nijenborgh 4
9747 AG Groningen
The Netherlands

[2]Unilever Research
Olivier van Noortlaan 120
3133 AT Vlaardingen
The Netherlands

Summary

Quercetin 2,3-dioxygenase is a copper-containing enzyme that catalyzes the insertion of molecular oxygen into polyphenolic flavonols. Dioxygenation catalyzed by iron-containing enzymes has been studied extensively, but dioxygenases employing other metal cofactors are poorly understood. We determined the crystal structure of quercetin 2,3-dioxygenase at 1.6 Å resolution. The enzyme forms homodimers, which are stabilized by an N-linked heptasaccharide at the dimer interface. The mononuclear type 2 copper center displays two distinct geometries: a distorted tetrahedral coordination, formed by His66, His68, His112, and a water molecule, and a distorted trigonal bipyramidal environment, which additionally comprises Glu73. Manual docking of the substrate quercetin into the active site showed that the different geometries of the copper site might be of catalytic importance.

Introduction

Dioxygenases play a key role in the complex degradation pathway of aromatic and heteroaromatic compounds. Most of the enzymes that catalyze dioxygen incorporation in aromatic substrates are nonheme, iron-containing dioxygenases. Among them, the intradiol-type catechol dioxygenases utilize mononuclear Fe^{3+} centers for enzymatic activity, while the extradiol-type enzymes use Fe^{2+} ions [1]. The different redox states of the iron are associated with different reaction mechanisms. Intradiol-type dioxygenases activate the metal-bound substrate, whereas the extradiol-type dioxygenases activate the O_2 bound to the Fe^{2+} ion. Copper- [2, 3], manganese- [4], and magnesium-containing [5] dioxygenases have also been isolated. However, structural knowledge on how dioxygenation is carried out by those noniron centers is totally lacking.

Copper-containing proteins are known to participate in a variety of O_2-processing reactions. In biological systems, copper has only been observed as Cu^{+} or Cu^{2+}. Copper is remarkably flexible regarding its coordination mode, and the variety of copper sites found in proteins reflects this versatility [6, 7]. While Cu^{+} is preferably complexed by cysteine and methionine residues, Cu^{2+} is mostly ligated by histidine, serine, threonine, or tyrosine residues and by H_2O. The preferred coordination geometry is tetrahedral for Cu^{+} ions and square planar for Cu^{2+} ions. Based on their spectroscopic properties, the copper sites have traditionally been classified into three groups. The type 1 copper center is characterized by a strong absorption band near 600 nm. The absorption band is responsible for the intense color of the blue oxidases and of the blue electron transfer proteins. Generally, it is coordinated in a trigonal bipyramidal or a distorted tetrahedral geometry. The type 2 copper center is colorless. It is found in mono- and dioxygenases, nonblue copper oxidases, and Cu,Zn superoxide dismutase. The most common ligand in type 2 copper sites is histidine. The stereochemical requirements are less critical in type 2 copper centers than in type 1, allowing for a larger array of coordination geometries. The type 3 copper, present in hemocyanin and tyrosinase, is mostly associated with redox reactions. It consists of a pair of Cu^{2+} ions and is characterized by a strong absorption band at 300 nm. The reactivity of the copper centers in proteins is influenced by their surrounding environment, which fine-tunes the redox potential of the metal and confers substrate specificity [8].

Quercetin 2,3-dioxygenase (2,3QD; EC 1.13.11.24) is a type 2 copper-dependent enzyme [9] expressed by *Aspergillus* species when grown on complex aromatic compounds, such as rutin or quercetin. It is able to disrupt the O-heteroaromatic ring of flavonols, yielding the corresponding depside (phenolic carboxylic acid ester) and carbon monoxide (Figure 1) [10]. The difficult breakage of two carbon-carbon bonds makes this reaction chemically challenging. Studies on the 2,3QDs from *Aspergillus flavus* [11, 12] and *Aspergillus niger* DSM 821 [9] have shown that the enzyme does not require any additional organic cofactors for catalysis. The *A. flavus* 2,3QD has a molecular mass of 111 kDa, with a sugar content of 27.5% and a metal content of 2 mol of copper per mole of enzyme [11]. The 2,3QD isolated from *A. niger* DSM 821 is composed of three different subunits, with molecular masses of 63–67, 53–57, and 31–35 kDa, respectively. It has a carbohydrate content of 46%–54% and contains 1.0–1.6 mol of copper per mole of enzyme [9]. No information is available on the amino acid sequences of these latter two enzymes, besides their biochemical characterization.

In this report, we present the crystal structure of the copper dioxygenase 2,3QD from *Aspergillus japonicus* determined at 1.6 Å resolution. The structure of 2,3QD, which is the first of a noniron dioxygenase, firmly estab-

[3]Correspondence: bauke@chem.rug.nl
[4]These authors contributed equally to this work.

Key words: dioxygenase; copper; cupin; glycoprotein; X-ray structure; quercetin

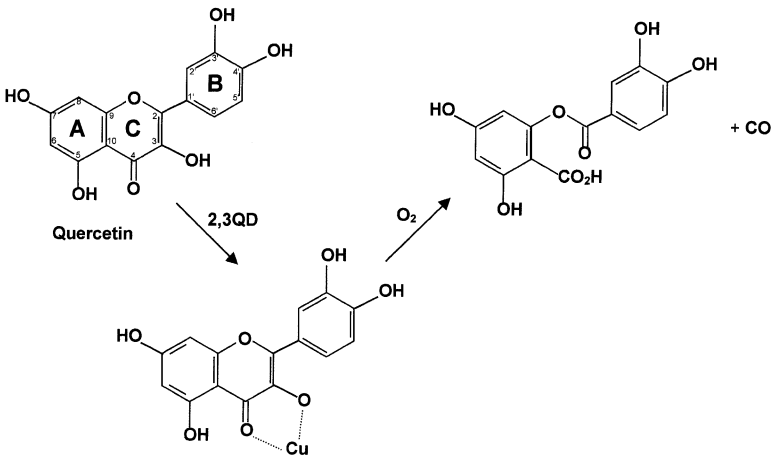


Figure 1. Reaction Catalyzed by Quercetin 2,3-Dioxygenase

The proposed substrate-Cu complex is shown with quercetin (5,7,3′,4′-tetrahydroxyflavonol).

lishes a relationship with members of the cupin superfamily. Docking of the substrate quercetin into the active site provides valuable information about the amino acid residues of potential catalytic importance.

Results and Discussion

A. japonicus quercetin 2,3-dioxygenase (2,3QD) is a glycoprotein, with a molecular mass of about 50 kDa. It is composed of 350 amino acid residues and contains a single copper ion per monomer, as revealed by atomic absorption spectroscopy. A partially deglycosylated form was purified and crystallized in space group $P2_1$, with four molecules in the asymmetric unit. Its crystal structure was determined by multiple isomorphous replacement supplemented with anomalous scattering (MIRAS), using three derivatives and the native copper atom as anomalous scatterers. The final crystallographic and free [13] R factor values, calculated with data between 50.0 and 1.6 Å resolution, are 16.2% and 18.9%, respectively. The four 2,3QD monomers (A, B, C, and D) display a two-domain structure and could be superimposed with an average root mean square (rms) difference of 0.31 Å for 331 Cα atoms. While the C-terminal ends of all polypeptide chains are well defined, the first two N-terminal residues of molecules A and B and the first four of molecules C and D were not included in the model because of lacking electron density. Similarly, a stretch of nine residues (155–163) could not be built in any of the four molecules, leaving the catalytic site solvent exposed. Two *cis* proline residues (Pro129 and Pro195) could be identified in each monomer. A total of 1337 amino acid residues, 4 copper ions, 30 carbohydrate residues, 32 molecules of ethylene glycol from the cryoprotectant, and 1756 water molecules were included in the final model. A summary of data collection, phasing, and refinement statistics is given in Tables 1 and 2.

Quaternary Structure and Glycosylation

At functionally relevant pH values (pH 5.0–7.0), 2,3QD is a homodimer of about 100 kDa, containing ~25% (w/w) of N-linked glycan chains. The structure, at pH 5.2, shows that a homodimer is formed by two monomers interacting almost perpendicularly to each other (Figure 2). In this dimer, the copper centers are separated by 40 Å. Upon dimer formation, an extensive surface of about 2000 $Å^2$ is buried. The dimerization interface is mainly constituted of residues located in loops, and interactions occur via hydrogen bonds, often mediated by solvent molecules. The two dimers in the asymmetric unit are identical within coordinate error and can be superimposed with an rms difference of 0.2 Å for all main chain atoms.

Table 1. Data Processing and Phasing Statistics[a]

Crystal		Resolution (Å)	Completeness (%)	Redundancy	R_{sym}	Number of Sites	R_{cullis}	Phasing Power	Figure of Merit
Native 1		50–2.16	88.6 (97)	5.6	5.7 (13.6)	—	—	—	—
Cu	Anomalous					4	—	0.46	0.13
Native 2		50–1.60	95.0 (90)	9.6	4.1 (38.3)	—	—	—	—
Na_2PtCl_6	Isomorphous	50–2.16	87.0 (96)	3.0	7.3 (11.1)	8	66.6	1.00	0.24
	Anomalous					0	—	—	—
K_2PtCl_4	Isomorphous	50–2.16	79.9 (89)	2.5	7.9 (18.5)	12	61.2	1.18	0.23
	Anomalous					12		0.55	0.24
$KAuBr_4$	Isomorphous	50–2.16	87.0 (95)	2.9	4.9 (10.3)	16	55.8	1.47	0.25
	Anomalous					16		1.05	0.20

Phasing statistics are calculated with data to 2.7 Å resolution. No anomalous signal was detectable for the Na_2PtCl_6 derivative. The overall figure of merit was 0.55. Statistics for the highest resolution shells are indicated in parentheses.

[a] See [44] for the definitions of standard crystallographic quantities.

Table 2. Refinement and Model Statistics

Polypeptide Chain	A	B	C	D	Total
Number of residues	334	339	334	330	1337
Number of copper atoms	1	1	1	1	4
Number of sugar residues					
N-acetylglucosamine	6	5	5	5	21
Mannose	5	2	2	—	9
Number of water molecules					1756
Number of ethylene glycol molecules					32
Number of nonhydrogen atoms					12,599
Resolution (Å)			50–1.6		
R_{cryst} (%)[a]			16.2		
R_{free} (%)[a]			18.9		
Mean B value for all atoms (Å^2)			20.62		
Mean B value for protein atoms (Å^2)			17.95		
Mean B value for copper ions (Å^2)			16.91		
Rms deviation from ideal geometry[b]					
Bond lengths (Å)			0.009		
Bond angles (°)			1.48		

[a] See [46] for the definitions of standard crystallographic quantities. R_{free} [13] was calculated using 10% of the data.
[b] Ideality is defined according to [47].

2,3QD is secreted by *A. japonicus* in the extracellular medium in a glycosylated form. Deglycosylation treatment was necessary to produce diffraction-quality crystals. Endoglycosidase-H, which should cleave the carbohydrate chain after the N-linked N-acetylglucosamine residue, reduced the sugar content to about 3%–5% (w/w) without affecting enzyme activity and quaternary assembly (data not shown). Five potential Asn-X-Thr glycosylation sites were identified in the sequence. The crystal structure confirms N-linked glycosylation at all predicted sites, with one N-acetylglucosamine (GlcNAc) visible at residues Asn90, Asn109, Asn142, and Asn248. Surprisingly, a well-defined branched heptasaccharide was visible at Asn191, which could be modeled as shown in Figure 3. In 2,3QD, three out of five modified asparagine residues are located in β turns, according to the general expectations for glycoproteins [14, 15]. The other two (Asn142 and Asn248) are inserted at the N-terminal ends of β strands 12 and 18, respectively.

In general, attempts to determine the conformation of carbohydrate chains on the surface of glycoproteins have been hampered by the intrinsic flexibility of the oligosaccharides. Therefore, finding a well-defined covalently linked glycan in our structure was very surprising. Besides increasing the solubility of a protein in the extracellular medium, a putative function of N-linked glycosylation is to enhance correct protein folding and assembly [16]. Indeed, the glycan N linked to residue Asn191 could promote stable oligomerization of 2,3QD, as it is located at the dimer interface. Protein-carbohydrate interactions, often mediated by water molecules, contribute to the extraordinary order of the N-linked glycan. The oligosaccharide packs against β turn 185–191 of the other molecule of the dimer. The GlcNAc2 N-acetyl carbonyl oxygen atom is hydrogen bonded to the side chain of Thr189, and the mannose2 O2 atom forms a water-mediated hydrogen bond with the main chain carbonyl group of Pro185. Another water-mediated hydrogen bond is present between the GlcNAc1 O6 atom and the side chain of Gln83 of the other monomer in the dimer. The two N-linked carbohydrates run in opposite directions at the dimer interface and do not interact with each other. In neither of the monomers is the β(1,4) glycosidic bond between GlcNAc1 and GlcNAc2 accessible for cleavage by endoglycosidase-H. This could explain the unsuccessful deglycosylation at the Asn191 site, supporting that this 2,3QD dimer is also formed in solution and is not a crystallization artifact.

Monomer Structure

The 2,3QD monomer has approximate dimensions of 30 × 45 × 50 Å. It is composed of two structurally similar domains positioned face to face around a pseudo 2-fold symmetry axis (Figure 4). The two domains, which are joined by a linker of 60 amino acids (residues 146–205), can be superimposed with an rms difference of 1.6 Å for 120 Cα atoms. The N-terminal domain (residues 1–145) shares about 20% sequence identity with the C-terminal domain (206–350). Apart from a salt bridge between Arg108 and Asp246, the interactions at the domain-domain interface are mainly hydrophobic. As illustrated by the topology diagram (Figure 4), each domain comprises two antiparallel β sheets, with eight strands forming a β sandwich, and two short α helices. In the N-terminal domain, there is an additional β strand (β strand 5), which builds up part of the catalytic site. In both domains, similar hydrophobic cavities are present. In the N-terminal domain, this cavity corresponds to the catalytic site and contains the mononuclear copper center, located at about 10 Å from the protein surface. The C-terminal domain lacks the metal center. Sequence and structural alignments of the two domains show that the copper-coordinating histidine residues (see below) are re-

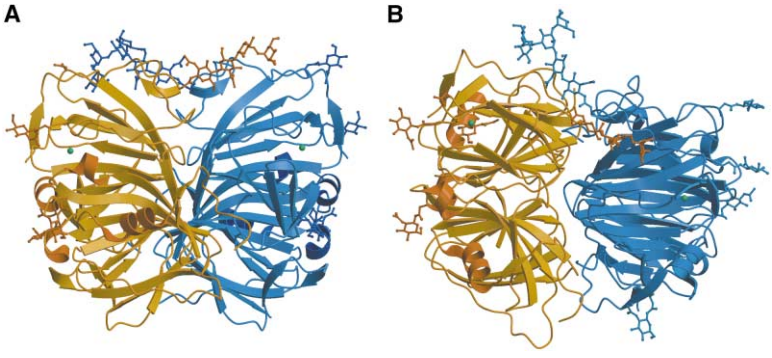


Figure 2. Dimer Structure

(A) Ribbon representation of the 2,3QD dimer. (B) A side view of the dimer, rotated about 45° from (A) about an axis parallel to the 2-fold axis. The monomers are colored blue and gold. The copper atoms are shown as green spheres. The N-linked carbohydrate residues are shown as ball and stick models. This figure was generated using the programs MOLSCRIPT [46] and RASTER3D [47].

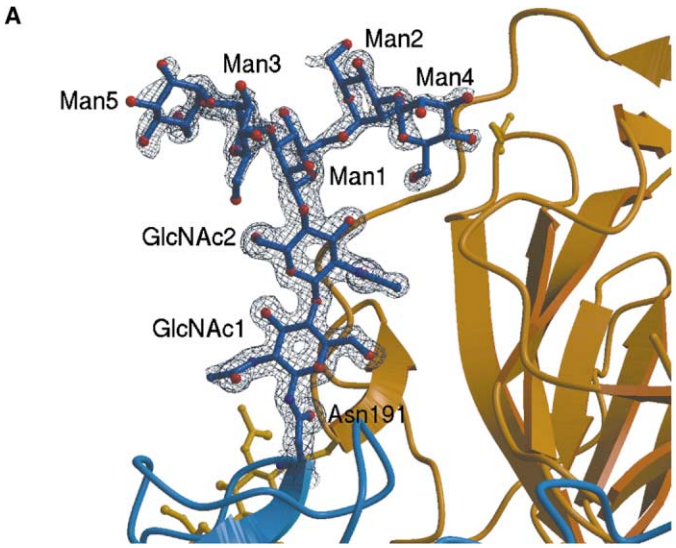
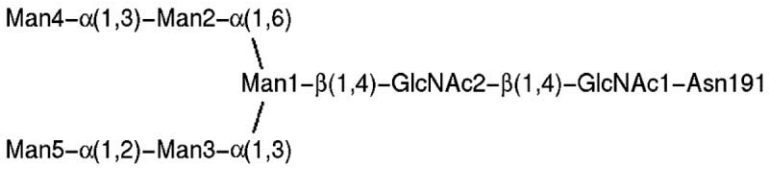


Figure 3. Glycosylation Site at the Dimer Interface of 2,3QD

(A) $2F_o - F_c$ electron density map (contoured at 1 σ) corresponding to the oligosaccharide N linked to Asn191. The oligosaccharide was modeled as shown in (B). The carbohydrates and residue Asn191 are shown in ball and stick representation (GlcNAc, N-acetylglucosamine; Man, Mannose). The figure was prepared with the programs BOBSCRIPT [48] and RASTER3D [47].

B



placed by other aromatic residues in the C-terminal domain (Figure 5).

Sequence and Structure Comparison

Based on its amino acid sequence and three-dimensional structure, 2,3QD can be classified within the cupin superfamily [17]. This superfamily includes functionally diverse proteins that are found in Archaea, Eubacteria, and Eukaryota. It is characterized by two signature sequences, PG-$(X)_5$-HXH-$(X)_4$-E-$(X)_6$-G and G-$(X)_5$-PXG-$(X)_2$-H-$(X)_3$-N, separated by a variable stretch of 15–50 amino acid residues. In 2,3QD, these signature sequences are best conserved in the N-terminal domain. They are separated by 39 amino acids, and 2 amino acids are inserted in the first sequence motif (Figure 5). The conserved histidine and glutamate correspond to the copper-coordinating residues.

Most cupins are only known from gene cloning or sequencing projects, and very few of them have been functionally characterized. Structural information shows that they contain a motif of six antiparallel β strands located within a conserved β barrel structure [17–20]. A structure comparison performed with the program DALI [21] detected high similarity between 2,3QD and the cupin enzymes germin and homogentisate 1,2-dioxygenase (HGO) [20, 22]. Germin is a 16 kDa Mn-containing oxalate oxidase, the structure of which can be superimposed on the N-terminal domain of 2,3QD with an rms difference of 1.8 Å for 91 Cα atoms. When the two structures are compared, the Mn site of germin, formed by three histidines, a glutamate residue, and two water molecules, matches with the copper site of 2,3QD. The Fe-containing enzyme HGO is composed of two homologous domains but, in contrast to 2,3QD, contains the metal binding site in the C-terminal domain [22]. The structure of HGO can be superimposed on that of 2,3QD with an rms difference of 2.0 Å for 121 Cα atoms; the metal binding domains superimpose with an rms difference of 1.3 Å for 59 main chain atoms.

The close structural similarities between 2,3QD, HGO, and germin suggest the existence of an evolutionary link. While the origin of copper type 1 proteins can be easily traced, starting from the small blue copper proteins present in chloroplasts and unicellular organisms, the phylogenetic origin of type 2 copper proteins is not yet fully understood [6]. Nevertheless, there are indications that the type 2 copper site in phenylalanine hydroxylase [23] and Cu,Zn superoxide dismutase [24] evolved from primordial Fe- and Mn-containing enzymes after copper became biologically available. The similarity of 2,3QD to the cupin metalloenzymes at the genetic and structural level lends further support to the hypothesis that type 2 copper sites have developed from preexisting metal binding centers.

Copper Coordination

The crystal structure of 2,3QD shows that the enzyme contains only one copper center per monomer, in agreement with results from metal elementary analysis. The metal center is solvent exposed. The copper ion is mainly coordinated by three histidine residues (His66, His68, and His112) and by a water molecule. The geometry is distorted tetrahedral. There is also an additional coordination in which the three histidines and Glu73

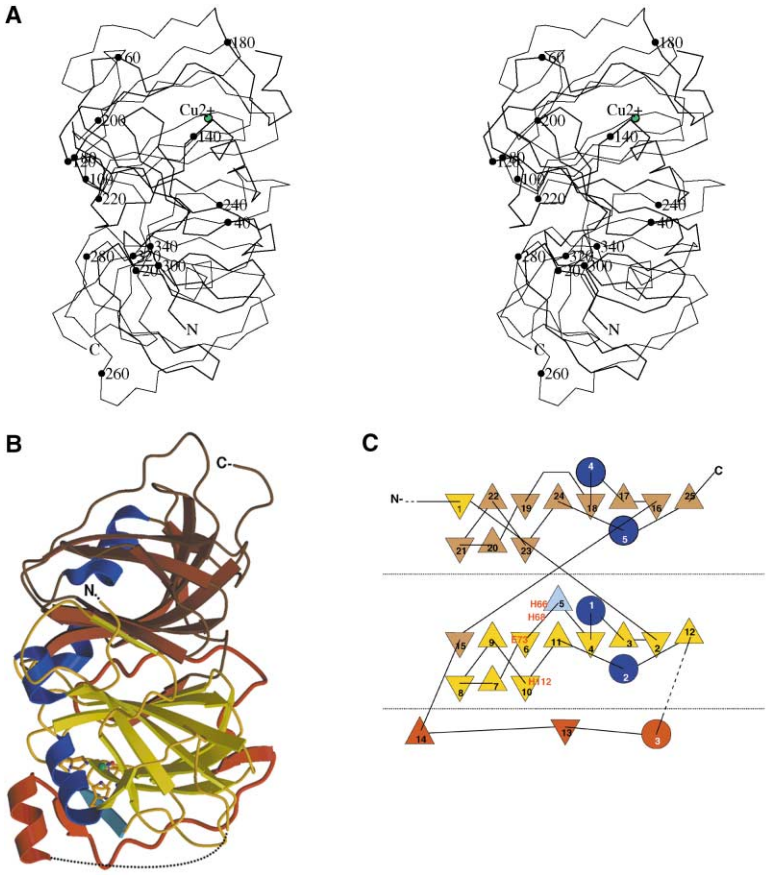


Figure 4. Overview of the 2,3QD Structure

(A) Stereo diagram of the monomer shown as a Cα trace. Every twentieth residue is labeled.

(B) Ribbon diagram of the monomer. The N-terminal domain is shown in yellow and the C-terminal domain in brown. α helices are shown in blue. The linker connecting the two domains is colored red. The copper is shown as a green sphere. The residues coordinating the copper are represented as ball and stick models. These figures were generated using the programs MOLSCRIPT [46] and RASTER3D [47].

(C) Topology diagram of the secondary structure of the 2,3QD monomer generated with the program TOPS [49] and hand edited afterwards. The colors correspond to those in (A). In both figures, the dotted lines indicate the flexible regions of the structure (residues 1–3 and 154–169).

interact with the metal and the water molecule is positioned further away from the copper (Figure 6). The geometry of this secondary coordination, estimated to represent at most 30% of the structure, is trigonal bipyramidal, with His66 and Glu73 as the axial ligands. In the main coordination, Glu73 points away from the metal, and the carboxylic Oϵ2 atom is hydrogen bonded to the coordinating water molecule (Figure 6). In the minor conformation, Glu73 coordinates the copper ion, and the water molecule is displaced to occupy an equatorial position at 2.6 Å from the metal. In both coordination arrangements, the Nϵ2 atoms of the histidine side chains are at a distance of about 2.1 Å from the Cu^{2+} ion and coordinate the metal similarly to what has been found in other type 2 copper sites [8]. The orientation of each histidine ligand is stabilized by a hydrogen bond between the Nϵ1 atom of the imidazole ring and the main chain carbonyl oxygen of residues Leu64 (His66), Asp176 (His68), and Val110 (His112). A detailed description of the copper geometry is given in Table 3.

Electron paramagnetic resonance (EPR) measurements are in agreement with a heterogeneous coordination of the copper ion. Cu^{2+} contains an unpaired electron and can be detected by EPR, whereas Cu^{+} is EPR silent. The X-band EPR spectrum, at pH 6.0, shows a clear double spectral line (major line with EPR parameters $g_{//} = 2.330$ and $A_{//} = 13.7$ mT; minor line with $g_{//} = 2.290$ and $A_{//} = 12.5$ mT), indicating that the Cu^{2+} environment is also heterogeneous in solution (Kooter et al., submitted).

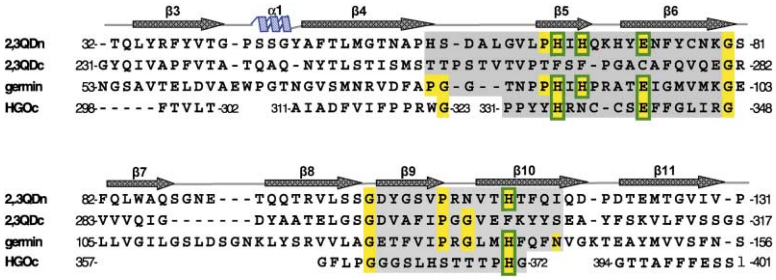


Figure 5. Structure-Based Sequence Alignment of 2,3QD with Members of the Cupin Superfamily

2,3QDn, N-terminal domain of 2,3QD; 2,3QDc, C-terminal domain of 2,3QD; germin, oxalate oxidase germin; HGOc, C-terminal domain of human homogentisate 1,2-dioxygenase. The alignment is based on 3D comparison with 2,3QD. Only the regions adjacent to the cupin signature motifs are shown. The cupin motifs are shaded in gray, and the conserved residues are highlighted in yellow. The metal coordinating residues are in green boxes.

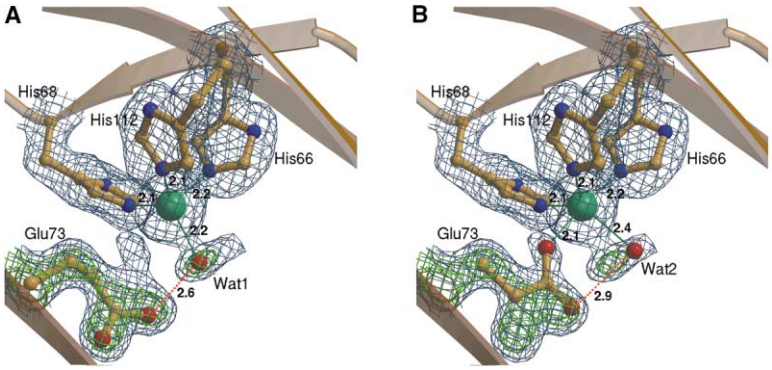


Figure 6. Copper Coordination in 2,3QD

(A) Tetrahedral geometry in which three histidine residues (His66, His68, and His112) and one water molecule (Wat1) are ligated to the copper.

(B) Trigonal bipyramidal geometry in which Glu73 is directly participating in the copper coordination, and the bound water molecule is shifted to a distal position (Wat2). The $2F_o - F_c$ electron density map is contoured at 1σ (blue) and $2\ \sigma$ (green). The atomic distances (in Å) represent the average value calculated from the four monomers in the crystallographic asymmetric unit. The copper is shown as a green sphere. The coordinating residues are represented in ball and stick models. This figure was prepared with the programs BOBSCRIPT [48] and RASTER3D [47].

Careful inspection of the σ_A-weighted $2F_o - F_c$ electron density maps showed that the electron density around the coordinating water molecule was not spherical but had a somewhat elongated shape (Figure 6). The possibility that such electron density belonged to a bound dioxygen molecule was investigated by Raman spectroscopy. However, no resonance at 704 cm^{-1}, indicative of an O-O stretch from a coordinated dioxygen, could be detected. Furthermore, substitution of the coordinating water with a dioxygen molecule during refinement resulted in high B values for the oxygen atoms (>40 $Å^2$) and produced high peaks of negative electron density around the copper. Since the quality of the diffraction data and the refinement statistics were good and most parts of the structure were well defined by the electron density, we conclude that the ellipsoidal shape of the electron density results from alternative positions of the copper-bound water molecule.

Although it is known that metal coordination in macromolecules often exhibits a range of distortions, both copper arrangements seen in the 2,3QD structure are peculiar. A tetrahedral coordination is, in general, not typical for Cu^{2+} ions. Instead, as shown by model compounds, such coordination together with the trigonal planar geometry is preferred by Cu^{+} ions [6, 25]. To exclude that radiation damage or photoreduction of the copper produced the tetrahedral coordination observed here, we have subjected a crystal of 2,3QD to multiple data collections at beamline ID14-EH4 (EMBL Grenoble). No significant changes of the ratio between noncoordinating/coordinating Glu73 were observed for an absorbed dose exceeding 2×10^7 Gy [26] (R.A. Steiner and R.B. Ravelli, unpublished data). Only partial general decarboxylation was detected as a result of radiation damage [27].

Carboxylate ligation has never before been observed in natural copper proteins [6–8, 28]. It has been proposed that, for some iron redox centers, coordination by a carboxylate residue might be responsible for modulating small energy barriers [29, 30]. The structure of 2,3QD shows, for the first time, that a carboxylic amino acid can function as a direct copper ligand in a protein. The two metal coordination spheres visible in the 2,3QD structure may represent specific catalytic states of the enzyme (see below).

Table 3. Copper Coordination Geometry

Bond (Å)/Angle(°)[a]	Tetrahedral	Trigonal Bipyramidal
Cu-Nϵ2 (His66)	2.2	2.2
Cu-Nϵ2 (His68)	2.1	2.1
Cu-Nϵ2 (His112)	2.1	2.1
Cu-O_{wat}	2.2	2.4[b]
Cu-Oϵ1 (Glu73)	—	2.1
Nϵ2 (His66)-Cu-Nϵ2 (His68)	99	99
Nϵ2 (His66)-Cu-Nϵ2 (His112)	95	95
Nϵ2 (His66)-Cu-O_{wat}	109	94
Nϵ2 (His66)-Cu-Oϵ1 (Glu73)	—	174
Nϵ2 (His68)-Cu-Nϵ2 (His112)	114	114
Nϵ2 (His68)-Cu-O_{wat}	111	136[c]
Nϵ2 (His68)-Cu-Oϵ1 (Glu73)	—	86
Nϵ2 (His112)-Cu-O_{wat}	123	108[d]
Nϵ2 (His112)-Cu-Oϵ1 (Glu73)	—	86
Oϵ1 (Glu73)-Cu-O_{wat}	—	80

[a] The indicated distances and angles represent the average values calculated from the four monomers in the asymmetric unit. The standard deviation from the average values is below 0.1 Å for all bond distances, except where noted below. The standard deviation for the coordination angles is in the range of 1°–5°, except where noted below.

[b] Standard deviation = 0.2 Å.

[c] Standard deviation = 12°.

[d] Standard deviation = 10°.

Modeling of the Substrate into the Active Site

2,3QD catalyzes the dioxygenation of quercetin (5,7,3′,4′-tetrahydroxyflavonol) and other flavonoids, resulting in the fixation of two oxygen atoms into the substrate accompanied by a ring-opening step (Figure 1). The reactivity of the substrates is greatly influenced by the distribution of the hydroxyl substituents [12]. For example, the absence of C3′-OH doubles the reaction rate, whereas lack of C7-OH and C4′-OH has a drastic negative effect [12]. The C3-hydroxyl and the C4-carbonyl groups as well as the presence of a double bond between C2 and C3 are essential for catalysis [10, 12, 31]. Since the earliest biochemical characterization of a 2,3QD [12], it was proposed that the substrates could bind in the active site, chelating the copper with the 3-hydroxyl and 4-carbonyl groups. This hypothesis was supported by later biomimetic studies and by the evidence that substrates are able to protect the enzyme from inactivation by metal chelators [10].

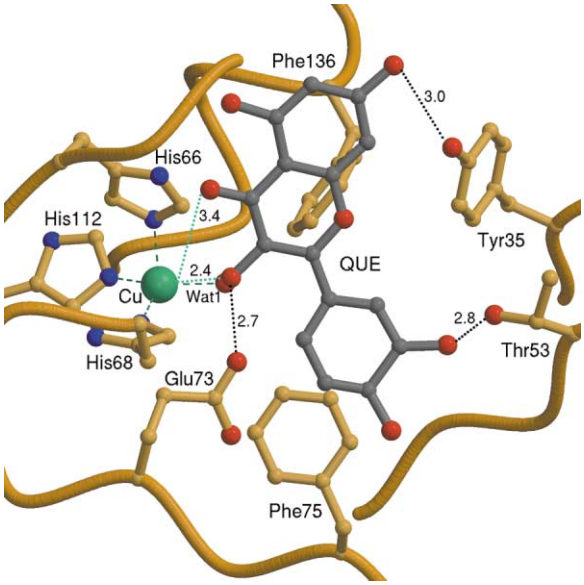


Figure 7. Model Complex with Quercetin

Diagram of the 2,3QD active site illustrating a possible mode of substrate binding. The position of the quercetin molecule (QUE, gray) has not been defined by diffraction data, and the indicated atomic distances (in Å) are based exclusively on manual modeling of the substrate molecule in the active site. The copper binding site and some of the surrounding side chains are shown as ball and stick diagrams.

Based on this knowledge, we have modeled a substrate molecule in the catalytic site of 2,3QD (Figure 7). Quercetin possesses five hydroxyl groups that probably replace part of the water structure upon binding in the active site. The manual docking was guided by the position of solvent molecules present in the active site of all four crystallographically independent molecules. The orientation of the quercetin molecule was adjusted to maximize the number of favorable protein-substrate interactions. The model shows that the molecule can bind to the copper in a monodentate manner, without requiring major conformational changes of the protein or the substrate structure. Tyr35 and Thr53 are in a favorable position to interact with C7-OH and C3′-OH, in agreement with an enhancing effect of these groups on the oxygenation rate. Substitution at C8 decreases activity; the model shows that there is not enough space for any atom other than hydrogen at this position. Several aromatic chains, including Tyr35, Phe75, and Phe136, surround the copper center and might stabilize substrate binding. The B ring of quercetin is positioned in a cavity adjacent to the copper site and points toward Gly125. Replacement with any other amino acid would be incompatible with the proposed substrate binding mode. A comparison between the N- and C-terminal domains shows that Gly125 is replaced by Phe322 in the C-terminal domain. In the model, the B ring of the quercetin occupies the position corresponding to that of the aromatic side chain of Phe322.

Studies with nonenzymatic model systems have suggested that the oxygenolysis of quercetin could be base catalyzed, with substrate activation achieved by deprotonation of the C3-OH group (Figure 1). As a result, negative charge develops at the C2 atom, activating the substrate toward electrophilic attack by O_2 [10, 31, 32]. If the enzyme-catalyzed reaction proceeds in a similar way, an amino acid base is expected to assist in substrate deprotonation. In 2,3QD, Glu73 is the only residue in a favorable position to function as the active site base that abstracts the proton from the reactive C3-OH group of the substrate. Site-directed mutagenesis studies showed that a Glu73Gln mutation resulted in an EPR active enzyme, with more than 1000-fold decreased activity, confirming that this residue is important (I. Kooter, personal communication).

Thus, our structural results together with the absence of O_2 bound to the Cu^{2+} ion are consistent with a catalytic pathway in which the substrate is bound to the metal center, and subsequent proton abstraction, promoted by Glu73, activates the complex for electrophilic attack by O_2. However, an alternative mechanism in which Glu73 modulates the redox potential of the copper and, in this way, may affect the substrate susceptibility to oxygen cannot be excluded at this stage.

Biological Implications

The incorporation of O_2 into biological substrates by oxygenases is an important reaction initiating the degradation of aromatic substrates. It is most commonly catalyzed by enzymes that utilize iron as a cofactor. The reaction catalyzed by the glycoprotein quercetin 2,3-dioxygenase (2,3QD) is unique in that it involves a type 2 Cu^{2+} center as catalyst. To gain insight into copper-catalyzed oxygenation, we have determined the crystal structure of 2,3QD from *Aspergillus japonicus* at 1.6 Å resolution.

The crystal structure revealed that 2,3QD is a dimeric enzyme stabilized by an N-linked glycan bound to Asn191. This shows that glycosylation of proteins is not only important for increasing protein solubility but may also fulfill specific structural roles. Unexpectedly, the Cu^{2+} ion exhibits two distinct coordination environments. In the first one, Cu^{2+} is coordinated by His66, His68, His112, and a water molecule. In the second one, Glu73 is an additional ligand, shifting the metal-bound water molecule to another position. EPR analysis of 2,3QD in solution confirms this dual nature of the Cu^{2+} environment. A carboxylate in the coordination sphere of copper in a natural enzyme has never been observed before and thus extends the repertoire of observed Cu binding sites in proteins. Moreover, the alternate positions of the Glu73 side chain may be of catalytic relevance for 2,3QD, since model building suggests that the side chain is close enough to abstract a proton from the substrate's 3-OH group.

Finally, the structure of 2,3QD revealed that the enzyme belongs to the cupin superfamily, which includes the iron-containing homogentisate 1,2-dioxygenase and the Mn-containing oxalate oxidase, the structures of which were solved recently. The conservation of the type 2 copper site of 2,3QD and the iron and manganese binding sites present in those enzymes suggests that these cupin metallo enzymes are evolutionarily related, despite their different metal binding preferences.

Experimental Procedures

Protein Purification

For 2,3QD expression, a large-scale fermentation (100 liters) of *Aspergillus japonicus* was performed using quercetin (10 g/l) as inducer. Secreted 2,3QD was purified from the culture medium by a 70% (w/v) ammonium sulfate fractionation. The precipitate was discarded, and the supernatant was concentrated by ultrafiltration using an Amicon cell (filter cutoff 10 kDa). The concentrated 2,3QD solution was applied onto a Cu-chelating Sepharose Fast Flow column and eluted with an imidazole gradient. Fractions containing 2,3QD were pooled and further purified by size-exclusion chromatography on a Pharmacia Superdex 200 column eluted with 30 mM MES buffer and 0.5 M NaCl (pH 6.0). The obtained sample was subsequently loaded onto a MonoQ column and eluted with an NaCl gradient. 2,3QD fractions were pooled and applied to a Pharmacia Sephadex G-25 column and eluted with 2.5 mM piperazine buffer (pH 5.0). Finally, 2,3QD-containing fractions were further purified by chromatofocusing on a Pharmacia MonoP column eluted with a polybuffer composed of 1:10 Pharmalyte74 and 1:100 Ampholyte, at pH 3.4. Ampholytes were removed by applying the protein sample onto a Pharmacia Superdex 75 column. The purity of 2,3QD was assessed by SDS-polyacrylamide gel electrophoresis. The obtained 2,3QD sample had a specific activity of 177 U/mg, measured as described previously [11]. Deglycosylation was performed by treating the purified 2,3QD with endoglycosidase-H (Boehringer-Mannheim) under nondenaturing conditions, following the manufacturer's instructions.

Spectroscopic Measurements

Atomic Absorption Spectroscopy analyses were performed with plasma emission spectrometry using a Perkin Elmer Model Plasma 1000.

Raman spectroscopy measurements were performed on a Jobin-Yvon Raman microspectrometer equipped with a CCD detector. An excitation wavelength of 633 nm was chosen to detect the O-O stretch frequency shift at 704 cm^{-1}.

Crystallization

For crystallization, 2,3QD was concentrated to 12–15 mg/ml in 20 mM MES buffer (pH 6.0). Initial crystallization conditions were found by using the sparse matrix approach [33]. Crystals of 2,3QD were grown at room temperature from hanging drop vapor diffusion setups using 22%–26% (w/v) PEG 8000 and 200 mM ammonium sulfate in 100 mM citrate buffer (pH 5.2) as a precipitant. The colorless crystals belong to space group $P2_1$, with unit cell parameters a = 108.55 Å, b = 55.78 Å, c = 123.68 Å, and $\beta = 98.31°$. For the soaking of heavy metal compounds, crystals were first transferred to a mother liquor containing 30% (w/v) PEG 8000, 20 mM ammonium sulfate, and 20 mM citrate buffer (pH 5.2). For cryoprotection, crystals were transferred for a few seconds in the reservoir solution enriched with ethylene glycol (25% v/v).

Data Collection and Structure Determination

Data from native crystals and crystals soaked in Na_2PtCl_6, K_2PtCl_4, and $KAuBr_4$ were collected at 100 K at beamline D2AM at the ESRF, Grenoble. Wavelengths were chosen to optimize the anomalous signal. In addition, a native 1.6 Å resolution data set was collected at beamline BW7B of the EMBL outstation at DESY, Hamburg. Heavy atom derivatives and low-resolution native data were processed using XDS [34] and programs from the BIOMOL package (University of Groningen). The high-resolution native data were processed with the HKL suite [35]. Programs from the CCP4 package [36] were used for various calculations. Phases could be calculated up to a resolution of 2.7 Å with programs from the PHASES package [37]. Data collection and phasing statistics are presented in Table 1. The phasing procedure started with the identification of the heavy atom sites. Later, the Cu sites of 2,3QD (four per asymmetric unit) could be unambiguously detected in cross-phased anomalous difference Fourier maps calculated from the native data. Phases were refined against the isomorphous and anomalous data. The overall figure of merit was 0.55 at 2.7 Å. The program DM [38] was used for solvent flattening and histogram matching for data between 50 and 2.7 Å. The electron density allowed very clearly the identification of the four molecules in the asymmetric unit, and the noncrystallographic symmetry (NCS) relationship could be determined. The NCS operators, refined with the program IMP of the RAVE package [39], were used to perform 4-fold averaging of the electron density with DM. Electron density maps obtained after this procedure were of exceptional quality and could be interpreted readily.

Model Building and Refinement

The structure of 2,3QD was built using the programs O [40] and QUANTA (Accelrys). Residues 4–151, 179–350, and the copper atom could be unambiguously placed in density. The model was initially refined against the 2.7 Å native data using the program X-PLOR [41]. For monitoring the refinement by the R_{free} method [13], 10% of the data was set aside. To reduce model bias due to the high noncrystallographic symmetry, the test set was taken in thin resolution shells. The initial R_{factor} was 46.1% (R_{free} 45.5%). After two subsequent rounds of refinement and manual model building, clear electron density for an N-acetylglucosamine residue was visible at four asparagine residues (109, 142, 191, and 248). Subsequently, the model was refined against data to 1.6 Å using the program CNS [42]. Bulk solvent correction and overall B factor correction were included in every refinement cycle. NCS restraints were initially imposed but later released, resulting in a clear improvement of the crystallographic R factor and R_{free} values. Sugar residues and solvent molecules were inserted over several refinement cycles. Final refinement, performed with data from 50 to 1.60 Å resolution, resulted in a crystallographic R factor of 16.2% and an R_{free} of 18.9%. The final model consists of a total of 1337 amino acids (four protein molecules of 334, 339, 334, and 330 residues), 4 copper atoms (one per molecule), 21 N-acetylglucosamine and 9 mannose residues, 32 molecules of ethylene glycol, and 1756 water molecules. Double conformations were included in the last refinement rounds and could be clearly assigned to the side chains of 39 amino acids and to 4 water molecules. A central stretch of amino acids remained undefined in the electron density in all four crystallographically independent molecules, and, consequently, residues 155–168, 155–163, 155–166, and 154–169 were not built in molecule A, B, C, and D, respectively. Similarly, the first two N-terminal residues were omitted in molecules A and B, and the first four were omitted in molecules C and D. The quality of the model was assessed using the program PROCHECK [43]. All residues display allowed stereochemical geometry, with 87% of the 937 nonproline and nonglycine residues in the most favored regions, 11.8% in the additionally allowed regions, and 1.1% in the generously allowed regions of the Ramachandran plot. The rms differences from ideal geometry were 0.009 and 1.48 Å for bonds and angles, respectively. Refinement statistics are listed in Table 2.

Acknowledgments

We thank the staff at beamline D2AM at the ESRF, Grenoble, and the staff at the EMBL Outstation at DESY, Hamburg, for access to synchrotron data collection and assistance, the ESRF for support of the work at the ESRF, the European Union for support of the work at the EMBL, Hamburg, through the HCMP Access to Large Installation Project, contract number CHGE-CT93-0040. R.A.S. acknowledges support by the Netherlands Foundation for Chemical Research (CW) with financial aid from the Netherlands Organization for Scientific Research (NWO).

Received: September 6, 2001
Revised: December 5, 2001
Accepted: December 5, 2001

References

1. Que, L., Jr. (1999). Oxygen activation at nonheme iron centers. In Bioinorganic Catalysis, J. Reedijk and E. Bouwman, eds. (New York: Marcel Dekker, Inc.), pp. 269–321.
2. Oka, T., and Simpson, F.J. (1971). Quercetinase, a dioxygenase containing copper. Biochem. Biophys. Res. Commun. *43*, 1–5.
3. Sharma, H.K., and Vaidyanathan, C.S. (1975). A new mode of ring cleavage of 2,3-dihydroxybenzoic acid in *Tecoma stans*

(L.). Partial purification and properties of 2,3-dihydroxybenzoate 2,3-oxygenase. Eur. J. Biochem. *56*, 163–171.
4. Boldt, Y.R., Sadowski, M.J., Ellis, L.B.M., Que, L., Jr., and Wackett, L.P. (1995). A manganese-dependent dioxygenase from *Arthrobacter globiformis* CM-2 belongs to the major extradiol dioxygenase family. J. Bacteriol. *177*, 1225–1232.
5. Gibello, A., Ferrer, E., Martin, M., and Garrido-Pertierra, A. (1994). 3,4-Dihydroxyphenylacetate 2,3-dioxygenase from *Klebsiella pneumoniae*, a Mg(2+)-containing dioxygenase involved in aromatic catabolism. Biochem. J. *301*, 145–150.
6. Abolmaali, B., Taylor, H.V., and Weser, U. (1998). Evolutionary aspects of copper binding centers in copper proteins. Struct. Bond. *91*, 91–190.
7. Holm, R.H., Kennepohl, P., and Solomon, E.I. (1996). Structural and functional aspects of metal sites in biology. Chem. Rev. *96*, 2239–2314.
8. Karlin, S., Zhu, Z.Y., and Karlin, K.D. (1997). The extended environment of mononuclear metal centers in protein structures. Proc. Natl. Acad. Sci. USA *94*, 14225–14230.
9. Hund, H.K., Breuer, J., Lingens, F., Huttermann, J., Kappl, R., and Fetzner, S. (1999). Flavonol 2,4-dioxygenase from *Aspergillus niger* DSM *821*, a type 2 CuII-containing glycoprotein. Eur. J. Biochem. *263*, 871–878.
10. Speier, G. (1991). Quercetin 2,3-dioxygenase mimicking chemistry. In Dioxygen Activation and Homogeneous Catalytic Oxidation, L.I. Simándi, ed. (Amsterdam: Elsevier Science).
11. Oka, T., Simpson, F.J., Child, J.J., and Mills, C. (1971). Degradation of rutin by *Aspergillus flavus*. Purification of the dioxygenase quercetinase. Can. J. Microbiol. *17*, 111–118.
12. Oka, T., Simpson, F.J., and Krishnamurty, H.G. (1972). Degradation of rutin by *Aspergillus flavus*. Studies on specificity, inhibition, and possible reaction mechanism of quercetinase. Can. J. Microbiol. *18*, 493–508.
13. Brünger, A.T. (1992). Free R-value: a novel statistical quantitiy for assessing the accuracy of crystal structures. Nature *355*, 472–475.
14. Beintema, J.J. (1986). Do asparagine-linked carbohydrate chains in glycoproteins have a preference for beta-bends? Biosci. Rep. *6*, 709–714.
15. O'Connor, S.E., and Imperiali, B. (1996). Modulation of protein structure and function by asparagine-linked glycosylation. Chem. Biol. *3*, 803–812.
16. Helenius, A., and Aebi, M. (2001). Intracellular functions of N-linked glycans. Science *291*, 2364–2369.
17. Dunwell, J.M., Khuri, S., and Gane, P.J. (2000). Microbial relatives of the seed storage proteins of higher plants: conservation of structure and diversification of function during evolution of the cupin superfamily. Microbiol. Mol. Biol. Rev. *64*, 153–179.
18. Lawrence, M.C., Izard, T., Beuchat, M., Blagrove, R.J., and Colman, P.M. (1994). Structure of phaseolin at 2.2 Å resolution. Implications for a common vicilin/legumin structure and the genetic engineering of seed storage proteins. J. Mol. Biol. *238*, 748–776.
19. Ko, T.P., Day, J., and McPherson, A. (2000). The refined structure of canavalin from jack bean in two crystal forms at 2.1 and 2.0 Å resolution. Acta Crystallogr. D Biol. Crystallogr. *56*, 411–420.
20. Woo, E.J., Dunwell, J.M., Goodenough, P.W., Marvier, A.C., and Pickersgill, R.W. (2000). Germin is a manganese containing homohexamer with oxalate oxidase and superoxide dismutase activities. Nat. Struct. Biol. *7*, 1036–1040.
21. Holm, L., and Sander, C. (1998). Touring protein fold space with Dali/FSSP. Nucleic Acids Res. *26*, 316–319.
22. Titus, G.P., Mueller, H.A., Burgner, J., Rodriguez De Cordoba, S., Peñalva, M.A., and Timm, D.E. (2000). Crystal structure of human homogentisate dioxygenase. Nat. Struct. Biol. *7*, 542–546.
23. Onishi, A., Liotta, L.J., and Benkovic, S.J. (1991). Cloning and expression of *Chromobacterium violaceum* phenylalanine hydroxylase in *Escherichia coli* and comparison of amino acid sequence with mammalian aromatic amino acid hydroxylases. J. Biol. Chem. *266*, 18454–18459.
24. Bordo, D., Djinović, K., and Bolognesi, M. (1994). Conserved patterns in the Cu,Zn superoxide dismutase family. J. Mol. Biol. *238*, 366–386.
25. Orpen, G., Brammer, L., Allen, F.H., Kennard, O., and Watson, D.G. (1989). Tables of bond lengths determined by X-ray and neutron diffraction. Part 2. Organometallic compounds and coordination complexes of the d- and f-block metals. J. Chem. Soc., Dalton Trans., S1–S83.
26. Henderson, R. (1995). The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. Q. Rev. Biophys. *28*, 171–193.
27. Ravelli, R.B., and McSweeney, S.M. (2000). The 'fingerprint' that X-rays can leave on structures. Structure *8*, 315–328.
28. Harding, M.M. (2001). Geometry of metal-ligand interactions in proteins. Acta Crystallogr. D Biol. Crystallogr. *57*, 401–411.
29. Rardin, R.L., Tolman, W.B., and Lippard, S.J. (1991). Monodentate carboxylate complexes and the carbolxylate shift: implications for polymetalloprotein and function. New J. Chem. *15*, 417–430.
30. Logan, D.T., Su, X.D., Aberg, A., Regnstrom, K., Hajdu, J., Eklund, H., and Nordlund, P. (1996). Crystal structure of reduced protein R2 of ribonucleotide reductase: the structural basis for oxygen activation at a dinuclear iron site. Structure *4*, 1053–1064.
31. Barhács, L., Kaizer, J., and Speier, G. (2001). Kinetics and mechanism of the stoichiometric oxygenation of the ionic zinc(II) flavonolate complex [Zn(fla)(idpa)]ClO4 (fla = flavonolate; idpa = 3,3′-iminobis(N,N-dimethylpropylamine)). J. Mol. Catal. *172*, 117–125.
32. Balogh-Hergovich, E., Kaizer, J., and Speier, G. (2000). Kinetics and mechanism of the Cu(I) and Cu(II) flavonolate-catalyzed oxygenation of flavonol. Functional quercetin 2,3-dioxygenase models. J. Mol. Catal. *159*, 215–224.
33. Jancarick, J., and Kim, S.-H. (1991). Sparse matrix sampling: a screening method for crystallization of proteins. J. Appl. Crystallogr. *24*, 409–411.
34. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Crystallogr. *26*, 795–800.
35. Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.
36. CCP4 (Collaborative Computational Project 4) (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D. Biol. Crystallogr. *50*, 760–763.
37. Furey, W., and Swaminathan, S. (1995). PHASES: a program package for processing and analysis of diffraction data from macromolecules. Methods Enzymol. *277*, 590–620.
38. Cowtan, K., and Main, P. (1996). Phase combination and cross validation in iterated density modification calculations. Acta Crystallogr. D Biol. Crystallogr. *52*, 43–48.
39. Kleywegt, G.J., and Jones, T.A. (1994). Halloween masks and bones. In From First Map to Final Model, S. Bailey and R.W.D. Hubbard, eds. (Warrington, UK: SERC Daresbury Laboratory), pp. 59–66.
40. Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for binding protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.
41. Brünger, A.T. (1992). X-PLOR: A System for Crystallography and NMR, Version 3.1 (New Haven, CT: Yale University Press).
42. Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.
43. Laskowski, R.A., Moss, D.S., and Thornton, J.M. (1993). Main-chain bond lengths and bond angles in protein structures. J. Mol. Biol. *231*, 1049–1067.
44. Drenth, J. (1999). Principles of Protein X-ray Crystallography, Second Edition (Berlin: Springer).
45. Engh, R.A., and Huber, R. (1991). Accurate bond and angle parameters for X-ray structure refinement. Acta Crystallogr. A *47*, 392–400.
46. Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.

47. Merritt, E.A., and Bacon, D.J. (1997). Raster3D photorealistic molecular graphics. Methods Enzymol. *277*, 505–524.
48. Esnouf, R.M. (1997). An exensively modified version of MolScript that includes greatly enhanced coloring capabilities. J. Mol. Graph. *15*, 133–138.
49. Westhead, D.R., Slidel, T.W., Flores, T.P., and Thornton, J.M. (1999). Protein structural topology: automated analysis and diagrammatic representation. Protein Sci. *8*, 897–904.

Accession Numbers

Atomic coordinates have been deposited in the Protein Data Bank at Rutgers University under accession code 1JUH.