Life Sciences 91 (2012) 115–126

Contents lists available at SciVerse ScienceDirect

Life Sciences

journal homepage: www.elsevier.com/locate/lifescie

Homology modeling and phylogenetic relationships of catalases of an opportunistic pathogen *Rhizopus oryzae*

Beáta Linka [a], Gerda Szakonyi [a], Tamás Petkovits [c], László G. Nagy [c], Tamás Papp [c], Csaba Vágvölgyi [c], Sándor Benyhe [b], Ferenc Ötvös [b,*]

[a] Institute of Pharmaceutical Analysis, Faculty of Pharmacy, University of Szeged H-6720 Szeged, Somogyi u. 4, Hungary
[b] Institute of Biochemistry, Biological Research Centre, Hungarian Academy of Sciences, H-6701 Szeged, P.O. Box 521, Hungary
[c] Department of Microbiology, Faculty of Science and Informatics, University of Szeged, H-6726 Szeged, 52 Közép fasor, Hungary

ARTICLE INFO

Article history:
Received 29 February 2012
Accepted 15 June 2012

Keywords:
Catalase
Homology modeling
3D structure
R. oryzae
Phylogenetic relationships

ABSTRACT

Aims: A homology modeling methodology was developed and used to obtain the 3D structures for four putative catalases of *Rhizopus oryzae* in order to assess their functionality.
Main methods: Homology models were built using different modeling strategies using non-protein compounds as steric constraints, a symmetry constraint to force identical chains and an additional loop modeling algorithm. Percent structural overlap values (SO) were calculated for each model–template pair to qualify the homology models.
Key findings: Comparing the different modeling strategies by the SO values revealed that the quality of the models, i.e. the similarity to the template was greatly increased in the presence of the prosthetic groups, modeling multiple protein chains together, enforcing symmetrical chains and applying additional loop modeling. For the best homology models achieved this way, the SO values express similar evolutionary relationships between the proteins modeled and the templates that were previously established by phylogenetic analysis. In three out of the four catalases of *R. oryzae* the highest quality models, the active center, i.e. the heme molecule and the surrounding amino acids showed a spatial arrangement identical to that observed experimentally in other catalases. The remaining protein is missing an 11 residue long fragment and has mutated residues within the active center.
Significance: Better homology models can be obtained with templates chosen by phylogenetic relationship, although building an accurate model needs structural constraints too. Calculating the structural overlap between the models and the templates may also help to find the appropriate templates.

Introduction

Rhizopus oryzae, a member of the class Zygomycetes, is the main causative agent of zygomycosis which is an invasive, rapidly evolving and often fatal infection in patients having a heavily immunocompromised health condition. These opportunistic, mainly rhinocerebral fungal infections are of increasing clinical interest due to the high mortality rate (75–90%) (Ribes et al., 2000; Eucker et al., 2001). The therapeutic ways are narrow including an early diagnosis followed by aggressive surgical and antifungal treatment (Hemashettar et al., 2011). Moreover, these fungi have an intrinsic resistance against the recently available antifungal agents. Neutrophil cells play an important role among the defense mechanisms against the invasive fungal infections by effectively killing the hyphae with hydrogen peroxide (Kolotila and Diamond, 1990). Therefore, the microbial catalase production against the host's hydrogen peroxide may be regarded as a possible virulence factor in patients with an impaired neutrophil function (Shibuya et al., 2006). Consequently, in a future curative treatment the functionality of catalases of *R. oryzae* may also be targeted which raises the need for their detailed investigation.

The family of catalases is divided into groups of heme and non-heme catalases. Besides the bifunctional catalase–peroxidases the monofunctional catalases constitute the largest group within the heme catalase family (Bernroitner et al., 2009). The monofunctional catalases are classified into two subgroups by phylogenetic analysis named as small-subunit catalases and large-subunit catalases with a molecular mass of 55 to 69 kDa and 75 to 84 kDa, respectively. The small-subunit enzymes have associated heme b and some of them have NADPH. The large-subunit enzymes characterized so far have associated heme d and none have NADPH (Klotz and Loewen, 2003). Most catalases are active in homotetrameric form. Former phylogenetic analyses of the catalase genes revealed the existence of three main clades that were segregated early in the evolution of this gene family through at least two gene-duplication events (Zamocky et al., 2008; Klotz and Loewen, 2003). Bacterial catalases are found in all three distinct groups. Clade 1 and 3 contain

* Corresponding author. Tel.: +36 62599635; fax: +36 62433506.
E-mail address: otvos@brc.hu (F. Ötvös).

0024-3205/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.lfs.2012.06.016

exclusively small-subunit prokaryotic sequences, while clade 2 contains exclusively large-subunit sequences (Klotz et al., 1997). The two subgroups of the small-subunit catalases are located in the phylogenetic tree distinctly, namely in clade 1 and 3. The fungal catalases are found in clade 2 and 3. The phylogenetic analyses suggest a highly conserved protein sequence of catalases resulting in highly conserved 3D structures too. The large-subunit catalases have a 50–70 residue long N-terminal and an about 150 residue long C-terminal extension compared to the small-subunit ones preserving the tertiary structure assembly common in all monofunctional catalases. The catalytic center and surroundings are highly conserved (Zamocky and Koller, 1999).

Lacking experimental structures, our aim was to construct accurate 3D models of the four hypothetical catalases of *R. oryzae* by homology modeling in order to gain insight into their catalytic center and to find amino acids as possible targets for further drug design. To improve model accuracy, an efficient homology modeling strategy was developed.

Materials and methods

Sequence analysis

The putative amino acid sequences were obtained scanning the *R. oryzae* genome database (*Rhizopus oryzae* Sequencing Project, Broad Institute of Harvard and MIT, 2004, http://www.broadinstitute.org/annotation/genome/rhizopus_oryzae/MultiHome.html) for catalases. One annotated and three putative catalase sequences were found. To characterize the proposed protein and to predict the signal peptide, the programs ProtParam (Gasteiger et al., 2005) and SignalP (Bendtsen et al., 2004) were used, respectively. Domain search and prediction were performed using the Motif Scan (MyHits) program (Pagni et al., 2007). For phylogenetic analyses, amino acid sequences of 36 monofunctional catalases (Appendix 5) were downloaded from GenBank; raw sequence data of zygomycetes hypothetical proteins were obtained from the genome sequence databases of *R. oryzae*, *Mucor circinelloides* (DoE Joint Genome Institute; *M. circinelloides* CBS277.49 v1.0; http://genome.jgi-psf.org/Mucci1/Mucci1.home.html) and *Phycomyces blakesleeanus* (DoE Joint Genome Institute P. blakesleeanus v1.1; http://genome.jgi-psf.org/Phybl1/Phybl1.home.html). The amino acid sequences were aligned using the Probalign algorithm (Roshan and Livesay, 2006) with default settings. The resulting alignments were used to infer trees using Bayesian MCMC algorithm implemented in the program package MrBayes 3.2.1 (Ronquist and Huelsenbeck, 2003) run as a standalone version under Windows®. Best fit substitution models were estimated by the Reversible-Jump MCMC approach; Markov Chains were run for five million generations sampling every 100th generation. The first 30,000 trees were discarded as burn-in, and the remaining trees were used to compute a 50% Majority Rule consensus tree; Bayesian Posterior Probabilities (BPP) ≥ 0.95 were considered significant. Members of the Reversible-Jump MCMC analysis selected the WAG model implemented in MrBayes as the best fit model to the data, with a posterior probability of 1.00.

Homology modeling

Homology modeling combines computational chemistry and bioinformatics tools to calculate atomic-resolution model of the target protein from its amino acid sequence and one or more experimental three-dimensional structures of related homologous proteins, i.e. the templates. Homology modeling consists of the following steps: (i) template selection by database search, (ii) sequence alignment between the target and template proteins, (iii) model building and refinement, and (iv) model evaluation (Liu et al., 2011). Due to the sequential completion of steps i–iv any error propagates to the later steps decreasing the model accuracy. Alignment errors are unrecoverable therefore both finding the proper template and making a good alignment need special attention.

Homology modeling was performed by the program package MODELLER 9v8) (Šali and Blundell, 1993) running under Linux distributions Kubuntu 11.4 and Mandriva 2011). The model building algorithm of MODELLER includes an automatic sequence alignment followed by modeling the backbone, side chains and disulfide bonds in accordance with the sequence alignment and other statistical potential related restraints embedded in the program. Finally, MODELLER finishes the model structures by a short simulated annealing molecular modeling using the Charmm22 molecular force field (http://salilab.org/modeller/9v7/manual/node442.html). Additional features of MODELLER were utilized as follows: using multiple chains and multiple templates; involvement of the non-protein component used as rigid body constraint (heme b or d, depending on what the template contained, see Appendix 1); symmetry constraint i.e. forcing identical geometry for each chain; applying the "dopehr_loopmodel" module instead of the faster "allhmodel" module to improve the flexible loop regions; and rating the models by the percent structural overlap values (SO) using the "iterative_structural_align" module in order to select the most appropriate templates for multitemplate modeling. In the preparation of the template PDB files for modeling homotetramers the "MakeMultimer.py" (http://watcut.uwaterloo.ca/cgi-bin/makemultimer) python program was used to restore the symmetrical homotetramer coordinates where only the asymmetric unit was given together with the symmetry operator. Visualization of the structures was performed by the Visual Molecular Dynamics (VMD) program package (Humphrey et al., 1996). Finally, the quality assessment of the structures obtained by homology modeling was performed by the program PROCHECK (Laskowski et al., 1993).

Procedure to obtain 3D models

Fig. 1 shows the relation between the different modeling methods used. The complexity of the methods increases from the upper left corner to the lower right corner and the arrows point toward the expected improvements of the model quality. Details of the methods are as follows:

A: Homology modeling of a single chain (monomer) of the catalases with each template:

1. Simplest model obtained by only using the protein atoms of the template (Table 3, Appendices 2–4: monomer, chain only).
2. Improved model using the heme molecule as a constraint (significant improvement of the models; Table 3, Appendices 2–4: monomer, chain only + hem).
3. Further improvement by applying the loop modeling module (Table 3, Appendices 2–4: monomer, chain only + hem + loop modeling).

B: Modeling the tetramers with each template:

1. Tetramer model obtained by only using the protein atoms of single templates (Table 3, Appendices 2–4: tetramer, chain only).
2. Improved models using the heme molecule as a constraint (significant improvement of the models; Table 3, Appendices 2–4: tetramer, chain only + hem).
3. Further improvement by applying the loop modeling module (Table 3, Appendices 2–4: tetramer, chain only + hem + loop modeling).

C: The same as in B with additional symmetry constraint to force the four chains to be identical (Table 3, Appendices 2–4: tetramer + symmetry + chain only; tetramer + symmetry + chain only + hem; tetramer + symmetry + chain only + hem + loop modeling) (Fig. 6).

Real-time quantitative PCR (qPCR) analysis

Transcription level of the examined genes was analyzed using the qPCR method. Total RNA samples were purified from mycelia cultivated

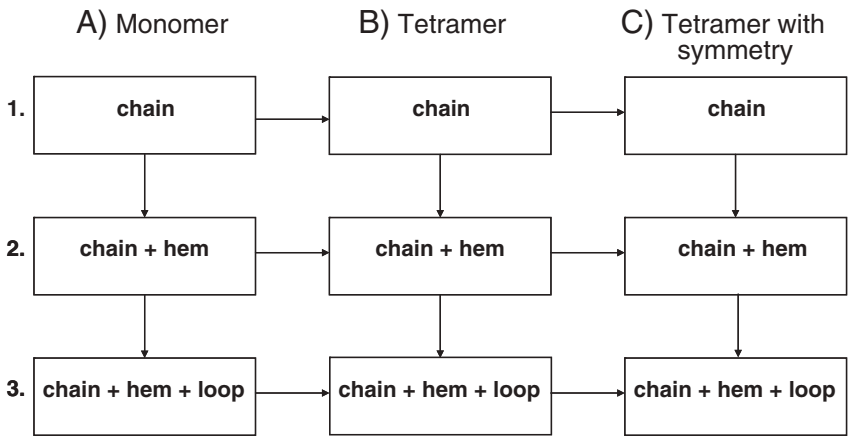

Fig. 1. The modeling strategies used in homology modeling.

in YEG medium with the E.Z.N.A. Total RNA kit II (Omega Bio-Tek) according the recommendations of the supplier. In the reverse transcription reactions, the RevertAidTM H Minus M-MuLV Reverse Transcriptase kit (Fermentas) using random hexamer and oligo(dT)18 primers was applied. QPCR experiments were performed in iQ5 and CFX96 real-time PCR detection systems (Bio-Rad) using the iQTM SYBR Green Supermix (Bio-Rad). The amplification conditions were as follows: an initial denaturing step of 3 min at 95 °C was followed by 39 cycles of 15-s denaturation (95 °C) and 30-s annealing (60 °C) and 15-s extension (72 °C). Sequences of forward primers and reverse primers were as follows: qcat1F (5′-acc ctg ata ctc acc gtc atc gtc-3′) and qcat1R (5′-cgc tga tga tta gag aca cgg cac-3′); qcat2F (5′-cct gat gga acc tca tcc cgc t-3′) and qcat2R (5′-gca acg gcc tta cca gag aca g-3′); qcat3F (5′-ctt ttg ttc gct tct cta ccg tcc-3′) and qcat3R (5′-ggc atc cag atc cca gtt tcc-3′); qcat4F (5′-aca agc ctg aac ctg ttt ctc cca-3′) and qcat4R (5′-agc ttc ttg cca tgt ttg cct ctc-3′); qRO3G_12558actinF (5′-gcc att caa gct gtc ctt tcc-3′) and qRO3G_12558actinR (5′-cca tcA cca gag tca aga acg a-3′). Data were normalized using an actin-2 (gene locus number: RO3G_12558) reference. The relative quantification of the gene expression was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

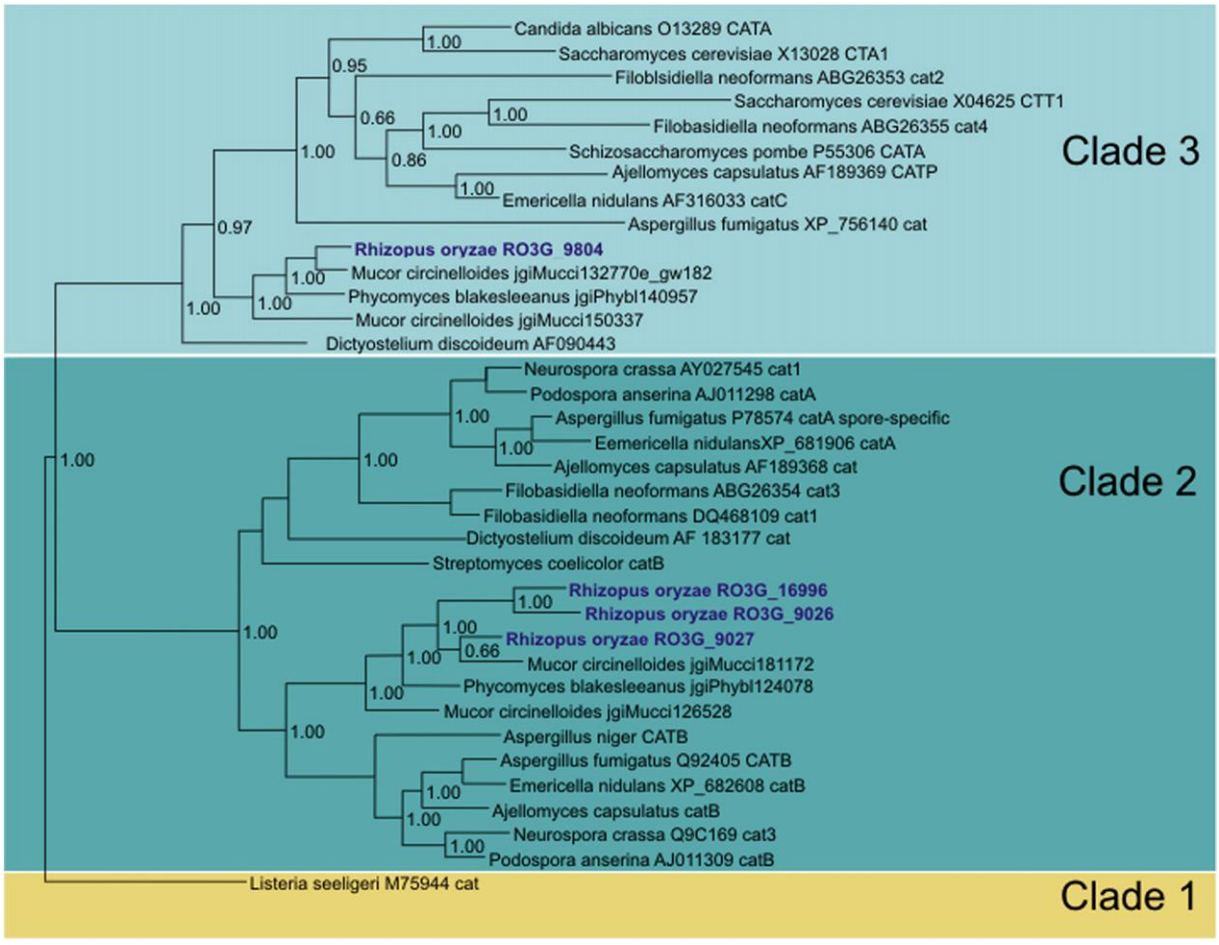

Fig. 2. The phylogenetic tree based on Bayes analysis of the amino acid sequences. The 50% Majority Rule consensus tree obtained by Bayes analysis.

Table 1
Structural information of the four catalases.

Protein	CAT1	CAT2	CAT3	CAT4
Calculated molecular weight (kDa)	54.564	80.130	81.232	79.658
Length (residues)	484	710	718	710
Signal peptide	–	–	1–20 aa	–
Possible microbody C-terminal targeting signal	482–484 aa	–	–	–
Distal site of the prosthetic heme group	H54, S93, V95, N127, F132	H72, S111, V113, N145, F150	H90, S129, V131, N-, F-	H72, S111, V113, N145, F150
Proximal site of the prosthetic heme group	Y337, R344	Y359, R366	Y366, R373	Y359, R366
Distal site of the NADPH group	H173, R182, H214, H284	H195, R204, H236, H-,	H202, R211, H343, H-	H195, R204, H236, H-
Proximal site of the NADPH group	L421, L-, E-	L-, L449, E-	L-, L457, E-	L-, L449, E-
Tetramerization domain	1–50 aa	1–68 aa	21–86 aa	1–68 aa
Clade	3	2	2	2

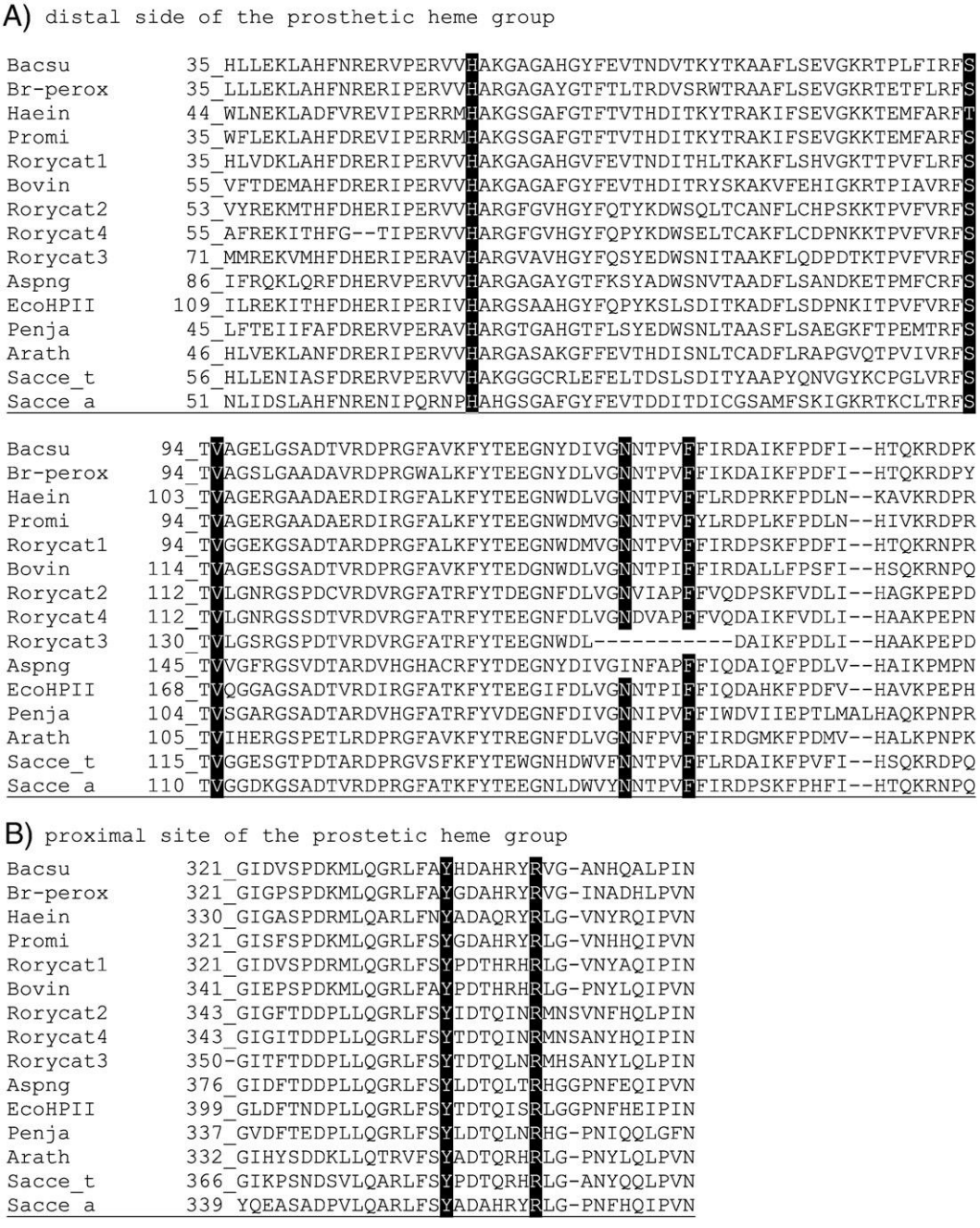

Fig. 3. Multiple sequence alignment of typical catalases and the four *R. oryzae* catalases in the areas of the distal and proximal side of heme and NADPH. The functionally important residues are in black boxes.

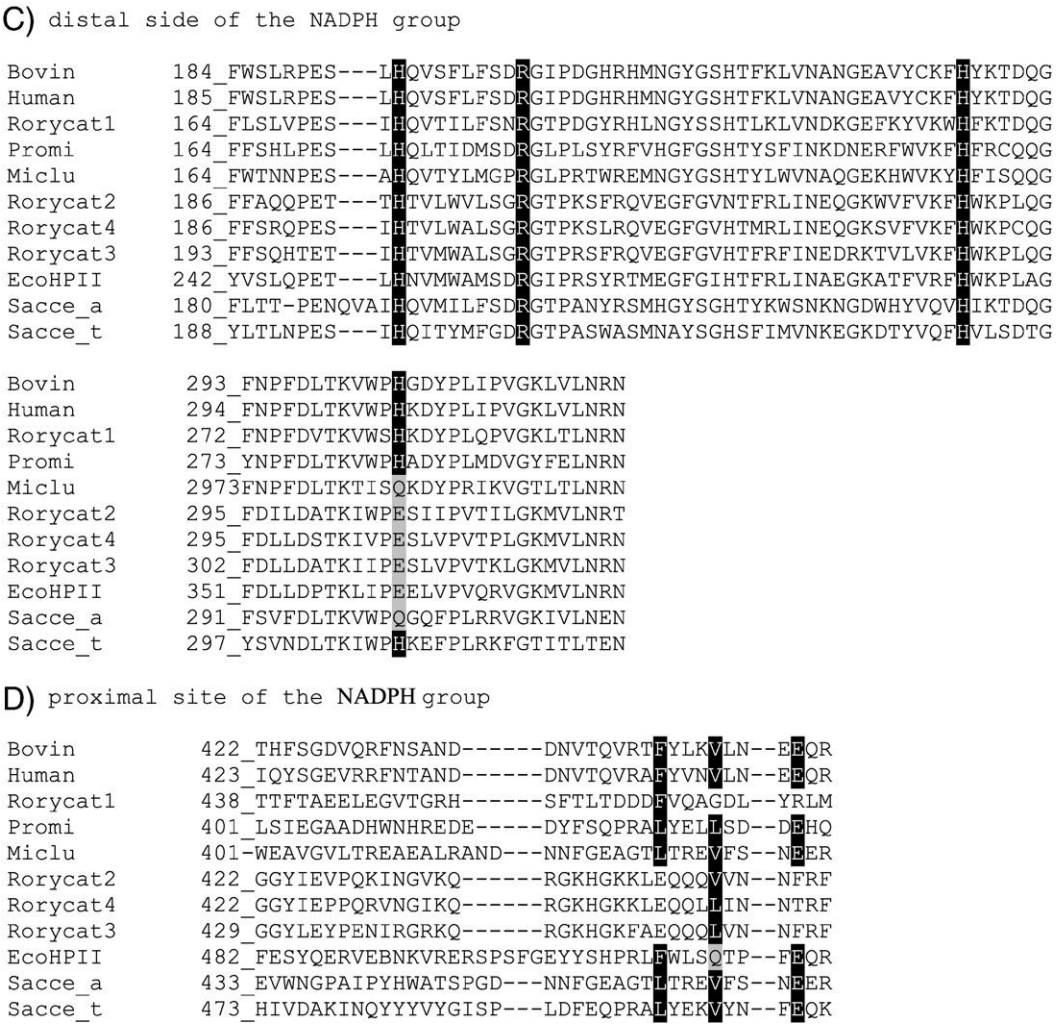

Fig. 3 (*continued*).

Results

Investigation of the phylogenetic relationships of catalase genes of R. oryzae

Numerous genes coding catalase enzymes were identified, cloned and characterized from different organisms including several yeast and human pathogen filamentous fungi. The whole genome of *R. oryzae* was published recently (*Rhizopus oryzae* Sequencing Project, Broad Institute of Harvard and MIT, 2004, http://www.broadinstitute.org/annotation/genome/rhizopus_oryzae/MultiHome.html) and four protein sequences were found to be homologous to monofunctional catalases denoted here as CAT1, CAT2, CAT3 and CAT4 (corresponding gene loci RO3G_9804.3, RO3G_09026.3, RO3G_09027.3 and RO3G_16995.3, referred as cat1, cat2, cat3 and cat4, respectively).

The amino acid sequences of the hypothetic catalases CAT1, CAT2, CAT3 and CAT4 were subjected to a phylogenetic analysis against the heme-containing fungal monofunctional catalase family. These catalases are members of clades 2 and 3 according to the classification of catalases by Klotz et al. (1997), Klotz and Loewen (2003)) and cover the fungal enzymes. According to this the phylogenetic analysis involved the protein sequences of the fungi monofunctional catalases from the NCBI GenBank and the genome databases of *M. circinelloides* and *P. blakesleeanus*. Catalase A of *Listeria seligeri* belonging to clade 1 was used as an outgroup. The resulting consensus tree is presented in Fig. 2. The comparison with the phylogenetic trees (Klotz et al., 1997; Klotz and Loewen, 2003) show that the hypothetic catalases of *R. oryzae* are expectedly located in clades 2 and 3 of catalases. CAT1 is positioned in clade 3 and because this clade involves the fungal peroxisomal enzymes close to CAT1 in the tree (*Mucor circinelloides*, *Phycomyces blakesleeanus*, *Aspergillus fumigatus*) it may be suggested that CAT1 is a peroxisomal catalase. CAT2, CAT3 and CAT4 are members of the large-subunit clade 2 catalases in the phylogenetic tree (monomers ranging from 79 to 80 kDa). CAT2 and CAT4 formed a sister group suggesting that they are results of a recent gene duplication event.

Sequence analysis of the catalase proteins

According to its molecular mass (54.5 kDa monomers) and the localization in the phylogenetic tree, CAT1 is a member of the small-subunit catalases in clade 3. Furthermore, it contains an SLA tripeptide sequence at the extreme C-terminus (Table 1) which is characteristic to the peroxisomal targeting signal 1 (PTS1) for sorting proteins to the peroxisome (Gould et al., 1988). This is in an agreement with the fact that many of the small-subunit catalases contain PTS1 signal (Klotz et al., 1997; Klotz and Loewen, 2003). As it is expected a common, ca. 70 residue long N-terminal threading arm can be found in all four proteins (Liu and Eisenberg, 2002; Putnam et al., 2000). This fragment ends with an ERIPE motif toward the C-terminus except CAT4 which has a fragment GTIPE. Only CAT3 contains a 20 residue long fragment at the extreme N-terminus (Table 1) which may have a possible extracellular role indicated by the programs ProtParam (Gasteiger et al., 2005) and SignalP (Bendtsen et al., 2004), although it was not specified. A former multiple

Table 2
Source of the catalases used in the multiple sequence alignment.

Abbreviation	Organism	Specification	Accession #
Bacsu	*Bacillus subtilis*		NCBI ID: P26901
Br-peroxa	*Streptomyces violaceus*		NCBI ID: P33569
Haein	*Haemophilus influenzae*		NCBI ID: P44390
Promi	*Proteus mirabilis*		NCBI ID: P42321
Bovin	*Bos taurus*	Bovine, liver	NCBI ID: P00432
Aspng	*Aspergillus niger*		NCBI ID: P55303
EcoHPII	*Escherichia coli*	HPII	NCBI ID: P21179
Penja	*Penicillium janthinellum*		NCBI ID: P81138
Arath	*Arabidopsis thaliana*		NCBI ID: P25819
Sacce_t	*Saccharomyces cerevisiae*	Catalase T	NCBI ID: P06115
Sacce_a	*Saccharomyces cerevisiae*	Catalase A	NCBI ID: P15202
Rorycat1	*Rhizopus oryzae*	CAT1 (catalase)	RO3G_09804.3
Rorycat2	*Rhizopus oryzae*	CAT2 (hypothetical protein)	RO3G_09026.3
Rorycat3	*Rhizopus oryzae*	CAT3 (hypothetical protein)	RO3G_09027.3
Rorycat4	*Rhizopus oryzae*	CAT4 (hypothetical protein)	RO3G_16995.3

sequence alignment revealed conserved structural motifs in catalases to coordinate the heme and NADPH molecules (Zamocky and Koller, 1999). Based on those results a multiple sequence alignment was performed including catalases of *R. oryzae* and those in (Zamocky and Koller, 1999). The presence or absence of the conserved residues at the distal and proximal sides of both heme and NADPH is shown in Fig. 3 and in Table 2. The conserved residues at the proximal side of the heme ring were found in all four catalases while an 11 residue long fragment is lacking in CAT3 involving the possible essential Asn and Phe at the distal side. Regarding the NADPH binding sites, multiple amino acids are missing at the distal side and none can be found at the proximal side in all four proteins.

Homology modeling

Representative structures obtained by the different modeling met hods are shown for CAT1 as target and the template catalase from *Bos taurus* (PDB code 3RGP) which was best scored by the PSI-BLAST search for CAT1. The same algorithms were applied for the other target and template proteins too. CAT1 was used to develop the modeling methodology because it is the shortest one out of the four targets (Table 1) and most of the template proteins selected have a similar length making the modeling easier. Additionally, this catalase proved to be the most active one in every stage of the life cycle of *R. oryzae* in preliminary qPCR experiments shown in Fig. 4.

Template selection from database

The four assumed catalase sequences of *R. oryzae* were aligned against the protein sequences in the Protein Data Bank repository using the NCBI web server. Using the PSI-BLAST algorithm with the BLOSUM62 replacement matrix and default parameters 43 candidates were retrieved with acceptable scores. Out of these, except one representative, a series of single amino acid mutants from *Escherichia coli* was abandoned because using all of them would not result in further improvements in the modeling. The resulting set of 28 catalases from different species was used for the modeling.

Ranking the templates by structural alignment

This procedure was intended to choose the most suitable templates to be used in the final multi-chain-multitemplate homology modeling of the homotetramers. Homology models of the monomeric chain of the target were built with every template and then the templates were structurally aligned with the corresponding models. The calculated SO values express how well the target can be modeled with the different templates and will be used to rank both the models and the suitability of the templates for modeling the target.

Comparison of the different modeling strategies

Homology modeling when utilizing only the protein chains of the templates resulted in many low quality models according to the SO values (Table 3, Appendices 2–4). The involvement of the heme rings, as physically well established constraints, resulted in a great

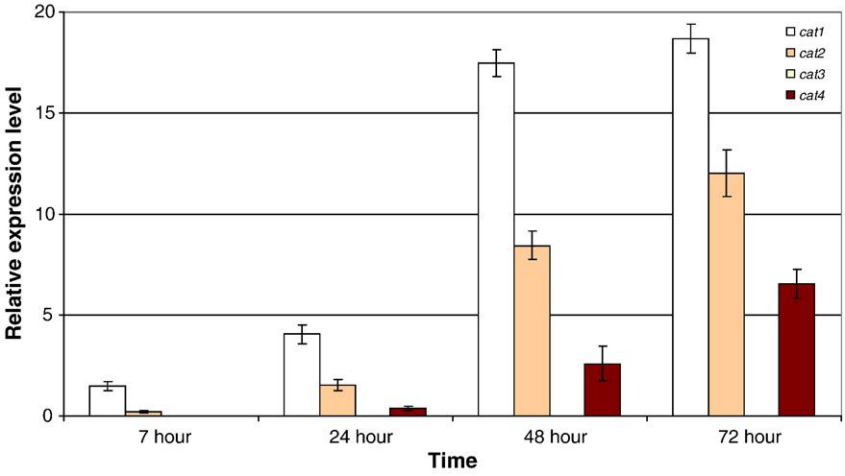

Fig. 4. Relative expression levels of the four catalase genes obtained by qPCR.

Table 3
Percent structural overlap between the homology models and the templates.

			PDB ID	Length (aa)	Monomer			Tetramer			Tetramer + symmetry		
					Chain only	+ hem	+ loop modeling	Chain only	+ hem	+ loop modeling	Chain only	+ hem	+ loop modeling
Bos taurus (cattle)	Purwar et al. (2011).	1	3RGP_A	499	91.73	98.34	98.96	98.96	98.96	98.96	98.96	98.96	98.96
Bos taurus (cattle)	Fita and Rossmann (1985).	2	7CAT_A	506	90.49	98.14	98.76	98.44	98.76	98.76	98.76	98.76	98.76
Bos taurus (cattle)	Foroughi et al. (2011).	3	3NWL_A	527	92.56	97.75	98.96	98.96	98.96	98.96	98.96	98.96	98.96
Human erythrocyte	Putnam et al. (2000).	4	1DGH_A	498	93.38	98.55	98.76	98.655	98.65	98.655	98.65	98.65	98.65
Human erythrocyte	Putnam et al. (2000).	5	1DGG_A	497	93.38	98.76	98.76	98.76	98.76	98.76	98.76	98.76	98.76
Human erythrocyte	Ko et al. (2000).	6	1QQW_A	527	91.73	98.76	98.96	99.01	99.01	99.01	99.01	99.01	99.01
Human erythrocyte	Putnam et al. (2000).	7	1DGH_B	498	93.38	98.55	98.76	98.65	98.65	98.65	98.65	98.65	98.65
Enterococcus. faecalis	Frankenberg et al. (2002).	8	1SI8_A	484	**94.93**	**99.78**	**99.78**	**99.78**	**99.78**	**99.78**	**99.78**	**99.78**	**99.78**
Pichia angusta	Penya-Soler et al. (2011).	9	2XQ1_A	509	92.56	96.69	98.14	98.24	98.08	98.038	98.29	91.21	93.43
Helicobacter pylori	Loewen et al. (2004).	10	1QWL_A	505	**91.73**	**95.24**	**99.79**	**99.79**	**99.79**	**99.79**	**99.79**	**99.79**	**99.79**
Helicobacter pylori	Loewen et al. (2004)	11	2A9E_A	505	91.11	93.59	98.96	98.96	98.96	98.96	98.96	98.96	98.96
Proteus mirabilis	Andreoletti et al. (2003).	12	1H6N_A	484	**91.78**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Proteus mirabilis	Andreoletti et al. (2003).	13	1M85_A	484	**92.42**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Vibrio salmonicida	Riise et al. (2007).	14	2ISA_A	483	**91.47**	**99.16**	**99.58**	**99.58**	**99.58**	**99.58**	**99.58**	**99.58**	**99.58**
Proteus mirabilis	Andreoletti et al. (2003).	15	1E93_A	484	**92.42**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Proteus mirabilis	Andreoletti et al. (2009).	16	3HB6_A	484	**92.64**	**99.15**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Proteus mirabilis	Andreoletti et al. (2001).	17	1H7K_A	483	**92.21**	**99.15**	**99.78**	**99.78**	**99.78**	**99.78**	**99.78**	**99.78**	**99.78**
Saccharomyces cerevisiae	MateÂ et al. (1999b).	18	1A4E_A	488	97.72	97.52	98.76	98.76	98.65	98.76	98.70	98.76	98.76
Exiguobacterium oxidotolerans	Hara et al. (2007).	19	2J2M_A	491	**91.66**	**99.37**	**99.79**	**99.79**	**99.79**	**99.79**	**99.79**	**99.79**	**99.79**
Pseudomonas syringae	Carpena et al. (2003).	20	1M7S_A	484	91.94	97.1	98.14	97.925	97.925	97.92	98.03	98.08	97.92
Micrococcus luteus	Murshudov et al. (2002).	21	1GWH_A	503	91.11	96.9	99.17	99.38	99.38	99.23	99.38	99.38	99.17
Micrococcus luteus	Murshudov et al. (1992).	22	1HBZ_A	498	90.9	94.42	99.17	99.32	99.32	99.17	99.38	99.38	99.17
Micrococcus luteus	Murshudov et al. (2002).	23	1GWF_A	503	89.25	90.7	98.14	98.34	98.34	98.445	98.34	98.34	98.39
Escherichia coli K-12	Jha et al. (2011).	24	3P9S_A	753	–	–	–	–	–	–	–	–	–
Escherichia coli K-12	Jha et al. (2011).	25	3P9R_A	754	–	–	–	–	–	–	–	–	–
Escherichia coli	MateÂ et al. (1999a).	26	1QF7_A	753	–	–	–	–	–	–	–	–	–
Escherichia coli K-12	Jha et al. (2011).	27	3P9P_A	753	–	–	–	–	–	–	–	–	–
Escherichia coli	Bravo et al. (1995).	28	1IPH_A	753	75.61	75.58	84.25	75.56	80.625	86.36	77.99	78.81	86.62
Escherichia coli	MateÂ et al. (1999a).	29	1CF9_A	753	–	–	–	–	–	–	–	–	–
Escherichia coli	Chelikani et al. (2003).	30	1P81_A	753	–	–	–	–	–	–	–	–	–
Escherichia coli	Chelikani et al. (2003).	31	1QWS_A	753	–	–	–	–	–	–	–	–	–
Escherichia coli	Melik-Adamyan et al. (2001).	32	1GGK_A	753	–	–	–	–	–	–	–	–	–
Escherichia coli	Chelikani et al. (2003).	33	1P7Z_A	753	–	–	–	–	–	–	–	–	–
Escherichia coli	Chelikani et al. (2003).	34	1P80_A	753	–	–	–	–	–	–	–	–	–
Escherichia coli	Melik-Adamyan et al. (2001).	35	1GG9_A	753	–	–	–	–	–	–	–	–	–
Escherichia coli	Melik-Adamyan et al. (2001).	36	1GGJ_A	753	–	–	–	–	–	–	–	–	–

(continued on next page)

Table 3 (*continued*)

			PDB ID	Length (aa)	Monomer			Tetramer			Tetramer + symmetry		
					Chain only	+ hem	+ loop modeling	Chain only	+ hem	+ loop modeling	Chain only	+ hem	+ loop modeling
Escherichia coli	Chelikani et al. (2003).	37	1P7Y_A	753	–	–	–	–	–	–	–	–	–
Escherichia coli K-12	Jha et al. (2011).	38	3P9Q_A	753	–	–	–	–	–	–	–	–	–
Escherichia coli	Melik-Adamyan et al. (2001).	39	1GGH_A	753	–	–	–	–	–	–	–	–	–
Neurospora crassa	Díaz et al. (2009).	40	3EJ6_A	688	81.61	83.88	93.38	88.38	89.92	93.5	89.87	89.66	94.26
Neurospora crassa	Díaz et al. (2004).	41	1SY7_A	715	84.5	88.84	94	87.80	89.87	92.198	88.16	89.46	92.76
Penicillium janthinellum	Borovik et al. (2011).	42	2XF2_A	688	80.78	79.13	93.38	91.11	91.11	93.69	91.67	91.67	94.26
Penicillium vitale	Alfonso-Praetor et al. (2007).	43	2IUF_A	688	77.89	77.89	92.97	90.38	90.38	93.12	90.07	90.07	94.15

Templates used for later homotetramer modeling of CAT1 are in boldface.
Quality of the homology models ranked on a 0–100 scale by the "iterative_structural_align" module.
PDB: code of the protein in the protein structure database.
Monomer: **chain only**: monomer model without heme, **+ hem**: monomer model in the presence of heme, **+ loop modeling**: monomer model with loop modeling in the presence of heme.
Tetramer: **chain only**: tetramer model without heme, **+ hem**: tetramer model in the presence of heme; **+ loop modeling**: tetramer model with loop modeling in the presence of heme.
Tetramer + symmetry: **chain only**: tetramer model with symmetry constraint and without heme, **+ hem**: tetramer model with symmetry constraint in the presence of heme; **+ loop modeling**: tetramer model with symmetry constraint and loop modeling in the presence of heme.

improvement of the models. It is important to note that the presence of heme not only improved the structural alignment between the model and its template but also changed the rank order between the templates for the specific target. The models were further improved by performing loop modeling as well, although the rank order of the templates did not change any more. The quality of the models obtained by the different modeling methods was characterized by the SO values and summarized in Table 3. The evolution of the model quality corresponding to the modeling strategies is shown in Fig. 5A. The structure obtained by using only the protein

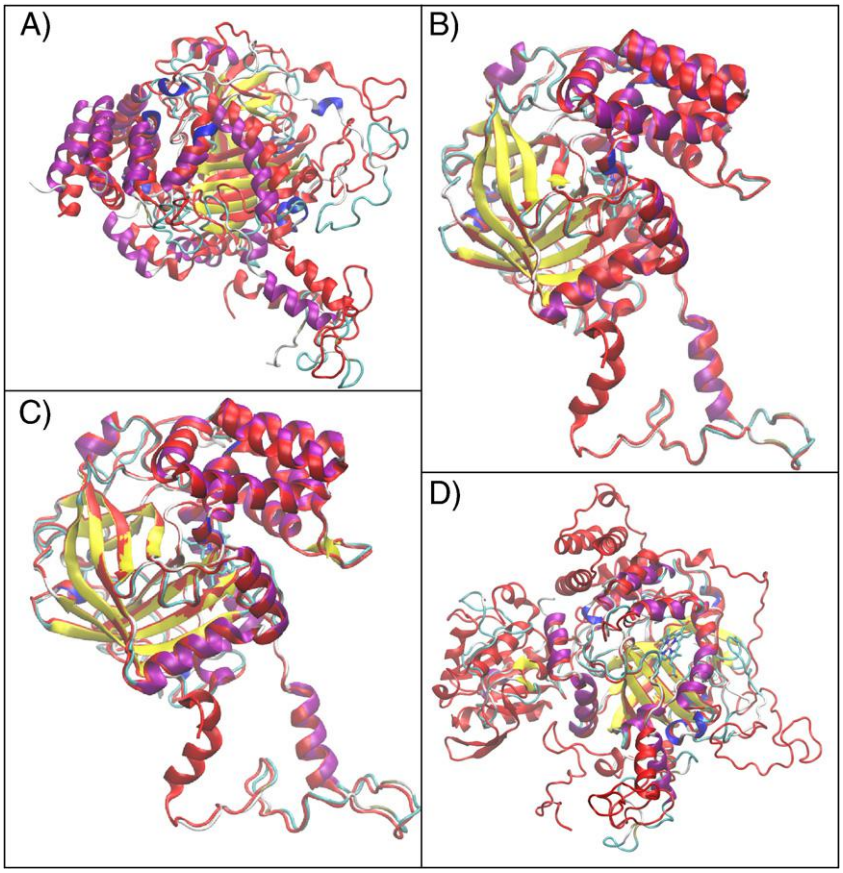

Fig. 5. A: Homology model of CAT1 without heme (template is in red); B: homology model of CAT1 with heteroatoms (heme) (template is in red); C: homology model by loop modeling in the presence of heme (template is in red); D: homology model of CAT1 using template 1IPH, a 200 residue longer catalase from *E. coli*.

Table 4
The templates selected by BLAST and by structural alignment.

R. oryzae protein	Length (aa)	BLAST score	PDB-ID	Length (aa)	Selection by structural alignment		Length (aa)
					Organisms	PDB-ID	
CAT1	484	*Bos taurus*	3RGP	499	*Helicobacter pylori*	1QWL	505
					Proteus mirabilis	1H6N	484
CAT2	710	*Escherichia coli* K-12	3P9S	753	*Neurospora crassa*	3EJ6	688
					Penicillium janthinellum	2XF2	688
CAT3	718	*Penicillium janthinellum*	2XF2	688	*Penicillium vitale*	2IUF	688
					Penicillium janthinellum	2XF2	688
CAT4	710	*Escherichia coli* K-12	3P9S	753	*Neurospora crassa*	3EJ6	688
					Micrococcus luteus	1GWF	503

chain of the template has large fragments which do not overlap with the template. However, in the presence of heme even the farther, outward parts of the protein chain aligned much better with the template (Fig. 5B). Using the loop modeling algorithm in the presence of heme resulted in further improvement in the structural alignment (Fig. 5C). The models are available upon request.

According to this measure i.e. the SO values from structural alignment, the homology modeling resulted in a different rank order of the templates for CAT1, CAT2 and CAT4 compared to that obtained by the PSI-BLAST sequence similarity search (Table 4). For example, the template 3RGP (*Bos taurus*) was best scored by BLAST for CAT1 while, after loop modeling, the homology model was given a value of 97.33, behind templates 1M85 and 3HB6 (*Proteus mirabilis*) with a value of 100. The order of the templates by the BLAST search and structural alignment was only identical for CAT3. Since the aim of the study is to obtain the 3D structures of the catalases of *R. oryzae*, the homology modeling of the homotetramers was performed with the templates ranked highest by the preliminary structural alignment of the homology models using the monomers. Fig. 5D shows the homology model of CAT1 based on 1IPH, a 200 residue longer template (*E. coli*). Although the highly conserved domains are fitted quite well, the use of proteins with such a longer chain resulted in models with a decreased accuracy.

To obtain the functional homotetramer form of the catalases the four chains should be modeled together. Modeling the homotetramer of CAT1 by the use of a single template (3RGP) without the assumption of identical chains resulted in slightly different geometry of the chains as seen in Fig. 6A. Applying the symmetry constraint (uniformity of the chains) resulted in a structure ordered slightly better (Fig. 6B). In both cases the heme molecules were included in all monomer units and the loop modeling algorithm was applied. The homotetramer models obtained with the highest SO value with and without the symmetry constraint are shown in Figs. 6D and C, respectively. As expected, the symmetry enforcement with loop modeling and the presence of heme resulted in a better model.

The homology models of the other *R. oryzae* catalases, CAT2, CAT3 and CAT4 obtained by the same strategies were usually rated by lower SO values compared with those of CAT1. Additional loop modeling usually resulted in weaker models where the templates were shorter. This is possibly the consequence of the 200–220 residue shorter chains of the templates due to which long chains of the target were modeled without the corresponding fragments of the template. Unexpectedly, the tetramer models without heme obtained higher SO values than the monomer models. The presence of heme and the symmetry constraint, however, improved the quality of the models similarly to the case of CAT1. Only three templates (3EJ6, 1HBZ,

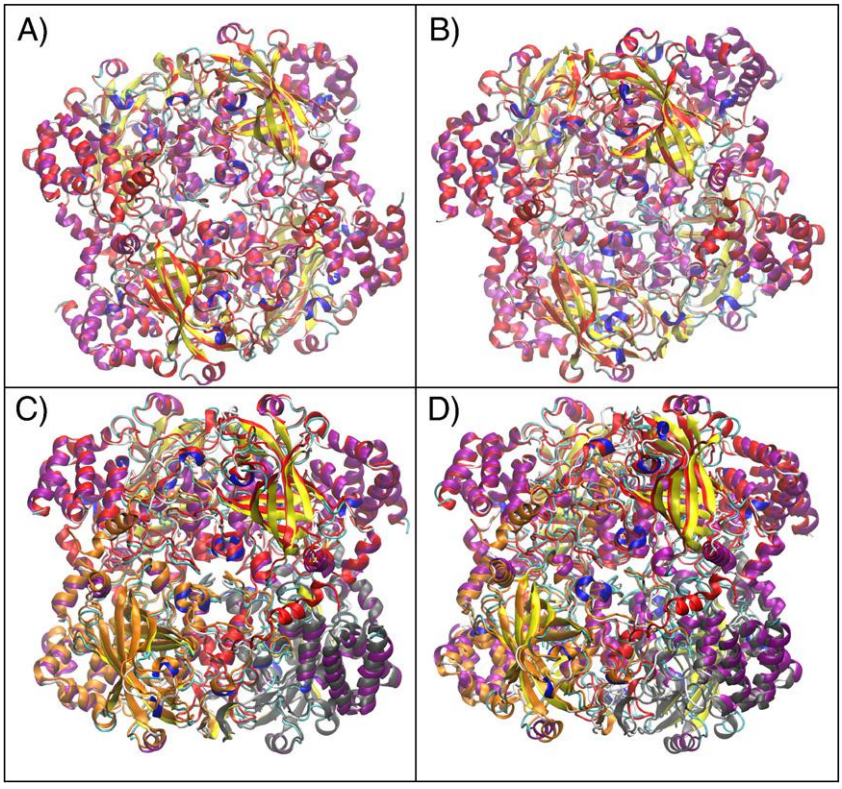

Fig. 6. A: Homology model of CAT1 tetramer using template 3RGP without symmetry constraint; B: homology model of CAT1 tetramer using template 3RGP with symmetry constraint; C: homology model of CAT1 tetramer using 9 templates without symmetry constraint; D: homology model of CAT1 tetramer using 9 templates in the presence of heme with symmetry constraint and loop modeling.

1GWF) resulted in analogous improvements of the models using the different modeling strategies as obtained for the models of CAT1. It may be explained with the similar length of the targets and the templates: the lengths of CAT2, CAT3 and CAT4 are between 710 and 718 while most templates are in the range of 477 to 503, expect 3EJ6, 1HBZ and 1GWF to have a length between 689 and 736. In the opposite case modeling CAT1, where the protein was fitted with longer templates, a similar incompatibility was not experienced.

The most important parts of the 3D models are the active centers, i.e. the immediate environment of the heme molecule which is shown in Fig. 7. These amino acids are the distal His and the proximal Tyr coordinating the Fe ion, an Arg intercalating between the carboxyl groups of heme and four hydrophobic residues around the parallel edge of the heme ring (Asn, Phe, Phe and Met). CAT1 have an equivalent arrangement of the residues than that found in 3RGP (Figs. 7B and A, respectively). CAT2 and CAT4 have a Leu in place of Met (Figs. 7C and E, respectively). In CAT3 the Asn and the closer Phe are absent due to the missing 11 residue long fragment being present in other catalases and there is an Asp in place of Met (Fig. 7D). The distance between the ε—N atom of the distal His and the Fe atom of heme was reproduced in the best rated models in Fig. 7 with acceptable error. The values obtained in angstroms for CAT1, CAT2, CAT3, CAT4 and the corresponding templates, 3EJ6, 1M85, 2IUF, 3EJ6 were 4.77, 4.92, 4.87, 4.67 and 4.83, 4.79, 4.40, 4.79, respectively. This distance in 3RGP is 4.66 Å.

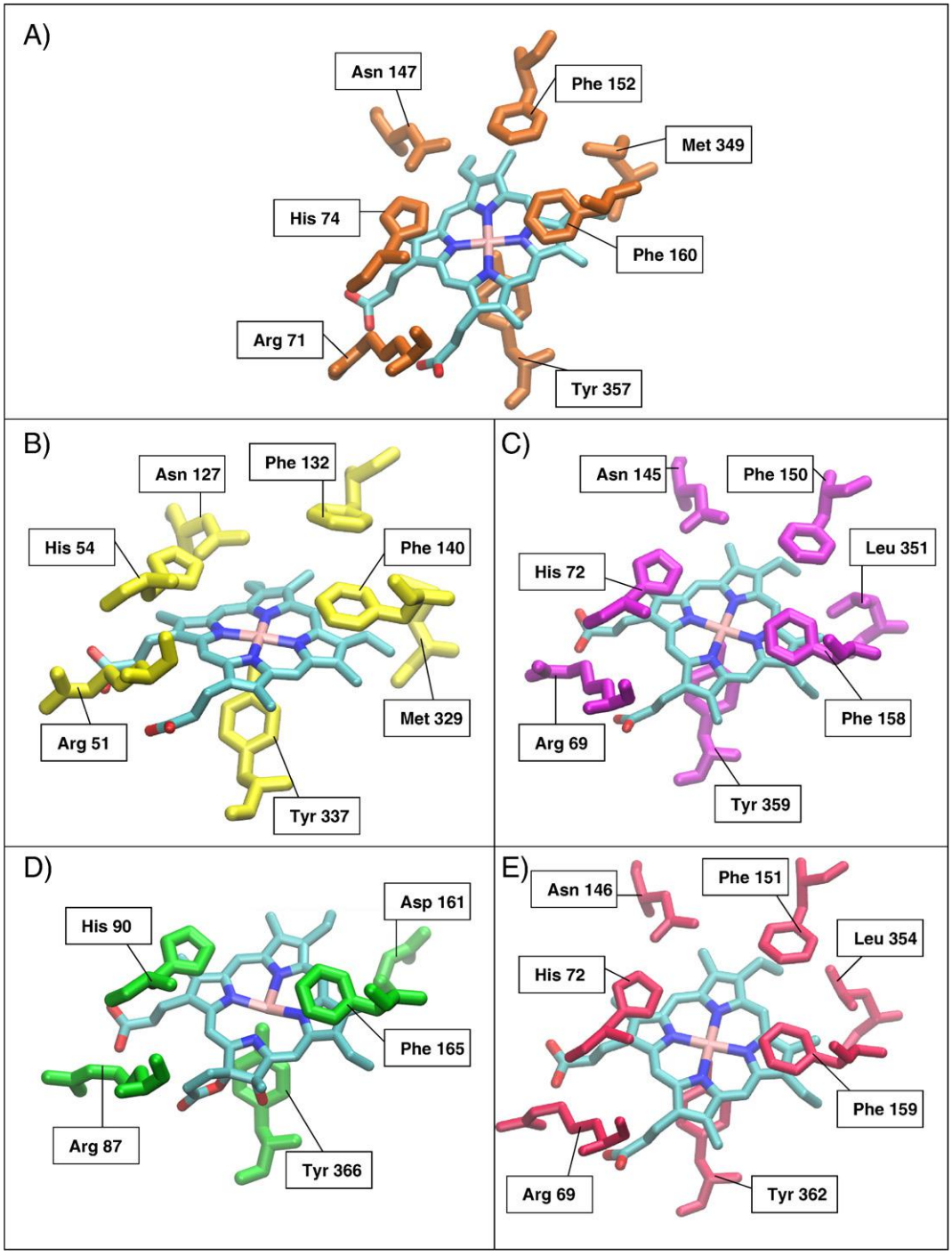

Fig. 7. The active center of catalases. A: *B. taurus* (PDB ID: 3RGP), B: *R. oryzae* CAT1, C: *R. oryzae* CAT2, D: *R. oryzae* CAT3, E: *R. oryzae* CAT4.

Real-time quantitative PCR (qPCR) analysis

The expression of the four catalase genes of *R. oryzae* was investigated in different life cycles. The actin gene was used as an internal standard and expression levels of the four catalase genes at a 4 h cultivation time were used as biological controls. RNA samples were extracted at 4 and 7 h after the inoculation when sporangiospores started to germinate and hyphae started to grow, respectively. Further samples were taken at 1, 2 and 3 days of life. Cat1 was the highest activity gene in all samples. At 7 h cat2 was induced and in the later samples the expression of cat4 was also observed. The highest transcription level was observed for all three genes at the 3 days of life time. The transcription activity of cat1 was ca. 2.5 and 3.5 times higher than that of cat2 and cat4, respectively. The cat3 gene did not show a significant expression level in any life cycles investigated (Fig. 4).

Discussion

The 28 experimental catalase structures were obtained by a PSI-BLAST search of the NCBI database and used for homology modeling of the catalases of *R. oryzae* and to develop an efficient modeling methodology as well. The target proteins were modeled with different modeling methods involving loop modeling and structurally important constraints such as the cofactor heme, symmetry of the chains. Instead of visual comparison the quality of the models was measured by the percent structural overlapping values. In the case of monomers the presence of heme caused a great improvement in the quality of the models and a further refinement was observed by loop modeling if the template had an equal or longer protein chain. If the templates were shorter than the target i.e. a large-subunit catalase was modeled with a small-subunit one, the quality of the models were obviously worse and the loop modeling further decreased the model quality. The flanking fragments without the corresponding parts of the template were not modeled correctly, although the highly conserved small subunit part was modeled quite well. In the latter case, however, an unexpected improvement was observed when the tetramers were modeled, in the presence of all four protein chains, heme and symmetry option. This implies the assumption that the multiple protein chains in the tetramer modeling may be considered additional constraints to help the packing of the chains mutually and expected to improve the model quality especially together with the symmetry constraint. According to the above assumption the best modeling algorithm involves multiple chains (tetramer), cofactor (heme), symmetry option and loop modeling and the quality of the models obtained in this manner was rated by the SO values.

The structural alignment of proteins from different species is a tool of increasing importance in structural biology to establish evolutionary and/or functional relationships (Wilson et al., 2009; Madhusudhan et al., 2009). Our homology modeling strategy resembles to this procedure in that the catalases of *R. oryzae* were modeled using a series of templates and then the best model was selected by the quantitative measure of the pairwise structural alignment of each model with the corresponding template. Assuming that the suitability of the template to model the protein may be dependent of the evolutionary distance between them, the best homology model can be built by the most related template. In other words, evolutionary relationships may also be concluded between the proteins modeled and the templates using the SO values similar to the case of the structural alignment of experimental structures. CAT1 of *R. oryzae* is the member of clade 3 in the phylogenetic tree of the fungi catalases and it is a possible peroxisomal catalase. Genes CAT2, CAT3 and CAT4 belong to clade 2 and they are close to each other within the tree.

Using to the SO values above 99%, the classification of the corresponding proteins by structural alignment of the homology models with the templates led to similar results (Table 4). The 3D structures obtained for CAT1 show the highest similarity to the small-subunit bacterial catalases. CAT2, CAT3 and CAT4 showed higher similarity to the large-subunit fungi catalases. Furthermore, CAT2 and CAT4 showed a somewhat decreased similarity to the small-subunit bacterial catalases contrary to CAT3 which is in agreement with the fact that CAT2 and CAT4 are observed on sister arms of the phylogenetic tree separated from CAT3 indicating their common origin. CAT3 belongs to the large-subunit catalases with an SO value of 97% related to catalases of two Penicillium and a Neurospora species. The high similarity of the *R. oryzae* catalases to the bacterial ones is in harmony with the proposition of Klotz et al. (Klotz et al., 1997; Klotz and Loewen, 2003) assuming a lateral gene transfer from prokaryotes to eukaryotes, based on a phylogenetic analysis of the catalases of prokaryotes and eukaryotes. This is also supported by the result of the PSI-BLAST search which found CAT1 more similar to the catalases of higher order livings. The different rank order of the templates obtained by the homology modeling may be the consequence of the fact that the protein structures are usually more conserved than the amino acid sequences which is necessary to preserve the functionality. It is worth to note, however, that CAT2 and CAT4 showed higher similarity to the small-subunit bacterial catalases than to the large fungal ones in a tetramer modeling method without symmetry while CAT3 still resembled the large-subunit fungal catalases as can be seen in Appendices 2–4. (Without the symmetry option the models based on shorter and longer templates were comparable due to the avoidance of the disadvantage of loop modeling observed for the short template/long target cases.)

To investigate the active centers of the four *R. oryzae* catalases, those models were selected which possessed the highest SO values. The core of the monomers i.e. the active center shows a characteristic, extremely well conserved 3D arrangement of the amino acids which is supported by the fact that the SO value of the models of CAT2, CAT3 and CAT4 were near 100% even when the templates were more than 200 residue shorter (Appendices 2–4). The variations in the conserved amino acids of the active center may be related to the results of qPCR experiments of the catalase genes of *R. oryzae* where the least active one was cat3 (Fig. 4). Compared to the catalase from *B. taurus* (PDB code 3RGP) CAT1 has identical amino acids, in CAT2 and CAT4 there is only a Met-Leu mutation while CAT3 lacks two amino acids besides the Met-Asp mutation. NADPH is known to have a protective effect preventing the oxidative degradation of catalases and besides the essential His, a Met near the active center also plays an important role in maintaining the catalase functionality (Andreoletti et al., 2009). According to our results none of the four *R. oryzae* catalases is supposed to bind the NADPH including CAT1, although it is a small-subunit catalase. Only CAT1 has a Met in the active center, CAT2 and CAT4, the other two active enzymes do not. The lack of the protective NADPH cofactor and Met in two out of the three active catalases may offer targets in a future drug design.

Conclusion

The quality of the homology models were greatly improved by applying constraints for the protein chains such as cofactors, multiple chains, symmetry to obtain identical chains and by additional loop modeling. The maximal structural overlap between the model and the template was observed with the phylogenetically closest templates. According to this, more accurate models can be built with templates selected by phylogenetic relationships than by a PSI-BLAST search. The phylogenetic relationships take in effect when all possible constraints were applied in the modeling algorithm. However, not only the maximal model accuracy can be achieved in this manner, but the resulting models may revert to their closest phylogenetic relatives too. Consequently, accurate models were obtained for the *R. oryzae* catalases representing the highly conserved catalytic center except

CAT3 which has a missing fragment and multiple mutations in the essential amino acids that may be related to the low expression of CAT3.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.lfs.2012.06.016.

Conflict of interest statement

The content of the manuscript has no conflicting interests with funding from any persons or institutions.

Acknowledgements

This work was supported by grants of the Hungarian Scientific Research Fund (OTKA CK78566), the Hungarian Scientific Research Fund and the National Office for Research and Technology (OTKA CK80188) and The National Innovation Office ERC_HU_09 3D_TRPV1. The work was also supported by the Project named "TÁMOP-4.2.1/B-09/1/KONV-2010-0005 - Creating the Center of Excellence at the University of Szeged" supported by the European Union and co-financed by the European Regional Development Fund.

References

Alfonso-Praetor M, Borovik A, Carpena X, Murshudov G, Melik-Adamyan W, Fita I, et al. The structures and electronic configuration of compound I intermediates of *Helicobacter pylori* and *Penicillium vitale* catalases determined by X-ray crystallography and QM/MM density functional theory calculations. J Am Chem Soc 2007;129:4193–205.

Andreoletti P, Gambarelli S, Sainz G, Stojanoff V, White C, Desfonds G, et al. Formation of a tyrosyl radical intermediate in *Rhizopus oryzae* catalase by directed mutagenesis and consequences for nucleotide reactivity. Biochemistry 2001;40:13734–43.

Andreoletti P, Sainz G, Jaquinod M, Gagnon J, Jouve HM. High-resolution structure and biochemical properties of a recombinant *Proteus mirabilis* catalase depleted in iron. Struct Funct Genet 2003;50:261–71.

Andreoletti P, Mouesca JM, Gouet P, Jaquinod M, Capeillère-Blandin C, Jouve HM. Verdoheme formation in *Rhizopus oryzae* catalase. Biochim Biophys Acta 2009;1790:741–53.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 2004;340:783–95.

Bernroitner M, Zamocky M, Furtmüller PG, Peschek GA, Obinger C. Occurrence, phylogeny, structure, and function of catalases and peroxidases in cyanobacteria. J Exp Bot 2009;60(2):423–40.

Borovik A, Grebenko AI, Melik-Adamyan WR. X-ray investigation of *Penicillium vitale* catalase inhibited by aminotriazole. Crystallogr Rep 2011;56:590–5.

Bravo J, Verdaguer N, Tormo J, Betzel C, Switala J, Loewen PC, et al. Crystal structure of catalase HPII from *Escherichia coli*. Structure 1995;15:491–502.

Carpena X, Soriano M, Klotz MG, Duckworth HW, Donald LJ, Adamyan WM, et al. Structure of the clade 1 catalase, CatF of *Pseudomonas syringae*, at 1.8 Å resolution. Structure Function and Genetics 2003;50:423–36.

Chelikani P, Carpena X, Fita I, Loewen PC. An electrical potential in the access channel of catalases enhances catalysis. J Biol Chem 2003;278:31290–6.

Díaz A, Horjales E, Rudiño-Piñera E, Arreola R, Hansberg W. Unusual Cys–Tyr covalent bond in a large catalase. J Mol Biol 2004;342:971–85.

Díaz A, Valdés VJ, Rudiño-Piñera E, Horjales E, Hansberg W. Structure–function relationships in fungal large-subunit catalases. J Mol Biol 2009;3861:218–32.

Eucker J, Sezer O, Graf B, Possinger K. Mucormycoses. Mycoses 2001;44:253–60.

Fita I, Rossmann MG. The NADPH binding site on beef liver catalase. Biochemistry 1985;82:1604–8.

Foroughi LM, Kang YN, Matzger AJ. Polymer-induced heteronucleation for protein single crystal growth: structural elucidation of bovine liver catalase and concanavalin a forms. Cryst Growth Des 2011;1:1294.

Frankenberg L, Brugna M, Hederstedt L. *Enterococcus faecalis* heme-dependent catalase. J Bacteriol 2002;184:6351–6.

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein identification and analysis tools on the ExPASy Server. In: Walker JM, editor. The Proteomics Protocols Handbook, 52. Humana Press; 2005. p. 571–607.

Gould SJ, Keller G-A, Subramani S. Identification of peroxisomal targeting signals located at the carboxy terminus of four peroxisomal proteins. J Cell Biol 1988;107:897–905.

Hara I, Ichise N, Kojima K, Kondo H, Ohgiya S, Matsuyama H, et al. Relationship between the size of the bottleneck 15 Å from iron in the main channel and the reactivity of catalase corresponding to the molecular size of substrates. Biochemistry 2007;46:11–22.

Hemashettar BM, Patil RN, O'Donnell K, Chaturvedi V, Ren P, Padhye AA. Chronic rhinofacial mucormycosis caused by *Mucor irregularis* (*Rhizomucor variabilis*) in India. J Clin Microbiol 2011;49:2372–5.

Humphrey W, Dalke A, Schulten K. VMD — visual molecular dynamics. J. Mol. Graphics 1996;14:33–8.

Jha V, Louis S, Chelikani P, Carpena X, Donald LJ, Fita I, et al. Modulation of heme orientation and binding by a single residue in catalase HPII of *Escherichia coli*. Biochemistry 2011;50:2101–10.

Klotz MG, Loewen PC. The molecular evolution of catalytic hydroperoxidases: evidence for multiple lateral transfer of genes between prokaryota and from bacteria into eukaryota. Mol Biol Evol 2003;20:1098–112.

Klotz MG, Klassen GR, Loewen PC. Phylogenetic relationships among prokaryotic and eukaryotic catalases. Mol Biol Evol 1997;14:951–8.

Ko TP, Safo MK, Musayev FN, Di Salvo ML, Wang C, Wu SH, et al. Structure of human erythrocyte catalase. Acta Crystallogr D Biol Crystallogr 2000;56:241–5.

Kolotila MP, Diamond RD. Effects of neutrophils and in vitro oxidants on survival and phenotypic switching of *Candida albicans* WO-1. Infect Immun 1990;58:1174–9.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK — a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–91.

Liu T, Eisenberg YD. 3D domain swapping: as domains continue to swap. Protein Sci 2002;11:1285–99.

Liu T, Tang GW, Capriotti E. Comparative modeling: the state of the art and protein drug target structure prediction. Comb Chem High Throughput Screen 2011;14:1386–2073.

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using realtime quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. Methods 2001;25:402–8.

Loewen PC, Carpena X, Rovira C, Ivancich A, Perez-Luque R, Haas R, et al. Structure of *Helicobacter pylori* catalase, with and without formic acid bound, at 1.6 Å Resolution. Biochemistry 2004;43:3089–103.

Madhusudhan MS, Webb BM, Marti-Renom MA, Eswar N, Sali A. Alignment of multiple protein structures based on sequence and structure features. Protein Eng Des Sel 2009;22:569–74.

Maté MJ, Sevinc MS, Hu B, Bujons J, Bravo J, Switala J, et al. Mutants that alter the covalent structure of catalase hydroperoxidase II from *Escherichia coli*. J Biol Chem 1999a;274:27717–25.

Maté MJ, Zamocky M, Nykyri LM, Herzog C, Alzari PM, Betzel C, et al. Structure of catalase—a from *Saccharomyces cerevisiae*. J Mol Biol 1999b;268:135–49.

Melik-Adamyan W, Bravo J, Carpena X, Switala J, Maté MJ, Fita I, et al. Substrate flow in catalases deduced from the crystal structures of active site variants of HPII from *Escherichia coli*. Proteins 2001;44:270–81.

Murshudov GN, Melik-Adamyan WR, Grebenko AI, Barynin VV, Antson AA, Vainshtein BK, et al. Three-dimensional structure of catalase from *Micrococcus lysodeikticus* at 1.5 Å resolution. FEBS 1992;312:127–31.

Murshudov GN, Grebenko AI, Brannigan JA, Antson AA, Barynin VV, Dodson GG, et al. The structures of *Micrococcus lysodeikticus* catalase, its ferryl intermediate (compound II) and NADPH complex. Acta Crystallogr 2002;58:1972–82.

Pagni M, Ioannidis V, Cerutti L, Zahn-Zabal M, Jongeneel CV, Hau J, et al. MyHits: improvements to an interactive resource for analyzing protein sequences. Nucleic Acids Res 2007;35:W433–7.

Penya-Soler E, Vega MC, Wilmanns M, Williams CP. Structural features of peroxisomal catalase from the yeast *Hansenula polymorpha*. Acta Crystallogr Sect D 2011;67:690.

Purwar N, McGarry JM, Kostera J, Pacheco AA, Schmidt M. Interaction of nitric oxide with catalase: structural and kinetic analysis. Biochemistry 2011;50:4491–503.

Putnam CD, Arvai1 AS, Bourne Y, Tainer JA. Active and inhibited human catalase structures: ligand and NADPH binding and catalytic mechanism. J Mol Biol 2000;296:295–309.

Rhizopus oryzae Sequencing Project, Broad Institute of Harvard and MIT. http://www.broadinstitute.org/annotation/genome/rhizopus_oryzae/MultiHome.html 2004.

Ribes JA, Vanover-Sams CL, Baker DJ. Zygomycetes in human disease. Clin Microbiol Rev 2000;13:236–301.

Riise KE, Lorentzen MS, Helland R, Smalås AO, Leiros HKS, Willassen NP. The first structure of a cold-active catalase from *Vibrio salmonicida* at 1.96 Å reveals structural aspects of cold adaptation. Acta Crystallogr 2007;63:135–48.

Ronquist F, Huelsenbeck JP. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 2003;19:1572–4.

Roshan U, Livesay DR. Probalign: multiple sequence alignment using partition function posterior probabilities. Bioinformatics 2006;22:2715–21.

Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.

Shibuya K, Paris S, Ando T, Nakayama H, Hatori T, Latgé JP. Catalases of *Aspergillus fumigatus* and inflammation in aspergillosis. Jpn J Med Mycol 2006;47:249–55.

Wilson D, Pethica R, Zhou Y, Talbot C, Vogel C, Madera M, et al. SUPERFAMILY—sophisticated comparative genomics, data mining, visualization and phylogeny. Nucleic Acids Res 2009;37:D380–6.

Zamocky M, Koller F. Understanding the structure and function of catalases: clues from molecular evolution and in vitro mutagenesis. Prog Biophys Mol Biol 1999;72:19–66.

Zamocky M, Furtmüller PG, Obinger C. Evolution of catalases from bacteria to humans. Antioxid Redox Signal 2008;10:1527–47.