Pyranose oxidase identified as a member of the GMC oxidoreductase family

Mario Albrecht* and Thomas Lengauer

Max-Planck-Institute for Informatics, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany

Received on December 3, 2002; revised on January 9, 2003; accepted on January 27, 2003

ABSTRACT
Fungal pyranose oxidase is a flavoenzyme whose preferred substrate among several monosaccharides is D-glucose. After a comprehensive analysis of conserved features in a structure-based multiple sequence alignment of homologous proteins, we could classify this enzyme into the GMC oxidoreductase family. The identified homology also suggests a three-dimensional protein structure similar to the functionally related glucose oxidase.

INTRODUCTION
Pyranose oxidase (POX; EC 1.1.3.10, pyranose:oxygen 2-oxidoreductase), also known as glucose 2-oxidase, catalyzes the oxidation of D-glucose and other aldopyranoses at carbon-2 in order to yield 2-ketoaldoses, reducing O₂ to H₂O₂ (Nishimura et al., 1996; Giffhorn, 2000; Leitner et al., 2001). This enzyme is produced by several lignin-degrading basidiomycete fungi and is primarily localized in the hyphal periplasmic space of the cell. It is assumed to play an important functional role by supplying H₂O₂ as cosubstrate to lignin-decomposing peroxidases.

The preferred substrate of POX is both α- and β-D-glucose, while the functionally related enzyme glucose oxidase (GOX; EC 1.1.3.4, β-D-glucose:oxygen 1-oxidoreductase) acts solely on β-D-glucose (Wohlfahrt et al., 1999). In contrast to GOX, POX has not been recognized yet to belong to the family of GMC (glucose–methanol–choline) oxidoreductases (GMC-OXREDs), all of which bind flavin adenine dinucleotide (FAD) as cofactor (Hecht et al., 1993). The homotetrameric POX and homodimeric GOX from fungi also share a similar molecular size of about 70–80 kDa and a sequence length of approximately 600 amino acids. However, solely the 3D structure of GOX has been determined crystallographically.

Though POX shows only a very low, non-significant sequence identity to the GMC-OXRED family members, we could assign it to this family based on many conserved features, which are revealed by the combination of different bioinformatics methods and a multiple sequence alignment of homologs. The identified distant homology also suggests that POX adopts a 3D structure very similar to other GMC oxidoreductases. This result is of great interest in practice because POX has received remarkable attention both in biotechnological application for high yields of D-fructose and in food technology as a tool for glucose measurement (Giffhorn, 2000). Apart from that, POX is used for analytical purposes, for instance, in clinical chemistry as a diagnostic marker of diabetes (Nishimura et al., 1996).

MATERIALS AND METHODS
All protein sequences were retrieved from the SWISS-PROT/TrEMBL (SPTrEMBL) database (Bairoch and Apweiler, 2000). Their accession numbers are given in the caption of Figure 1. Two POX sequences from the fungi Coriolus versicolor (cvPOX) and Trametes hirsuta (thPOX) are publicly available; we extracted the latter sequence from the US Patent 6146865, ‘Nucleic acids encoding polypeptides having pyranose oxidase activity’ (1999), by Christensen et al. and submitted it to SPTrEMBL. Members of the GMC-OXRED family were identified with the help of the Pfam database of protein families (Bateman et al., 2002). Structural classifications of protein homologs in the PDB (Berman et al., 2000) were obtained from the SCOP database (Murzin et al., 1995; Conte et al., 2002). The secondary structure assignments were taken from the DSSP database (Kabsch and Sander, 1983).

To predict the secondary structure of the POX protein sequences, we used the two prediction servers PSIPRED (McGuffin et al., 2000) and SSPro (Pollastri et al., 2002). Both are known for their comparatively good performance of over 75% accuracy on average (Rost and Eyrich, 2001), although they are based on different computational methods. The BLAST/PSI-BLAST suite of programs (Altschul et al., 1990, 1997; Wheeler et al., 2002) was applied with standard parameters to search for homologous proteins in the SPTrEMBL and NCBI non-redundant databases.
Fig. 1. Structure-based multiple sequence alignment of pyranose oxidases with GMC oxidoreductases. Known and predicted secondary structures are depicted in the upper part (α-helices are represented by curled lines, β-strands by arrows). Alignment columns with strictly conserved residues are highlighted in dark gray boxes, those in which more than 70% of the residues are physico-chemically equivalent are shown in light gray boxes. Dashed lines above the alignment indicate major conserved regions. Triangles, asterisks, and solid circles mark groups of alignment columns with strictly conserved amino acids of functional relevance. The SPTrEMBL accession numbers are as follows (sequence identities of cvPOX to the other GMC oxidoreductases are given in parentheses): pagOX (14%), P81156; cvPOX, P79076; thPOX (85%), P59007; dpGDEH (21%), P18172; stGDEH (21%), Q93149; mGOX (21%), Q9U86; abADEH (21%), Q9U8X6; cgCHDEH (18%), P17444; pdHNL (18%), Q945K2; peAAO (20%), Q94219; bsCHOX (15%), P22637.
sequence databases. The multiple sequence alignment was constructed by means of CLUSTAL W (Thompson et al., 1994; Higgins et al., 1996). Pairwise structural alignments and root mean square deviations (RMSDs) were computed by the superposition program CE (Shindyalov and Bourne, 1998). We explored all standard fold recognition tools available via the meta server by Bujnicki et al. (2001, http://BioInfo.PL/Meta). This meta server contacts twelve state-of-the-art prediction servers (whose names are listed on the web site) and reports their results. Figure 1 of the multiple sequence alignment was prepared using the web tool ESPript (Gouet et al., 1999).

RESULTS AND DISCUSSION

The cvPOX sequence of C. versicolor has a high sequence identity of 85% to thPOX of T. hirsuta. A BLAST search of cvPOX against SPTreMBL reports a hypothetical GMC-type oxidoreductase protein 4Y-NJ, but weak search hit with an E-value of 0.0002. A PSI-BLAST search of five iterations in the NCBI non-redundant database returns many significant hits, which cover the whole cvPOX sequence. The best hits are GMC oxidoreductases from different species according to the Pfam classification, including choline dehydrogenases (best E-value 10−106 from Escherichia coli), polyethylene glycol dehydrogenases (best E-value 10−104 from Sphingopyxis terrae), glucose dehydrogenases (best E-value 10−102 from Drosophila pseudoobscura), alcohol dehydrogenases (best E-value 10−99 from Alcanivorax borkumensis), and glucose oxidases (best E-value 10−63 from Apis mellifera).

Not surprisingly, the fold recognition servers contacted consistently detect homologies with very good confidence values to the GMC-OXREDs aryl-alcohol oxidase from Pleurotus eryngii (peAAO), of which a 3D model structure (Varella et al., 2000) is deposited under the PDB identifier (PDBid 1qjn, and glucose and cholesterol oxidases, whose 3D structures have been determined crystallographically (Hecht et al., 1993; Wohlfahrt et al., 1999; Vrieland et al., 1991; Li et al., 1993; Yue et al., 1999). The SCOP superfamily of these proteins additionally includes the diverse enzymes p-hydroxybenzoate hydroxylase (PHBH), sarcosine oxidase, phenol hydroxylase, polyamine oxidase, L-aminoo acid oxidase monooxidase oxidase B, and hydroxynitrile lyase.

GMC-OXRED sequences consist of a N-terminal FAD-binding domain and a C-terminal substrate-binding domain. The FAD-binding region shares its 3D fold not only with flavoenzymes, but also with other proteins such as the GTP-dissociation inhibitor of Rab GTPases (Mattevi, 1998; Fraaije and Mattevi, 2000). However, the flavoenzymes differ in their substrate binding mode, and catalytically relevant residues are not always located at corresponding positions. It is believed that a hypothetical ancestral enzyme has bound to various substrates, yielding a low level of catalysis for each (Cavener, 1992). Mutational refinements in the course of the evolution probably led to the higher specificity of the nowadays existing proteins with distinct functions.

The GMC-OXRED family in the Pfam database currently contains 120 protein sequences. Interestingly, in contrast to all other enzymes described above, the pyranose oxidase is neither classified as oxidoreductase nor assigned to any other protein family. Indeed, a Pfam search with the cvPOX sequence matches the GMC-OXRED family with a weak E-value of 0.14, which is greater than the statistical significance threshold that Pfam allows for inclusion into a protein family. However, several weak hits from the Pfam search point to FAD-binding regions in different protein families.

In order to verify our findings, we computed a multiple sequence alignment of the two pyranose oxidases cvPOX and thPOX with selected sequences of the identified GMC oxidoreductases, see Figure 1. Based on pairwise structure alignments of the glucose oxidase from Penicillium amagasakiense (paGOX, PDBid 1gpe), which shares a sequence identity of 73% and an RMSD of 0.8 Å with another GOX from Aspergillus niger (PDBid 1cf3), with the cholesterol oxidase from Brevibacterium sterolicum (bsCHOX, PDBid 1coy, RMSD 2.9 Å) and with the hydroxynitrile lyase from Prunus dulcis (pdHNL, PDBid 1ju2, RMSD 2.7 Å), we improved the multiple sequence alignment manually by minor modifications to agree with the structure superpositions. In addition, we annotated the alignment with the known secondary structures of paGOX and bsCOX and the corresponding predictions by PSIPRED and SSPro for cvPOX, which are all in agreement with each other in many alignment parts.

As detailed in the following, the multiple sequence alignment reveals the conservation of structurally and functionally important residues throughout the POX sequence despite low sequence identities of 14–22% to other GMC oxidoreductases. Note that other members of the GMC-OXRED family also show low sequence identities with each other, for instance, paGOX has a sequence identity of only 17% to bsCHOX based on their structure superposition. The residues in the multiple sequence alignment are numbered according to the paGOX sequence and discussed along the comprehensive analysis on paGOX in Kiess et al. (1998). The alignment also includes conserved N-terminal sequence segments of POX, which are assumed to consist of a prepeptide as signal peptide and a prepeptide (Nishimura et al., 1996). The prepeptide is proposed to target the preproenzyme to the hyphal periplasmic space as it is the case for lignin peroxidase.

We identified three major conserved regions in POX, which are characteristic of GOX and other GMC
oxidoreductases (Kiess et al., 1998): the FAD-binding site composed of four separate subregions, the flavin attachment loop, and the contiguous substrate-binding region, see Figure 1.

The first subregion near the N-terminus from Y25 to the conserved motif E/DxG at positions 55–57 contains the ADP-binding $\beta\alpha\beta$-fold motif (Wierenga et al., 1983), which consists of perfectly conserved residues in a fixed order from G31 to L43 (Wierenga et al., 1986): GxGxxGxxxAxxL. (x for less conserved amino acids). The second subregion from R243 to G308 also contains a number of conversed residues that constitute another typical GMC-OXRED signature. It encompasses a highly conserved C-terminus of the $\beta$-strand preceding the strictly conserved G294. This part is situated between the diphosphate and the ribose of the FAD in the paGOX structure. The strictly conserved L246 belongs to the $\beta$-sandwich of the FAD-binding motif.

The third, strongly conserved sequence subregion of the FAD-binding domain comprises the amino acids from V535 to the strictly conserved P557, which is close to the phosphate-binding region in paGOX. The fourth subregion is the C-terminal helix formed by the residues from H563 to I579. Contacts from this helix to oxygen-2 of the flavin ring stabilize its negative polarization by the helix dipole.

The flavin attachment loop fixes the flavin ring and is characterized by a strongly conserved region from G101 to P118, which is specific for GMC oxidoreductases. For instance, the N111 residue of paGOX is located on the si-face of the isoalloxazine ring of FAD to maintain its bent conformation. The extended FAD-binding region particularly contains a histidine at position 563, which is replaced by asparagine in COX and other GMC oxidoreductases (Witt et al., 2000). This position is assumed to be part of the active site near the isoalloxazine ring of FAD in GOX (Kiess et al., 1998; Wohlfahrt et al., 1999), COX (Yue et al., 1999; Sampson, 2001), and HNL (Dreveny et al., 2001, 2002). Another key member of the substrate binding region of GMC oxidoreductases is the absolutely conserved histidine at position 520, which probably acts as general base catalyst and accepts protons from the substrate.

Further relevant residues involved in binding of glucose by GOX are Y73, S114, R516, and N518. While Y73 is also contained in POX and HNL, but not COX, the remaining three amino acids are not conserved. This finding appears to account for the broader substrate specificity of POX in contrast to GOX. However, as the sequence alignment is not based on the true structure of POX, residues at other positions may have adapted similar functional roles for interactions with the substrates.

For a more detailed discussion of the structural and functional properties of the remaining conserved residues in GOX and other GMC oxidoreductases, we refer the reader to other publications on these enzymes; a comprehensive comparison to POX is beyond the scope of this study.

CONCLUSIONS

We identified several lines of evidence that allow us to assign the fungal enzyme pyranose oxidase to the GMC oxidoreductase family. In particular, the high degree of conservation of structural and functional features between pyranose and glucose oxidase underpins their evolutionary relationship despite low sequence identity.

It is now straightforward to extract a pairwise sequence-structure alignment from the multiple alignment shown in Figure 1 in order to model the 3D structure of the pyranose oxidase based on the glucose oxidase structure. The high degree of sequence conservation of large regions containing the binding sites allows their accurate modeling. This is in contrast to less conserved, but functionally irrelevant sequence parts.

The structural model can be useful in the optimized design of pyranose oxidases with altered substrate activities and specificities for industrial and clinical applications. In particular, the 3D model may provide a structural basis to model substrate complexes as it has already been performed with the known structure of glucose oxidase (Meyer et al., 1998; Wohlfahrt et al., 1999) and other oxidoreductases such as HNL (Dreveny et al., 2002).

ACKNOWLEDGEMENTS

We are grateful to Dorothee Heckmann and Friedrich Giffhorn for bringing the problem discussed in this note to our attention. Part of this research has been funded by the Deutsche Forschungsgemeinschaft (DFG) under contract no. LE 491/14-1.

REFERENCES


