Structure of the monofunctional heme catalase DR1998 from Deinococcus radiodurans

Patrícia T. Borges, Carlos Frazão, Cecília S. Miranda, Maria A. Carrondo and Célia V. Romão

Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Oeiras, Portugal

Keywords
bacteria; clade 1; metalloprotein; radiation-resistant; reactive oxygen species

Correspondence
C. V. Romão, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Apartado 127, 2781-901 Oeiras, Portugal
Fax: +351214433644
Tel: +351214469665
E-mail: cmromao@itqb.unl.pt

M. A. Carrondo, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Apartado 127, 2781-901 Oeiras, Portugal
Fax: +351214433644
Tel: +351214469657
E-mail: carrondo@itqb.unl.pt

(Received 1 April 2014, revised 12 June 2014, accepted 24 June 2014)

doi:10.1111/febs.12895

Introduction

Deinococcus radiodurans is an aerobic organism with the ability to survive under conditions of high radiation doses or desiccation. As part of its protection system against oxidative stress, this bacterium encodes three monofunctional catalases. The DR1998 catalase belongs to clade 1, and is present at high levels under normal growth conditions. The crystals of DR1998 diffracted very weakly, and the merged diffraction data showed an Rsym of 0.308. Its crystal structure was determined and refined to 2.6 Å. The four molecules present in the asymmetric unit form, by crystallographic symmetry, two homotetramers with 222 point-group symmetry. The overall structure of DR1998 is similar to that of other monofunctional catalases, showing higher structural homology with the catalase structures of clade 1. Each monomer shows the typical catalase fold, and contains one heme b in the active site. The heme is coordinated by the proximal ligand Tyr369, and on the heme distal side the essential His81 and Asn159 are hydrogen-bonded to a water molecule. A 25-Å-long channel is the main channel connecting the active site to the external surface. This channel starts with a hydrophobic region from the catalytic heme site, which is followed by a hydrophilic region that begins on Asp139 and expands up to the protein surface. Apart from this channel, an alternative channel, also near the heme active site, is presented and discussed.

Database
Coordinates and structure factors have been deposited in the Protein Data Bank in Europe under accession code 4CAB

Abbreviations
a.u., asymmetric unit; ESRF, European Synchrotron Radiation Facility; NCS, noncrystallographic symmetry; PDB, Protein Data Bank; ROS, reactive oxygen species.
activities of *D. radiodurans* extracts indicated that this organism contains more than one catalase. In fact, the KatA knockout mutant was shown to be only slightly more sensitive to ionizing radiation than the wild-type, indicating that the organism is still able to detoxify the ROS formed [17–19]. Later, analysis of the *D. radiodurans* genome revealed the presence of three catalases, i.e. DR1998 (KatA), DRA0259, and DRA0146 [1,15,16], the first two of which are constitutively expressed [17,19,20].

Catalases (hydrogen peroxide:hydrogen peroxide oxidoreductase; EC 1.11.1.6) are widespread in aerobic organisms of the three life domains, but are also present in microaerophilic and anaerobic organisms such as *Campylobacter jejuni* and *Desulfovibrio gigas* [21,22]. These proteins can be divided into heme-containing enzymes, including monofunctional catalases or bifunctional catalase-peroxidases, and manganese catalases. For reviews, see, for example, [23–27]. Monofunctional catalases are further subdivided into clade 1 and clade 2. Clade 1 consists of the smaller-subunit catalases (55–69 kDa) with heme *b* as the metal cofactor, which are present in plants, proteobacteria, and firmicutes. Clade 2 consists of catalases with larger subunits (75–84 kDa), resulting from the presence of an additional C-terminal flavodoxin domain; their prosthetic group is mostly heme *d*, although some examples with heme *b* are also known. The clade 2 catalases are mainly present in bacteria and fungi. Clade 3 consists of small-subunit (43–75 kDa) catalases containing heme *b*. These are widespread in different archaea, bacteria, and eukaryotes, and some have NADPH as a cofactor [25,27,28].

The catalytic dismutation reaction of two hydrogen peroxide molecules into one dioxygen and two water molecules occurs in two stages. Briefly, in the first stage, the heme iron (Fe**III**-porphyrin) is oxidized by the first hydrogen peroxide molecule to form compound I, which contains an oxo-Fe**IV** porphyrin with a π-cation radical, with the release of one water molecule [29,30]. Compound I can be then reduced by a second hydrogen peroxide molecule, with the release of one dioxygen molecule and of the second water molecule [31]. The three amino acids involved in the catalytic center are conserved among catalases: the proximal ligand is a tyrosine that coordinates the heme iron; a histidine and an asparagine located on the opposite, distal side of the heme plane also play a role in the catalytic mechanism. Although catalase was one of the first proteins to be crystallized (in 1937), the first catalase crystal structures only became available 40 years later: the bovine liver [Protein Data Bank (PDB) 1CAT] and the *Penicillium vitale* (PDB 4CAT) catalases [11–14].

The three catalases from *D. radiodurans* are monofunctional, each belonging to a different clade: DR1998 (60 kDa) is in clade 1; DRA0259 (84 kDa) is in clade 2; and DRA0146 is a smaller 40-kDa catalase, proposed to be more related to eukaryotic catalases from plants than to bacterial catalases [1,15]. The sequence identity between DR1998 and DRA0259 is 25%, whereas DRA0146 has only 6% sequence identity with either of the other two catalases (Fig. 1). Nevertheless, some residues are conserved among the three proteins, namely the tyrosine proximal ligand and the distal histidine (Fig. 1). DR1998 is the only catalase to have been biochemically characterized, and shows a specific catalase activity of 68 800 U mg⁻¹ but no peroxidase activity [32]. This protein is present at higher levels in *D. radiodurans*, suggesting an important role in the detoxification of ROS [20,33].

This article reports the first crystal structure of a catalase from the radiation-resistant bacterium *D. radiodurans*; the DR1998 structure was solved and refined at 2.6 Å resolution. The structural analysis revealed an overall fold and quaternary structure that are common among catalases. At present, several crystal structures of monofunctional catalases from the three clades are available, of which only two are clade 1 monofunctional catalases: those from the proteobacterium *Pseudomonas syringae* (PDB 1M7S) and from the firmicute *Exiguobacterium oxidotolerans* (PDB 2J2M) [34,35]. Therefore, the DR1998 crystal structure contributes to elucidate the details of structural variability among the different clades of monofunctional catalases.

### Results and Discussion

#### Structure determination and quality

The DR1998 crystal structure was determined by molecular replacement with four molecules in the asymmetric unit (a.u.) [36]. Structural refinement proceeded smoothly, and led to *R* \(_{free}\)/*R* \(_{work}\) of 0.2224/0.2436 and a final *R* \(_{factor}\) of 0.2363, with all available diffraction data. However, rmsd values of bond distances and bond angles (Table 1) are typical of structures with lower resolution. This is not too surprising, and could be attributable to the unusual low signal-to-noise ratio of the original diffraction data (see Experimental procedures). There are 564 solvent molecules in the structure, including 10 chloride ions (Table 1). The assignment of the chloride ions was based on a comparative analysis of their atomic displacement parameters with the neighboring atoms, and also on the presence of close neighbors with positive charges.
The four molecules in the a.u. (chains A, B, C, and D) are arranged as two dimers, which, by crystallographic symmetry (chains A<sub>0</sub>, B<sub>0</sub>, C<sub>0</sub>, and D<sub>0</sub>), form two typical catalase tetramers: tetramer 1, formed by chains A, B, A<sub>0</sub>, and B<sub>0</sub>; and tetramer 2, formed by chains C, D, C<sub>0</sub>, and D<sub>0</sub> [36]. Superimposition of the independent molecules showed a low rmsd of ~0.20 Å between the superimposed Ca atoms, as expected for a crystal structure refined with medium-resolution data and noncrystallographic symmetry (NCS) restraints.

There was clear electron density for all chains, except for the first 30 residues and a loop around Gly520. As the N-terminus up to residue 29 contains six glycines and 12 hydrophilic residues, the absence of electron density at all N-termini is most probably due to conformational disorder. Whereas chains A and D showed weak density (below 1σ contour level) on residue 522, chains B and C also showed weak density over its nearest neighbors, residues 520–523 and 521–522, respectively. Diffraction data showed a remarkably low Wilson B-factor of 25 Å<sup>2</sup> for the working resolution limit, but it compared nicely with the range of the mean B-values of each chain, 23–29Å<sup>2</sup> (Table 1). The C-terminal parts of all chains (A, B, C, and D), residues 456–536, showed atomic displacement parameters of ~40 Å<sup>2</sup>, indicating their more flexible nature.

The Ramachandran plot showed that 94.9% of the residues have conformational main chain angles within the favored regions, and no outlier residues were observed. The rmsd between the superimposed Ca atoms of tetramer 1 and tetramer 2 is 0.24 Å. Therefore, for simplicity, only tetramer 1 will be analyzed in the subsequent sections.

**Overall structure of DR1998**

The monomer of DR1998 adopts the highly conserved catalase fold, which is divided into four structural regions (Figs 1 and 2): the N-terminal arm (residues 30–82) contains only one α-helix and His81, one of the essential catalytic residues known as distal histidine; the second region (residues 83–377) contains one antiparallel eight-stranded β-barrel, with topology β1β4β5β6β7A,Bβ8β9β10 (strand β7 includes a β-bulge, and is therefore referred to as β7A,B), and 10 α-helices (α2–α11); the wrapping loop (residues 378–455) connects the β-barrel domain with the C-terminal domain, and has almost no secondary structure elements, only a small helix, α12; and finally, the C-terminal region (residues 456–536) is a helical domain composed of four consecutive α-helices (α13, α14, α15, α16) (Figs 1 and 2).
Table 1. Data collection, processing and refinement statistics. Numbers in parentheses refer to the highest-resolution shell.

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Beamline</th>
<th>ESRF ID-29</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detector</td>
<td>PILATUS 6M</td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.82656</td>
<td></td>
</tr>
<tr>
<td>Data processing</td>
<td>XDS</td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>C2221</td>
<td></td>
</tr>
<tr>
<td>Unit cell parameters (Å)</td>
<td>a = 97.33, b = 311.88, c = 145.63</td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>49.4–2.6 (2.76–2.6)</td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>352 748 (48 698)</td>
<td></td>
</tr>
<tr>
<td>Unique reflections</td>
<td>68 047 (10 656)</td>
<td></td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.3 (97.5)</td>
<td></td>
</tr>
<tr>
<td>Multiplicity</td>
<td>5.2 (4.6)</td>
<td></td>
</tr>
<tr>
<td>Mosaicity (°)</td>
<td>0.049</td>
<td></td>
</tr>
<tr>
<td>CC1/2 (%)</td>
<td>97.1 (75.4)</td>
<td></td>
</tr>
<tr>
<td>Rwork (%)</td>
<td>22.2</td>
<td></td>
</tr>
<tr>
<td>Rfree (%)</td>
<td>24.4</td>
<td></td>
</tr>
<tr>
<td>rmstd for bond lengths (Å)</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>rmstd for bond angles (°)</td>
<td>0.063</td>
<td></td>
</tr>
<tr>
<td>Average chain B-factor (Å²)</td>
<td>25, 28, 24, 29</td>
<td></td>
</tr>
<tr>
<td>Number of molecules in a.u.</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Estimated solvent content (%)</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>Ramachandran plot</td>
<td>Residues in favored regions (%)</td>
<td>94.9</td>
</tr>
<tr>
<td></td>
<td>Residues in allowed regions (%)</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Residues in disallowed regions (%)</td>
<td>0</td>
</tr>
<tr>
<td>PDB code</td>
<td>4CAB</td>
<td></td>
</tr>
</tbody>
</table>

DR1998 oligomerizes as a homotetramer with 222 point-group symmetry, which is highly conserved among catalases [25,26]. The dimensions of the DR1998 tetramer along the three consensually defined orthogonal two-fold axes, P, Q, and R [13], are approximately 95, 70, and 95 Å, respectively (Fig. 3).

Superposition of either the DR1998 monomer or tetramer with monofunctional catalases with known crystal structures (Table 2) shows that both the tertiary and quaternary structures of DR1998 are very similar to those of the other family members. Cα superposition of monomers gave, on average, rmsd values of 1.27 Å with clade 1, 1.52 Å with clade 2 and 1.45 Å with clade 3 structures, respectively, whereas the Cα superposition of tetrarsers resulted in average rmsd values of 1.08, 1.48, and 1.37 Å, respectively. As expected, the lowest rmsd values were obtained for the phylogenetically most closely related sequences, those of the clade 1 catalases from *P. syringae* (PDB 1M7S) and *E. oxidotolerans* (PDB 2P2M) [34,35].

DR1998 tetramer 1 includes a total of 273 hydrogen bonds between its four subunits, of which almost 80% are formed between the R-related and Q-related subunits, with the remaining 20% resulting from interactions between P-related subunits (Fig. 3). This

---

* a CC1/2 = correlation between intensities from random half-datasets [69].
* Rsym = Σ|(|I(hkl)| − <I(hkl)>|ΣI(hkl)|, where |I(hkl)| is the observed intensity and <I(hkl)> is the average intensity of multiple observations from symmetry-related reflections [52].
* Rfree = Σ|(|M(hkl)| − |<M(hkl)>|)|ΣM(hkl)|, where M(hkl) is the data multiplicity, |I(hkl)| is the observed intensity, and <I(hkl)> is the average intensity of multiple observations from symmetry-related reflections. It is an indicator of the agreement between symmetry-related observations [53].
* Rwork = Σ|(|N(hkl) − 1|)|ΣN(hkl)|, where N(hkl) is the data multiplicity, |I(hkl)| is the observed intensity, and <I(hkl)> is the average intensity of multiple observations from symmetry-related reflections. It is an indicator of the precision of the final merged and averaged data set [54].
* Rwork = ΣFobs − Fcalc|ΣFobs, where Fobs and Fcalc are the amplitudes of the observed and the model calculated structure factors, respectively. It is a measure of the agreement between the experimental X-ray diffraction data and the crystallographic model.
indicates that, in the tetramer, the major interactions are between the R-related and Q-related subunits, as confirmed by the relative interface areas between P-related, Q-related and R-related subunits of 2211, 4412, and 5908/C23 A², respectively [37,38]. The N-terminal arm and the β-barrel domain are the regions that contribute most to the total number of hydrogen bonds in the tetramer, most of which are between these two domains (29%) and those involving the wrapping loop (48%).

**Heme active site and heme pocket**

The crystal structure of DR1998, shows that each monomer contains one heme b group. Each heme is deeply buried inside the tetramer, and is mainly surrounded by residues from the β-barrel domain (Figs 2 and 4). The distances between the irons of each monomer are 29 Å for the P-related subunits, 35 Å for the R-related subunits, and 44 Å for the Q-related subunits. The bond distances mentioned in the following description are the mean values of the calculated equivalent bond distances from chains A, B, C, and D.

The iron atoms in each heme are pentacoordinated (Fig. 4) by the four nitrogen atoms from the porphyrin ring, at a mean distance of 2.1 Å, and the phenolate oxygen atom from the proximal Tyr369 ligand, at a distance of 2.3 Å. The iron atom from the heme group is displaced from the plane of the pyrrole nitrogen atoms by 0.3 Å towards the oxygen atom of the Tyr369 phenolate. Besides coordinating the heme iron, the Tyr369 phenolate oxygen atom forms two hydrogen bonds, with Arg365 NE (3.1 Å) and NH2 (2.9 Å). Arg365 NH1 is hydrogen-bonded to Asn229 OD1 (2.4 Å), Asn229 ND2 is hydrogen-bonded to Asp359 OD2 (2.9 Å), and Asp359 OD1 is hydrogen-bonded (2.9 Å) to a solvent water molecule, W2 (Fig. 4), located at the bottom of a cavity that is open to the solvent. This hydrogen-bonded network is conserved among the monofunctional catalases, and it has been proposed to be involved in the stabilization of the electrostatic field at the catalytic center in the course of redox catalysis [39].

Both Arg365 and Asp359 are strictly conserved residues; however, Asn229 is usually a histidine in other catalases (Table 2). Another known exception occurs in clade 2, in the catalase from *Neurospora crassa* (PDB 1SY7) (Table 2) [40].

On the heme distal side, the imidazole ring of the essential catalytic His81 is nearly parallel to the heme pyrrole IV ring, in what is normally termed the His-IV orientation. This feature is common to the catalases in clade 1, whereas, in those in clade 3, the heme plane is rotated by 180°, resulting in a His-III orientation [25]. His81 NE2 is 5.1 Å distant from the iron atom, and 2.7 Å away from water W1, whereas His81 ND1 is hydrogen-bonded to Ser125 OE (2.8 Å) and also to Thr126 O (2.8 Å). W1 is the only water present in the main channel described below, and is observed for all chains in the crystal structure of DR1998 (Fig. 4). The His-IV orientation, together with the hydrogen bonds with His81 ND1, favors the imidazole tautomeric with a proton on ND1, whereas NE2 is oriented towards the iron atom that participates in the catalytic transfer of the hydrogen peroxide, as proposed by Fita and Rossmann [41]. Also on the heme distal side,
Table 2. Structural alignment between DR1998 monomer and related monofunctional catalase X-ray crystallography structures. The corresponding residues involved on the proximal heme side, heme propionate groups, heme distal side, main heme channel and NADPH-binding region are presented. Only one PDB code was chosen from each organism. Clade 1: D. radiodurans (PDB 4CAB), P. syringae (PDB 1M7S) [34], and E. axiotoolerans (PDB 2J2M) [35]. Clade 2: Es. coli (PDB 1GGE) [46], Pe. vitale (PDB 2UIF) [57], and N. crassa (PDB 1SYZ) [40]. Clade 3: beef liver (PDB 7CAT) [70], S. cerevisiae (PDB 1A4E) [43], human erythocyte (PDB 1DG) [39], Micrococcus lysodeikteicus (PDB 1HBZ) [56], Proteus mirabilis (PDB 1MB5) [71], Enterococcus faecalis (PDB 1S8) [72], Helicobacter pylori (PDB 2OF) [57], Vibrio salmonicida (PDB 2IUF) [73], Hansenula polymorpha (PDB 2XQ1) [74], Corynebacterium glutamicum (PDB 4E37) and Pseudomonas aeruginosa (PDB 4E37). The conserved residues among all the catalases are in bold type. In the structures with PDB codes 1M85, 2ISA, and 4E37, it was observed a methionine sulfone (Omt), at the position of Val80 in DR1998 (4CAB).

<table>
<thead>
<tr>
<th>Table 2. Structural alignment...</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proximal heme side</strong></td>
</tr>
<tr>
<td>Tyr369</td>
</tr>
<tr>
<td>Arg141</td>
</tr>
<tr>
<td>Asp329</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Heme distal side</strong></td>
</tr>
<tr>
<td>His52</td>
</tr>
<tr>
<td>Arg78</td>
</tr>
<tr>
<td>Arg123</td>
</tr>
<tr>
<td>Arg376</td>
</tr>
<tr>
<td>Asp364</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Heme propionate groups</strong></td>
</tr>
<tr>
<td>Val228</td>
</tr>
<tr>
<td>Val167</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Heme vinyl group</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>NADPH-binding pocket</strong></td>
</tr>
<tr>
<td>His205</td>
</tr>
<tr>
<td>Ser206</td>
</tr>
<tr>
<td>Glu202</td>
</tr>
<tr>
<td>Gly234</td>
</tr>
<tr>
<td>Lys239</td>
</tr>
<tr>
<td>Thr324</td>
</tr>
<tr>
<td>Arg307</td>
</tr>
<tr>
<td>Glu457</td>
</tr>
</tbody>
</table>

Asn159, which is considered to be an essential residue for the catalytic mechanism, is located at a distance of 3.3 Å from water W1.

The heme propionates form hydrogen bonds with three conserved arginines in the monofunctional catalases (Arg78, Arg123 and Arg376 in DR1998) (Table 2). Carboxylate oxygens O1D and O2D from propionate III are hydrogen-bonded to Arg376 NE (2.7 Å) and Arg376 NH2 (2.6 Å), whereas O1A and O2A from propionate IV are hydrogen-bonded to Arg78 NE (3.1 Å) and Arg123 NH1, respectively (2.7 Å) (Fig. 4).
Heme main channel

The dismutation reaction occurs at the heme distal site, which is accessed from the surface through a 25 Å long channel. In the tetramer, these channels run parallel to the R-axis, and their entries are 20 Å apart (Fig. 5).

This main channel can be divided into two parts. The first, a hydrophobic region near the catalytic site is, 10 Å in length with a diameter of 4 Å, and is lined mainly by the hydrophobic residues Val80, Val127, Pro140, Phe164, Phe165, Phe172, Leu175, and Ile176 (Fig. 5). The second channel region is more hydrophilic, starts on Asp139, and is less conserved in the family, with some residues, e.g. His129, Arg138, Asn188, Asn190, and Asn265, running through a further length of 15 Å with a diameter ranging between 4 and 15 Å (Fig. 5).

The electrostatic charge distribution on the first part of this channel is influenced by the negative charge of Asp139 at the top of the hydrophobic region, and the positive charge of the metal site at the bottom. This feature allows fast substrate access to the iron, and explains the high values for the fast-turnover \( k_{\text{cat}} \) (range \( 7 \times 10^4 \) to \( 1.6 \times 10^6 \) s\(^{-1}\)) found in catalases [26,42].

This region has been regarded as a molecular ruler, optimized to select hydrogen peroxide. Mutations in the monofunctional catalase of \textit{Saccharomyces cerevisiae} (PDB 1A4E) designed to enlarge the hydrophobic channel region resulted in lower enzymatic activity and increased peroxidase activity. This was attributed to the facilitation of active site access by larger substrates with a concomitant decrease in hydrogen peroxide selectivity.
Lateral channel

Besides the above-described main channel, a lateral channel connecting the catalytic center to the external surface has been described for other monofunctional catalases, e.g. that from *S. cerevisiae* [43,44]. In the vicinity of the active site, the solvent-accessible surface of DR1998 shows a second channel (lateral channel) (Fig. 6). This is a 16-A-deep cavity that reaches the hydrophobic Val228, located close to the vinyl group linked to pyrrole I. Three water molecules are located within this channel in each of the four independent protein chains in the crystal structure. This channel has been reported to be involved in the exit/entry of molecules [45]. However, in DR1998, Val228 blocks access from the cavity to the active site. The homologous residues in other catalases are Val/Ile for clades 1 and 2 and Ser/Gly for clade 3 (Table 2). It has been proposed that these residues (Table 2), with an unfavorable $\phi$, $\psi$ conformation [34,40,46], could have a significant role in catalysis. Routine analysis of conformational $\phi$ and $\psi$ angles in macromolecules was pioneered by PROCHECK [47]; meanwhile, not only have the quantity and quality of data in the PDB increased enormously, but also the accuracy and specificity of Ramachandran plots have been steadily improved [48,49]. A re-examination of the above-mentioned structures with MOLPROBITY [50], which uses a dataset of 1.6 million residues with backbone atomic displacement parameters of $< 30$ Å$^2$, showed that, in fact, all of those residues have a conserved conformation, and that they are all located in the allowed regions of the Ramachandran ($\phi$, $\psi$) plot. Therefore, the previous suggestion regarding their catalytic role was based on outdated estimates of the favorable $\phi$, $\psi$ conformations, and is no longer applicable.

In clade 3 small-subunit catalases, this lateral channel has been described as part of a bifurcation with another channel that runs from the essential asparagine in the heme distal side up to the NADPH-binding pocket. A search for the presence of this type of channel in DR1998 was unsuccessful, as the corresponding region is blocked by the C-terminus (residues 530–536) of the R-related subunit (Figs 6 and 7). Although NADPH binding was only observed in some clade 3 catalases, the corresponding binding pocket is completely accessible from the exterior in clade 1 and clade 3 catalases, and the amino acids that participate in NADPH binding are present in catalases from the different clades (Table 2). In the case of clade 2 catalases, such as the *Escherichia coli* protein, this pocket is blocked by residues 585–590 (Fig. 7). However, an R260A mutation located in the binding pocket region showed, in this case, an increase in catalytic activity, suggesting faster diffusion of substrates or reaction products through that alternative protein channel [45]. In DR1998, Trp214 is located in a structurally equivalent position to Arg260 in *Es. coli*, and is the single example of a tryptophan at this position in all of the catalases included in Table 2. Therefore, in DR1998, this part of the lateral channel is blocked by the C-terminal residues 530–536 from the R-related subunit and also by Trp214. Nevertheless, the function of alternative channels in DR1998 warrants further investigation.

Concluding remarks

*D. radiodurans* is a radiation-resistant organism with a genome that encodes three monofunctional catalases, one from each of the three known clades. This article describes the crystal structure of the clade 1 catalase DR1998. A sequence alignment between DR1998 (clade 1) and DRA0259 (clade 2) shows that most of the residues important for catalysis are conserved. However, some of these residues are not conserved in DRA0146 (Fig. 1). The present structure was determined from a protein crystal with a diffraction data
signal-to-noise ratio of only 2.3, which is quite unusual. Despite these weak diffraction data, the precision indicators of the merged data, processed mosaic spread, Wilson B-factor and refined structure atomic displacement parameters showed lower values than those usually found from datasets recorded with crystals diffracting to a similar resolution. Such an accomplishment would have been unthinkable a few years ago, and was only made possible by the combined use of high-brilliance synchrotron radiation sources, pixel-type detectors and data collection strategies leading to a high multiplicity of equivalent measurements of the same independent reflection set.

The analysis of the DR1998 structure highlights the structural conservation of the oligomerization state, a tetramer with 222 point-group symmetry, as well as of the monomer architecture described as the catalase fold. The hemes are buried within the tetramer, and the residues involved in the catalytic active center are also conserved, namely the proximal ligand Tyr369, and the distal ligands His81 and Asn159. The main channel that is considered to allow access of the hydrogen peroxide to the active site begins at the heme pocket, and continues through a narrow hydrophobic region that is followed by a wider and more hydrophilic region that starts at Asp139, and that allows the access of solvent molecules from the external surface. An alternative channel was also identified, starting at the external surface and approaching (but not reaching) the heme through a channel that ends in Val228.

The DR1998 structure is the first crystal structure of a catalase from the radiation-resistant bacterium *D. radiodurans*. It will be important in the future to determine the crystal structures of the other two catalases from this bacterium, and to complement these with kinetic studies. This will not only contribute towards understanding the detoxification mechanisms for hydrogen peroxide in *D. radiodurans*, but also allow a structural comparison between the three different catalases, which are representatives of each of the three known clades of the monofunctional catalases.

### Experimental procedures

#### Protein purification and crystallization

DR1998 was obtained from *D. radiodurans* as previously described [36]. A 300 L culture of *D. radiodurans* was grown at 30 °C to an attenuation at 600 nm of 2.0, in M53 medium containing 0.5% (w/v) yeast extract, 0.5% (w/v) glucose, 0.5% (w/v) NaCl, and 1% (w/v) casein. The cells were harvested and resuspended in 20 mM Tris/HCl (pH 7.2). The soluble fraction was separated from the membrane pellet by ultracentrifugation at 182 000 g for 6.5 h, and dialyzed at 4 °C in 20 mM Tris/HCl (pH 7.2). The cells were then broken in a French press at 9000 p.s.i. The soluble fraction was separated from the membrane pellet by ultracentrifugation at 182 000 g for 6.5 h, and dialyzed at 4 °C in 20 mM Tris/HCl (pH 7.2).

The purification protocol was performed at 4 °C and pH 7.2. The soluble fraction was applied to a DEAE-Fast Flow column (XK 50/30, 10 mL/min) equilibrated with 20 mM Tris/HCl and a 0–1 M linear NaCl gradient in the same buffer was applied. The fraction containing the heme UV-visible fingerprint (peak at 400 nm) eluted at 550 mM NaCl, and was dialyzed against 20 mM Tris/HCl. This fraction was then loaded onto a Q-Sepharose HP column (XK 26/10, 2 mL/min) equilibrated with 20 mM Tris/HCl, and eluted with a 0–1 M linear NaCl gradient. The heme-containing fraction eluted at 600 mM NaCl. A final purification step was performed in order to remove the remaining contaminants. The heme-containing fraction was applied to a Q-Sepharose HiTrap HP column (5 mL, 1.5 mL/min) equilibrated with 20 mM Tris/HCl, and eluted with a 0–1 M linear sodium acetate gradient. The final fraction containing DR1998 eluted at 600 mM sodium acetate.

SDS/PAGE was used to confirm the purity of the protein, which was then used in crystallization trials as described in [36]. Crystals were obtained at room temperature with the hanging-drop vapor diffusion method, by
mixing 1.0 µL of 12 mg·mL⁻¹ purified native protein solution with 1.0 µL of reservoir solution, containing 0.2 M magnesium acetate, 0.1 M sodium cacodylate (pH 6.5) and 20% poly(ethylene glycol) 8000. Crystals (approximately 250 × 70 × 15 µm) were cryoprotected with the reservoir solution supplemented with 25% glycerol prior to flash-cooling in liquid nitrogen [36].

**X-ray diffraction, crystal structure determination, and refinement**

Diffraction data were measured at 100 K at beamline ID29 of the European Synchrotron Radiation Facility (ESRF, Grenoble, France) by the use of X-rays with a wavelength of 0.826 Å and a PILATUS 6M detector. Data collection, processing, crystallographic characterization and structure determination have been described previously [36], and are summarized in Table 1.

Crystals of DR1998 diffracted so faintly that, even when an X-ray beam with a diameter of 30 µm at ESRF beamline ID29 was used [51], the individual (unmerged) intensities showed a signal-to-noise ratio of only 2.3. Owing to a data multiplicity of 5.2, the final merged data showed a final signal-to-noise ratio of 5.3 for a resolution up to 2.6 Å, where the mean diffraction intensity approached the noise level (Table 1). However, the signature of the very weak original data was apparent in the unusually high merging statistics, \( R_{\text{sym}} [52] \), \( R_{\text{meas}} [53] \), and \( R_{\text{pim}} [54] \) (Table 1), as compared with PDB data [55] obtained at a similar resolution, showing \( <R_{\text{sym/merge}} > 0.10(6) \) (http://resb-auto-check.rutgers.edu/dev/pdbitem/index.html).

Additionally, the mosaic spread and Wilson \( B \)-values of 0.05° and 25 Å², respectively, are comparable to those usually found in protein crystals diffracting to higher resolution. The low mosaic spread is certainly influenced by the small beam divergence of ESRF station ID29, where the crystal sample is positioned ~15 m away from the beam primary slits, suggesting that the initial, unusually low signal-to-noise ratio of the diffraction was the main contributor to the high resolution limit discrepancy. As the diffraction intensity at low resolution was already near its detection limit, and despite a low Wilson \( B \)-value, even a small decay caused by a modest resolution increment (up to 2.6 Å) was enough to bring the diffraction intensity below its detection threshold.

Initial phases were determined by molecular replacement with an ensemble of six catalase monomers from different organisms showing the highest sequence identities (34–50%) with DR1998 [36]: PDB entries 2J2M from *E. oxidotolerans* [35], 1M7S from *P. syringae* [34], 1GWE from *Micrococcus luteus* [56], 1DGF from *Homo sapiens* [39], 2UF from *Penicillium vitale* [57], and 1SY7 from *N. crassa* [40].

As the crystal structure contained four DR1998 monomers in the a.u., NCS between them was expected, and 3% of the data in thin-resolution shells were set aside for \( R_{\text{free}} \) monitoring [36]. The *TLEMSD* server (http://skuld.bmsc.washington.edu/~tlsmd) [58] was used to define polypeptide chain regions for translation, libration and screw refinement of atomic displacement parameters. *PHENIX.RUNEF* [59] was used for structure refinement of atomic positional, atomic isotropic and translation, libration and screw atomic displacement parameters, and included standard stereochemical and NCS restraints. During refinement, \( R_{\text{free}} \) was used to steer the relative weights of stereochemical restraints versus experimental data in the minimized function. Structure refinement was performed in iteratively repeated cycles of protein and solvent updating and refinement, alternated with inspection of \( \sigma_A \)-weighted 2\( |F_o| - |F_c| \) and \( |F_o| - |F_c| \) electron density maps for manual model improvement with *COOT* [60]. The stereochemistry of the refined structure was analyzed with *MOLPROBITY* [50]. Diffraction data and refinement statistics are presented in Table 1. Structure factors and associated structure coordinates were deposited in the Protein Data Bank in Europe [61]. The analysis of molecular channels was performed with *CAVER* [62], [63] and *MOLE* [64].

**Sequence alignment**

A multiple amino acid sequence alignment of the *D. radiodurans* catalases was performed with *CLUSTALX* [65] and edited with *GENEDOC* [66].

**Figures**

Figures were prepared with *PYMOL* [66,67,68].

**Acknowledgements**

This work was supported by grants and fellowships from Fundação para a Ciência e Tecnologia under the following projects: PTDC/BIA-PRO/100365/2008 (C. V. Romão) and PEst-OE/EB/60004/2013. P. T. Borges is the recipient of PhD grant SFRH/BD/85106/2012; C. S. Miranda was supported by a BI fellowship within FCT grant PTDC/BIA-PRO/100365/2008; and C. V. Romão is the recipient of SFRH/BD/94050/2013. We thank the ESRF (Grenoble, France) for support during the X-ray diffraction data collection, and P. Matias for critically reviewing the manuscript.

**Author contributions**

C.V.R., C.F. and P.T.B. designed research; C.V.R., C.F., P.T.B., C.S.M. performed research; C.V.R., C.F. and P.T.B. analyzed data; and C.V.R., C.F., P.T.B. and M.A.C. wrote the paper.
References


68 The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.


