Original article

The Laccase Engineering Database: a classification and analysis system for laccases and related multicopper oxidases

Demet Sirim, Florian Wagner, Lei Wang, Rolf D Schmid and Jürgen Pleiss*

Institute of Technical Biochemistry, University of Stuttgart, Allmandring 31, 70569 Stuttgart, Germany

*Corresponding author: Tel.: (+49) 711 685 63191; Fax: (+49) 711 685 63196 Email: juergen.pleiss@itb.uni-stuttgart.de

Laccases and their homologues form the protein superfamily of multicopper oxidases (MCO). They catalyze the oxidation of many, particularly phenolic substances, and, besides playing an important role in many cellular activities, are of interest in biotechnological applications. The Laccase Engineering Database (LccED, http://www.lcced.uni-stuttgart.de) was designed to serve as a tool for a systematic sequence-based classification and analysis of the diverse multicopper oxidase protein family. More than 2200 proteins were classified into 11 superfamilies and 56 homologous families. For each family, the LccED provides multiple sequence alignments, phylogenetic trees and family-specific HMM profiles. The integration of structures for 14 different proteins allows a comprehensive comparison of sequences and structures to derive biochemical properties. Among the families, the distribution of the proteins regarding different kingdoms was investigated. The database was applied to perform a comprehensive analysis by MCO- and laccase-specific patterns.

The LccED combines information of sequences and structures of MCOs. It serves as a classification tool to assign new proteins to a homologous family and can be applied to investigate sequence–structure–function relationship and to guide protein engineering.

Database URL: http://www.lcced.uni-stuttgart.de

Introduction

Multicopper oxidases (MCOs) catalyze the one-electron oxidation of their substrates with a concomitant four-electron reduction of molecular oxygen to water. MCOs consist of four enzyme families: laccases (EC 1.10.3.2), ascorbate oxidases (EC 1.10.3.3), ferroxidases (EC 1.16.3.1) and ceruloplasmin (EC 1.16.3.1). Functional studies have revealed that MCOs have two active sites: one blue type 1 (T1) copper site where the substrate is oxidized, and a trinuclear copper cluster [consisting of three type 2 (T2)/type 3 (T3) coppers] where oxygen is bound, activated and reduced (1). The electrons are transferred from the T1 site to the T2/T3 site via highly conserved amino acids which have previously been described in PROSITE (2, 3) as MCO-specific patterns,

further referred to as M2 and M4 (4, 5). In addition, laccase-specific signature sequences, namely L1 and L3, were generated from 100 plant and fungal laccase sequences. L1 and L3 have been suggested to be specific for laccases and were proposed to distinguish laccases from other MCOs (6). While there is only low overall sequence similarity, the structure and catalytic mechanism is conserved (7). Most MCOs consist of three cupredoxin domains, except for ceruloplasmin and some bacterial laccases which contain six or two domains, respectively (8). Depending on the number of domains, MCOs vary in size, from 300 to 1000 residues, and contain up to six copper ions (4).

Laccases, which constitute the largest subfamily of MCOs, are widely distributed among fungi, higher plants (9, 10), bacteria (11) and insects (12). In fungi, they are supposed to

be involved in lignin degradation (13), pigment production (14) and plant pathogenesis (15). In plants, their potential function is the biosynthesis of lignin (16). In bacteria, they are suggested to play a role in melanin production, spore coat resistance, morphogenesis and detoxification of copper (17). In particular laccases, which form the largest subgroup of MCOs, also have a high biotechnological potential as versatile catalysts in textile and in pulp and paper industries, as well as in food applications, bioremediation and organic synthesis (18, 19). However, their redox potential often is restricted and they react in a relatively non-specific way (20). Engineered laccases promise to have improved enzymatic properties such as activity, specificity and selectivity (21). It is expected that understanding the relationships between sequence, structure and function would greatly help the engineering of laccases. Therefore, we integrated data on MCO sequences and structures and built up the Laccase Engineering Database (LccED) using the data warehouse system DWARF (22). Previously, 350 MCOs were assigned to 10 superfamilies (23): (A) basidiomycete laccases, (B) ascomycete laccases, (C) insect laccases, (D) fungal pigment MCOs, (E) fungal ferroxidases, (F) fungal and plant ascorbate oxidases, (G) plant laccase-like MCOs, (H) copper resistance proteins (CopA), (I) bilirubin oxidases and (J) copper efflux (CueO) proteins. For the MCOs that lack the second domain, referred to as small laccases (SLAC), a distinct family was established based on SLAC from of *Streptomyces coelicolor* (24). Homologous MCO sequences were retrieved and assigned to families by sequence similarity. In order to assist comprehensive sequence analysis, reliable multisequence alignments were generated and annotated either by an automated pattern search or by information extracted from GenBank (25). In addition, family-specific HMM profiles (26) and a BLAST (27) interface are provided to allow an assignment of new sequences to families. Thus, the LccED is the first data resource that combines information on sequences, sequence alignments, annotations, and structures of MCOs.

Construction and content

Database construction

The LccED was established within the data warehouse system DWARF, which provides a data model for the integration of sequences and structures in a family-specific protein database, as well as tools for extracting and loading data from various data sources (22). Previously, more than 350 MCO sequences were assigned to 10 superfamilies (23). From this data set 248 sequences, for which a GenBank entry was available, were selected as seed sequences and assigned according to their initial classification by Hoegger *et al.* to the 10 superfamilies, which were named based on the origin of their seed sequences. An additional

superfamily was created for two domain laccases and SLAC from *Streptomyces coelicolor* was used as seed sequence for this family. Subsequently, for each seed sequence a BLAST search (27) was performed in the non-redundant sequence database at NCBI (http://ncbi.nlm.nih.gov) with an *E*-value of $E = 10^{-10}$. The *E*-value for the BLAST search was determined empirically. Therefore, for the more diverse bacterial families a higher *E*-value of $E = 10^{-5}$ was applied. Each BLAST hit was assigned to the superfamily of the respective seed sequence if the sequence identity was higher than 40%. Sequences within a superfamily were classified into homologous families based on multiple sequence alignments and phylogenetic trees, as calculated by CLUSTAL W (28). Information on source organism, sequence annotations and sequence was extracted by the sequence data loader of the DWARF system. The species information was adapted to the NCBI taxonomy. Different names denominating the same organism are listed as synonyms on the organism page. Sequences from the same source organism and sharing >98% identical residues were represented as one single protein entry. This assignment is implemented in an automated script, thus preventing that one protein from the same organism may occur in duplicate within the database and avoiding redundancy even if it may occur in GenBank.

In case of BLAST hits specifying a protein structure, the respective structure was extracted from the PDB (29), stored as structural monomers and secondary structure information was generated for each chain by DSSP (30). For all families, multiple sequence alignments were performed by CLUSTAL W (28) and manually checked to improve consistency and quality. Proteins which were not assignable to any superfamily and are not similar to the class of MCOs were removed from the database. BLAST hits which are proteins fragments or putative sequences which were longer than usual MCOs, or proteins with high sequence similarity to MCOs where chosen to be included if they locally align in the active site regions.

The LccED provides regular yearly updates using the automated update function of the DWARF system which applies a protein 'blacklist' of removed entries, keeping pace with the permanently growing GenBank data (25).

Contents

The LccED contains data on 2828 sequences and 2297 proteins. For 21 proteins from 10 different homologous families crystal structures are deposited, which results in a total of 82 structural monomers. The proteins were assigned to 11 superfamilies based on the origin of the seed sequences and to 56 homologous families based on phylogeny (Table 1). For each superfamily and homologous family, an annotated multiple sequence alignment, a phylogenetic tree and a family-specific HMM profile (http://hmmer.janelia.org/) were generated.

Table 1. LccED families, sequences and structures

Superfamily	Homologous families	Proteins	Structures
A (Basidiomycete Laccases)	4	201	13
B (Ascomycete Laccases)	6	421	6
C (Insect Laccases)	8	168	0
D (Fungal Pigment MCOs)	4	55	0
E (Fungal Ferroxidases)	6	117	6
F (Fungal and plant AOs)	6	137	8
G (Plant Laccases)	5	333	0
H (Bacterial CopA Proteins)	6	383	0
I (Bacterial Bilirubin Oxidases)	5	149	24
J (Bacterial CueO Proteins)	5	310	11
K (SLAC homologs)	1	18	3

Web interface

The LccED is publicly available on http://www.LccED.uni-stuttgart.de. It can be browsed by family, organism or structure. For each family, pre-calculated annotated multiple sequence alignments, phylogenetic trees and HMMs are provided. All protein entries in the alignments and trees are linked to their original NCBI entries. Functionally relevant amino acids are color coded, and further information is displayed upon moving the mouse over the respective residue in the multiple sequence alignment. The conservation degree of the alignment was calculated using PLOTCON (31). PFAM (32) links to all protein entries were added as far they were available. Further, the PFAM annotation to each protein can be accessed within the multiple sequence alignments by scrolling the mouse over the respective region. Phylogenetic trees are visualized by an in-house developed tree-visualizer which allows coloring each entry by properties such as homologous family (in superfamily-trees), organism, sequence length and kingdom of the source organism (Figure 1). Via a local BLAST interface, unknown MCO sequences can be classified by sequence similarity to the existing LccED entries. A tar-archive comprising all information on families, sequences, structures, multiple sequence alignments, trees and profiles can be downloaded.

Analysis of organism distribution and patterns

In this study, 2297 MCO proteins from a wide spectrum of source organisms were assigned to superfamilies and homologous families based on sequence similarity and phylogenetic analysis. A comprehensive analysis of the relationships between sequence similarity, source organism

and of patterns forming the binding sites of copper was performed in 2274 proteins of families A–K. The proposed patterns L1 (H-W-H-G-x(9)–D-G-x(5)–Q-C-P-I) and L3 (H-P-x-H-L-H-G-H) have been suggested to be specific for laccases, the patterns M2 (G-x-[FYW]-x-[LIVMFYW]-x-[CST]-x-{PR}-{K}-x(2)-{S}-x-{LFH}-G-[LM]-x(3)-[LIVMFYW], PROSITE entry PS00079) and M4 (H-C-H-x(3)-H-x(3)-[AG]-[LM], PROSITE entry PS00079) for MCOs (Figure 2). Pattern L1 includes one histidine which binds the T2 copper and one histidine which binds the T3 copper. Pattern M2 includes two further T3 copper ligands. Pattern L3 includes ligands of the T1, T2 and T3 coppers. Within pattern M4 three of the four ligands of the T1 centre and one T3 copper ligand are located (Figure 2). For annotation and evaluation purposes regular expressions generated from L1, L3, M2 and M4 were applied. Because of the short length of their sequences and the missing domain, proteins of the SLAC family were excluded from the analysis of the appearance of the pattern within the families (Supplementary Table S1) and the resulting numbers for false negatives (Supplementary Table S2).

Family A (Basidiomycete Laccases) contains exclusively fungal proteins, 91% are from basidiomycetes (homologous families A1–A4), 9% from ascomycetes (homologous family A2). Eighty percent are annotated as laccases in GenBank. Twenty-three percent of the proteins contain the pattern L1, 22% M2, 14% L3 and 46% M4. Seventeen percent of the entries are annotated as putative.

Family B (Ascomycete Laccases) contains 36% from ascomycetes. All of them cluster into the homologous family B1 and 62% are annotated as laccases in GenBank. The other proteins are all of bacterial origin (homologous families B2–B6) and 3% are annotated as laccases in GenBank. Yet, they show a considerable sequence identity of over 40% to ascomycetous laccases. Eighty-eight percent of the proteins contain the pattern L1, 92% M2, 49% L3 and 15% M4. Thirty-three percent of the entries are annotated as putative.

Family C (Insect Laccases) resulted in 78% proteins of insect origin (homologous families C1–C8). The remaining 22% consist of euechinoidea (in homologous family C1), cephalochordata and cnidaria (in homologous family C6). Thirty-eight percent are annotated as laccases in GenBank. Thirty percent of the proteins contain pattern L1, 75% M2, 75% L3 and 3% M4. Thirty-five percent of the entries are annotated as putative.

Family D (Fungal Pigment MCOs) contains exclusively fungal proteins. Thirty-six percent of the proteins are annotated in GenBank as laccases. Ninety percent of the proteins contain pattern L1, 78% M2, 82% L3 and 11% M4. Fifty-eight percent of the entries are annotated as putative.

Family E (Fungal Ferroxidases) contains exclusively fungal proteins. Seventeen percent of the proteins are annotated in GenBank as laccases. Eighty-three percent of the proteins
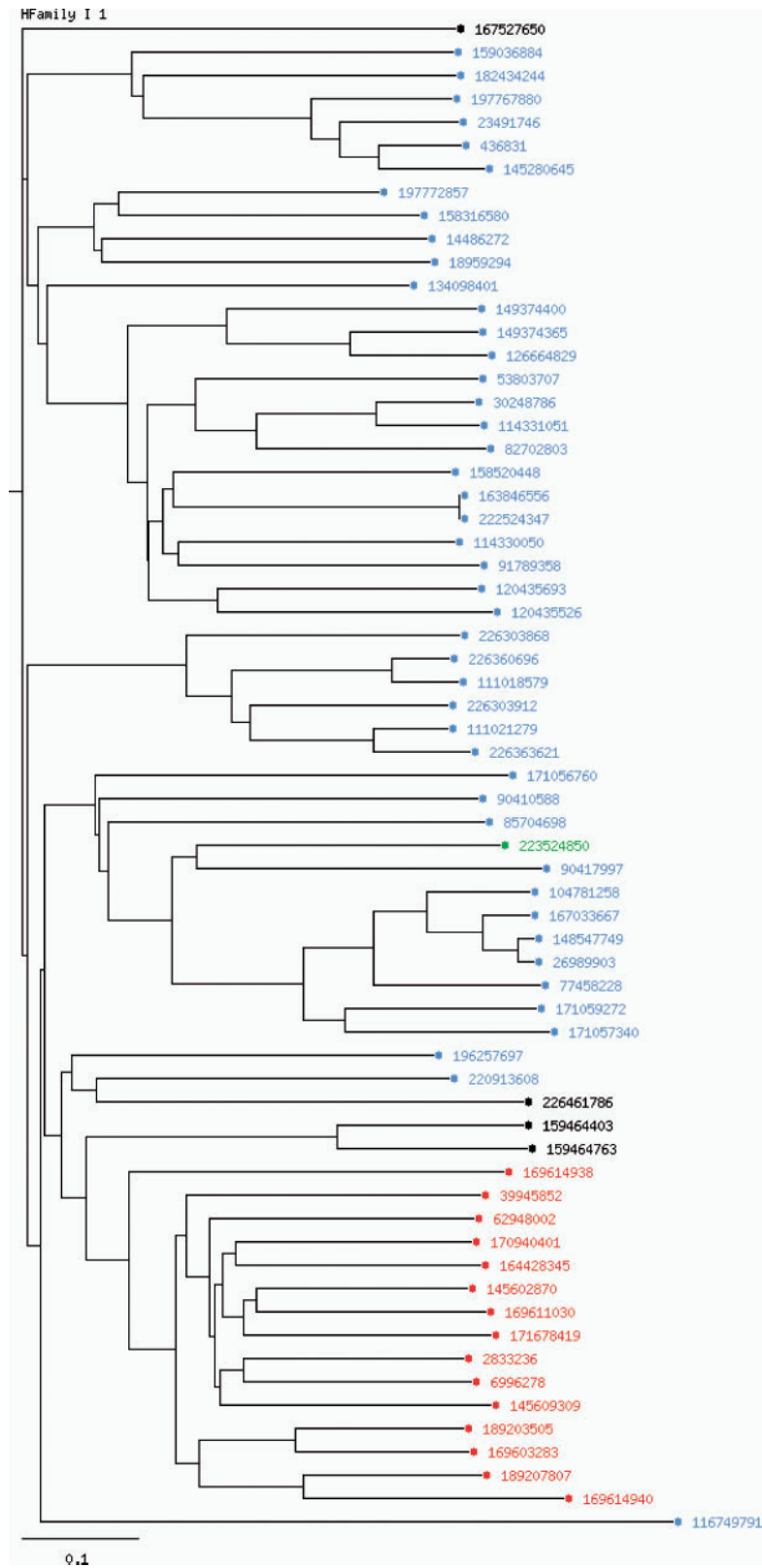
Figure 1. Phylogenetic tree for the homologous family I1 (Bilirubin oxidases).The chosen coloring option is 'by kingdom'. Entries of bacterial origin are shown in blue, fungal entries in red, plant proteins in green and non-specified entries are colored in black.
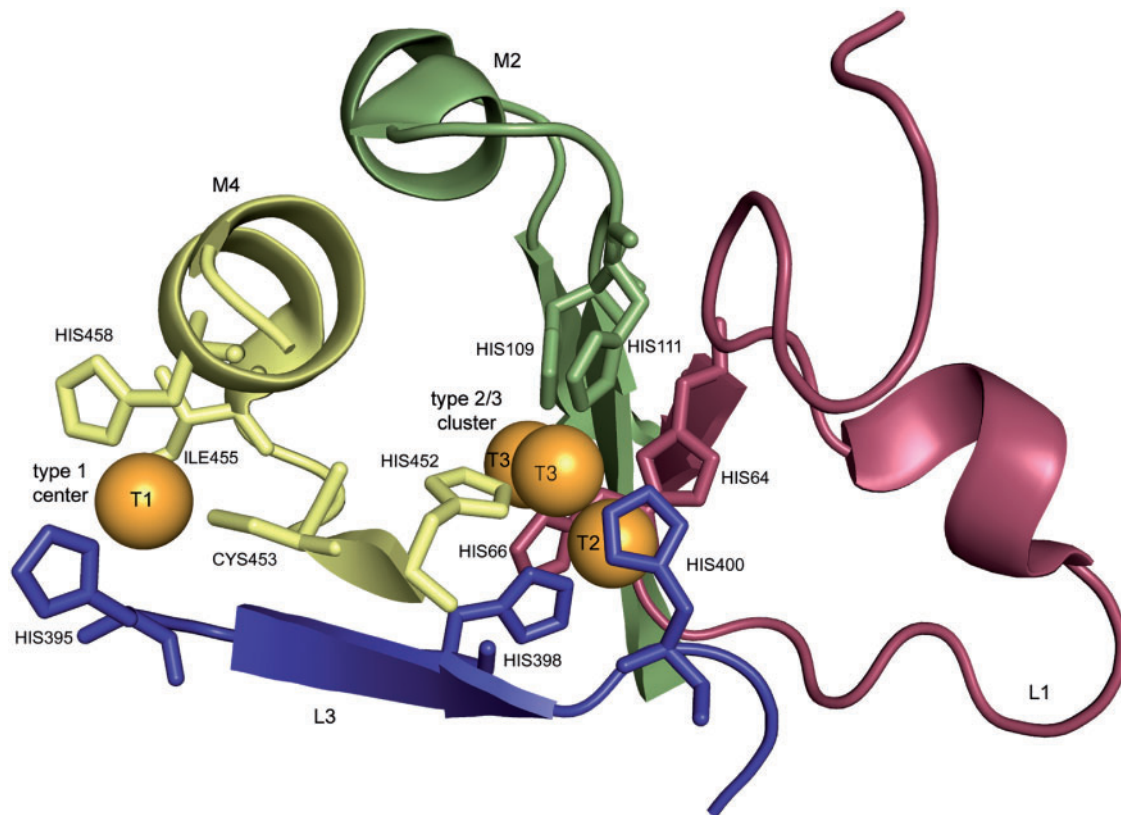
Figure 2. Copper binding residues of laccase from *Trametes versicolor* [PDB entry 1GYC, (39)].The copper centers are shown in orange, the residues that match the defined pattern L1, M2, L3, M4 are colored in red, green, blue, and yellow, respectively [visualization by PyMOL (40)].

contain pattern L1, 40% M2, 23% L3 and 3% M4. Forty-five percent of the entries are annotated as putative.

Family F (Fungal and Plant Ascorbate Oxidases) mainly contain proteins of plant origin (homologous families F2–F6). Twelve percent are of fungal origin and clustered all to the homologous family F1. Two percent are annotated as laccases in GenBank. In this family, 88% of the proteins contain pattern L1, 56% M2, 66% L3 and 66% M4. Sixty-two percent of the entries are annotated as putative.

Family G (Plant Laccases) exclusively contains proteins of plant origin and 83% are annotated as laccases in GenBank (homologous families G1–G5). Fifteen percent of the proteins contain pattern L1, 88% contain pattern M2, 77% contain pattern L3 and 2% contain pattern M4. Fifty-three percent of the entries are annotated as putative.

Family H (Bacterial CopA Proteins) contains proteins of which 98% were of bacterial origin (homologous families H1–H6). Eighty-three percent are annotated as laccases in GenBank. Fifty percent contain pattern L1, 50% M2, 42% L3 and 3% M4. Six percent of the entries are annotated as putative.

Family I (Bilirubin Oxidases) contains proteins of which 70% were of bacterial origin (homologous families I1–I5), 15% of plant origin (homologous family I3), 10% of fungal origin (homologous family I1) and 5% of unspecified source organism (Figure 1). Three percent are annotated in GenBank as laccases. Seventy percent contain pattern L1, 92% M2, 91% L3 and 65% M4. Twenty-six percent of the entries are annotated as putative.

Family J (Bacterial CueO Proteins) contains proteins of which 90% were of bacterial origin (homologous families J1–J4) and 10% of eukaryotic origin (homologous family J5). Twelve percent are annotated as laccases in GenBank. Seventy-four percent of the proteins contain pattern L1, 75% M2, 76% L3 and 3% M4. Thirty-nine percent of the entries are annotated as putative.

Family K (SLAC) which contains exclusively members of the 'small laccase family' which all are of bacterial origin, annotated as MCOs in GenBank and contain only pattern L4. Six percent of the entries are annotated as putative.

Besides the slight variations within the laccase and MCO sequence patterns almost all MCO sequences share the same highly conserved copper binding residues (Figure 2).

They could be identified and annotated by a manual validation of each family-specific multisequence alignment. Only within homologous families F5, F6, H3 and J2 these residues could not be detected.

Discussion

As suggested previously (23), the 10 MCO superfamilies were named by combining the name of the prevailing source organism and the putative enzymatic function. The overall distribution of source organisms among the families generally agreed with the initial classification. Since the assignment of a protein to a superfamily was exclusively based on sequence similarity, it was expected to find in the same family proteins from different source organisms, even from different kingdoms of life. Indeed, most families consisted of a majority of proteins belonging to one kingdom with a minority from other kingdoms, despite a sequence similarity as high as 40%. A systematic classification of all proteins by sequence similarity only was prerequisite to a reliable sequence alignment of superfamilies and to identify conserved, functionally relevant sequence patterns. However, a systematic analysis of previously described MCO- and laccase-specific patterns (4–6), which have been derived from a small number of MCOs and laccases, demonstrated their low sensitivity. This is also supported by the high number of false negatives which were retrieved by a manual analysis of the respective regions of the multisequence alignments (Supplementary Table S2). The MCO patterns M2 and M4 were only found in 15 and 65% of all MCOs (Supplementary Table S1), respectively. Nine percent of all MCOs contain both M2 and M4. To differentiate laccases from other MCOs is even more difficult. If we assume that sequence similarity is an indication of function similarities, there are four superfamilies which contain putative laccases. However, for these families the laccase-specific patterns L1 and L3 were only found in 45 and 37% of the sequences, respectively (Supplementary Table S1). Only 8% of all putative laccases contain all four patterns simultaneously. This low percentage of positive hits could either indicate that 'laccase superfamilies' contain MCOs without laccases activity, or it might be caused by the too restrictive patterns. As an alternative to patterns, sequence profiles are widely used to specify functionally related protein families (32, 33). Therefore, for each superfamily, a hidden Markov profile is provided, and the four copper binding regions are consistently annotated in the LccED.

To define discriminating rules for laccases, more detailed functional studies are needed. Newly gained information, for example, on redox potential or the effect of mutations, can be added to readily prepared tables and may be transferred on closely related proteins. It has been shown previously that a systematic classification of large protein families based on sequence similarity and comprehensive analysis tools as provided by the LccED serve as a reliable framework for studying sequence–structure–function relationships of enzyme families (34–36) and for the design of mutants or focused mutant libraries with improved biochemical properties (37, 38).

Conclusion

The LccED enables the systematic classification and analysis of MCO sequences and structures from different public sources. The integration of protein data in a relational database system has been used to study the molecular basis of biochemical properties and to investigate sequence–structure–function relationships. The LccED comes with a set of tools for phylogenetic analysis and classification. The annotated multisequence alignments allow the identification of the regions which house the copper atoms and other functionally relevant residues.

Availability

The LccED is available at http://www.LccED.uni-stuttgart.de. Via this web interface all sequences, alignments and trees are accessible and all data are supplied for download.

Supplementary Data

Supplementary data are available at *Database* online.

Acknowledgements

Funding

References

1. Solomon,E., Sundaram,U. and Machonkin,T. (1996) Multicopper oxidases and oxygenases. *Chem. Rev.*, **96**, 2563–2606.
2. Hulo,N., Bairoch,A., Bulliard,V. *et al*. (2008) The 20 years of PROSITE. *Nucleic Acids Res.*, **36**, D245–D249.
3. Sigrist,C.J., Cerutti,L., Hulo,N. *et al*. (2002) PROSITE: a documented database using patterns and profiles as motif descriptors. *Brief Bioinform.*, **3**, 265–274.
4. Messerschmidt,A. and Huber,R. (1990) The blue oxidases, ascorbate oxidase, laccase and ceruloplasmin. Modelling and structural relationships. *Eur. J. Biochem.*, **187**, 341–352.

5. Ouzounis,C. and Sander,C. (1991) A structure-derived sequence pattern for the detection of type I copper binding domains in distantly related proteins. *FEBS Lett.*, **279**, 73–78.

6. Kumar,S.V.S., Phale,P.S., Durani,S. *et al.* (2003) Combined sequence and structure analysis of the fungal laccase family. *Biotechnol. Bioeng.*, **83**, 386–394.

7. Nakamura,K. and Go,N. (2005) Function and molecular evolution of multicopper blue proteins. *Cell. Mol. Life Sci.*, **62**, 2050–2066.

8. Murphy,M.E., Lindley,P.F. and Adman,E.T. (1997) Structural comparison of cupredoxin domains: domain recycling to construct proteins with novel functions. *Protein Sci.*, **6**, 761–770.

9. Messerschmidt,A. (1997) *Multi-Copper Oxidases*. World Scientific, Singapore.

10. Mayer,A.M. and Staples,R.C. (2002) Laccase: new functions for an old enzyme. *Phytochemistry*, **60**, 551–565.

11. Alexandre,G. and Zhulin,I.B. (2000) Laccases are widespread in bacteria. *Trends Biotechnol.*, **18**, 41–42.

12. Dittmer,N.T., Suderman,R.J., Jiang,H. *et al.* (2004) Characterization of cDNAs encoding putative laccase-like multicopper oxidases and developmental expression in the tobacco hornworm, Manduca sexta, and the malaria mosquito, Anopheles gambiae. *Insect. Biochem. Mol. Biol.*, **34**, 29–41.

13. Bourbonnais,R. and Paice,M.G. (1990) Oxidation of non-phenolic substrates. An expanded role for laccase in lignin biodegradation. *FEBS Lett.*, **267**, 99–102.

14. Clutterbuck,A.J. (1972) Absence of laccase from yellow-spored mutants of Aspergillus nidulans. *J. Gen. Microbiol.*, **70**, 423–435.

15. Geiger,J.P, Nicole,M., Nandris,D. and Rio,B. (1986) Root-rot diseases of Hevea brasiliensis. 1. Physiological and biochemical aspects of host aggression. *Eur. J. Forest Pathol.*, **16**, 22–37.

16. O'Malley,D.M., Whetten,R., Bao,W. *et al.* (1993) The role of of laccase in lignification. *Plant J.*, **4**, 751–757.

17. Sharma,P., Goel,R. and Capalash,N. (2007) Bacterial laccases. *World J. Microbiol. Biotechn.*, **23**, 823–832.

18. Rodriguez Couto,S. and Toca Herrera,J.L. (2006) Industrial and biotechnological applications of laccases: a review. *Biotechnol. Adv.*, **24**, 500–513.

19. Riva,S. (2006) Laccases: blue enzymes for green chemistry. *Trends Biotechnol.*, **24**, 219–226.

20. Burton,S.G. (2003) Oxidizing enzymes as biocatalysts. *Trends Biotechnol.*, **21**, 543–549.

21. Kunamneni,A., Camarero,S., Garcia-Burgos,C. *et al.* (2008) Engineering and Applications of fungal laccases for organic synthesis. *Microb. Cell Fact.*, **7**, 32.

22. Fischer,M., Thai,Q.K., Grieb,M. *et al.* (2006) DWARF–a data warehouse system for analyzing protein families. *BMC Bioinformatics*, **7**, 495.

23. Hoegger,P.J., Kilaru,S., James,T.Y. *et al.* (2006) Phylogenetic comparison and classification of laccase and related multicopper oxidase protein sequences. *FEBS J.*, **273**, 2308–2326.

24. Machczynski,M.C., Vijgenboom,E., Samyn,B. *et al.* (2004) Characterization of SLAC: a small laccase from Streptomyces coelicolor with unprecedented activity. *Protein Sci.*, **13**, 2388–2397.

25. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J. *et al.* (2008) GenBank. *Nucleic Acids Res.*, **36**, D25–D30.

26. Durbin,R., Eddy,S., Krogh,A. and Mitchison,G. (1998) *Biological Sequence Analysis*. Cambridge University Press, Cambridge, UK.

27. Altschul,S.F., Madden,T.L., Schäffer,A.A. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

28. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.

29. Berman,H.M., Battistuz,T., Bhat,T.N. *et al.* (2002) The Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.*, **58**, 899–907.

30. Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

31. Rice,P., Longden,I. and Bleasby,A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.

32. Finn,R.D., Tate,J., Mistry,J. *et al.* (2008) The Pfam protein families database. *Nucleic Acids Res.*, **36**, D281–D288.

33. Servant,F., Bru,C., Carrere,S. *et al.* (2002) ProDom: automated clustering of homologous domains. *Brief Bioinform.*, **3**, 246–251.

34. Fischer,M. and Pleiss,J. (2003) The Lipase Engineering Database: a navigation and analysis tool for protein families. *Nucleic Acids Res.*, **31**, 319–321.

35. Knoll,M., Hamm,T.M., Wagner,F., Martinez,V. and Pleiss,J. (2009) The PHA Depolymerase Engineering Database: a systematic analysis tool for the diverse family of polyhydroxyalkanoate (PHA) depolymerases. *BMC Bioinformatics*, **10**, 89.

36. Sirim,D., Wagner,F., Lisitsa,A. and Pleiss,J. (2009) The Cytochrome P450 Engineering Database: integration of biochemical properties. *BMC Biochem.*, **10**, 27.

37. Seifert,A. and Pleiss,J. (2009) Identification of selectivity-determining residues in cytochrome P450 monooxygenases: a systematic analysis of the substrate recognition site 5. *Proteins*, **74**, 1028–1035.

38. Seifert,A., Vomund,S., Grohmann,K. *et al.* (2009) Rational design of a minimal and highly enriched CYP102A1 mutant library with improved regio-, stereo- and chemoselectivity. *Chembiochem*, **10**, 853–861.

39. Piontek,K., Antorini,M. and Choinowski,T. (2002) Crystal structure of a laccase from the fungus Trametes versicolor at 1.90-A resolution containing a full complement of coppers. *J. Biol. Chem.*, **277**, 37663–37669.

40. Delano,W.L. (2002) The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA.