Cloning and characterization of a cDNA encoding a cellobiose dehydrogenase from the white rot fungus *Phanerochaete chrysosporium*

Manuel Raices[a], Edenia Paifer[a], José Cremata[a], Raquel Montesino[a], Jerry Ståhlberg[b], Christina Divne[b], István J Szabó[c], Gunnar Henriksson[c], Gunnar Johansson[c], Göran Pettersson[c,*]

[a]*Centro de Ingeniéria Genética y Biotecnologia, Havana, Cuba*
[b]*Department of Molecular Biology, Uppsala University, Uppsala, Sweden*
[c]*Department of Biochemistry, University of Uppsala, Uppsala, Sweden*

Received 2 June 1995; revised version received 27 June 1995

Abstract The cDNA of cellobiose dehydrogenase (CDH) from *Phanerochaete chrysosporium* has been cloned and sequenced. The 5′ end was obtained by PCR amplification. The cDNA contains 2310 translated bases excluding the poly(A) tail. The deduced mature protein contains 770 amino acid residues and is preceded by a 18 residue long signal peptide. The regions of the amino acid sequence corresponding to the heme and FAD domains of CDH were identified as well as the nucleotide-binding motif, the disulfide pairing and a methionine residue chelating the heme iron. No homologous sequences were found for the heme domain, however, the FAD domain appears to be distantly related to the GMC oxidoreductase family.

Key words: Cellobiose dehydrogenase; Nucleotide sequence; Wood degradation; FAD; Heme; Nucleotide-binding motif

1. Introduction

White rot fungi are the only known organisms capable of efficient degradation of all wood components [1]. The problems in accessing and degrading a solid composite substrate such as wood are many, but they are, to a large extent, overcome through the production of a wide variety of extracellular enzymes. These enzymes display different catalytic activities in order to achieve complete and successful solubilisation of wood. Cellobiose dehydrogenase (CDH) is a hemoflavoenzyme which is secreted under cellulolytic conditions [2]. The most studied CDH is that of the white rot fungus *Phanerochaete chrysosporium*, however, the enzyme is produced by several other wood-degrading fungi as well. CDH oxidises cellodextrins, lactose and mannodextrins, and can employ a wide variety of electron acceptors including cytochrome *c*, quinones, triiodine ions, phenoxy radicals and complexed Fe(III) [2–6]. Interestingly, the enzyme binds strongly to cellulose, and the binding site has been found to be separate from the catalytic site [7,8]. The exact biological function of CDH is unknown, though it has been suggested that CDH takes part in wood degradation, or has a role in the catabolism of cellobiose [4,6,9]. Synergy between CDH and cellulases in the degradation of wood has been observed [10]. CDH can be cleaved with papain into two fragments containing heme and flavin as prosthetic groups, respectively. The flavin fragment retains the catalytic activity and the cellulose-binding properties of the intact enzyme [7].

*Corresponding author. Department of Biochemistry, BMC, University of Uppsala, Box 576, S-751 23 Uppsala, Sweden. Fax: (46) (18) 552 139. E-mail: gunnarh@bio.embnet.se

The nucleotide sequence reported in this paper has been deposited with the EMBL GenBank under the Accession Number X88897.

Abbreviations: CDH, cellobiose dehydrogenase; cDNA, complementary DNA; GMC, glucose/methanol/choline; FAD, flavin adenine dinucleotide; *P.c.*, *Phanerochaete chrysosporium*; PCR, polymerase chain reaction; PNGase F, peptide-*N*-glycosidase F; PVDF, polyvinyl difluoride; RACE, rapid amplification cDNA ends; RP-HPLC, reversed-phase high-pressure liquid chromatography.

2. Materials and methods

2.1. Organisms

Phanerochaete chrysosporium strain K3 was used in this study. *Escherichia coli* strains XL-1 blue S and XL-1 blue (Stratagene, USA) were used for transfection and growth of phage λ zap (Stratagene), and strain DH5α was used as host for pBluescript SK(−) (Stratagene), pT7BlueR (Novagene) and pUC 19.

2.2. Protein characterisation

Intact protein, as well as the FAD and heme fragments of CDH were prepared as described previously [7,11]. The disulfides were reduced with dithioerythritol and alkylated using 4-vinylpyridine [12]. Proteolysis was carried out with Lys C protease (Boehringer Mannheim, Germany), V8 protease (Sigma) or cyanogen bromide [12]. Peptides were purified by RP HPLC on a C-18 column (TJ Baker, USA) eluted with a 0–60% acetonitrile gradient in 0.1% trifluoroacetic acid, or by SDS-PAGE followed by electroblotting to a PVDF membrane (Millipore, USA) [13] and sequenced in an Applied Biosystems 470A gas-liquid phase sequencer equiped with an Applied Biosystems 120A PTH Analyser [14]. A test for free cysteines was performed by incubating CDH overnight in 6 M guanidine·HCl with excess of iodoacetic acid, followed by amino acid analysis.

2.3. Preparation of cDNA library

P.c. was cultivated for 26 h in an automatic 10 l fermentor using Norkrans medium with continuous addition of cellobiose to keep the fungus induced for CDH production [15]. The harvested mycelium was frozen in liquid nitrogen and grinded in a mortar to powder and suspended in 6 M guanidium thiocyanate solution. Total RNA was isolated using the CsCl gradient method [16]. Poly(A) RNA was purified either through oligo(dt)cellulose chromatography (Stratagene, USA) or Dynabeads (dt)25 (Dynal, USA). One band of ~2.4 kb was visualised in a Northern blot hybridisation (Hybond C membranes, Amersham) with [^{32}P]ATP labelled probes (Rediprime DNA-labelling protocol) specific for CDH (derived from peptide sequences). A cDNA gene library was constructed using 5 μg of the poly(A) RNA and the ZAP cDNA synthesis kit from Stratagene (USA) with an oligo(dt) *Xho*I primer. The cDNA was directionally cloned into *Eco*RI/*Xho*I digested λ ZAP arms and packaged using the Gigapack II packaging extract (Stratagene, USA).

2.4. Screening of the cDNA library

The library was subjected to immuno-screening using the Picoblue

SSDI 0014-5793(95)00758-X

kit (Stratagene, USA) and CDH specific, affinity purified [17], polyclonal antibodies. A second screening was done with oligo-nucleotide probes [18]. Isolated clones were excised to pBluescript II SK by phagemid rescue with F1 helper phage R408 (Stratagene, USA) and their size determined by restriction analysis [18]. Clones, positive in both immuno-screening and probe-screening, were purified by three repeated infections and screenings, and sub-cloned in pBluescript. Restriction pattern analysis (*Xho*I, *Eco*RI) suggested that all clones had

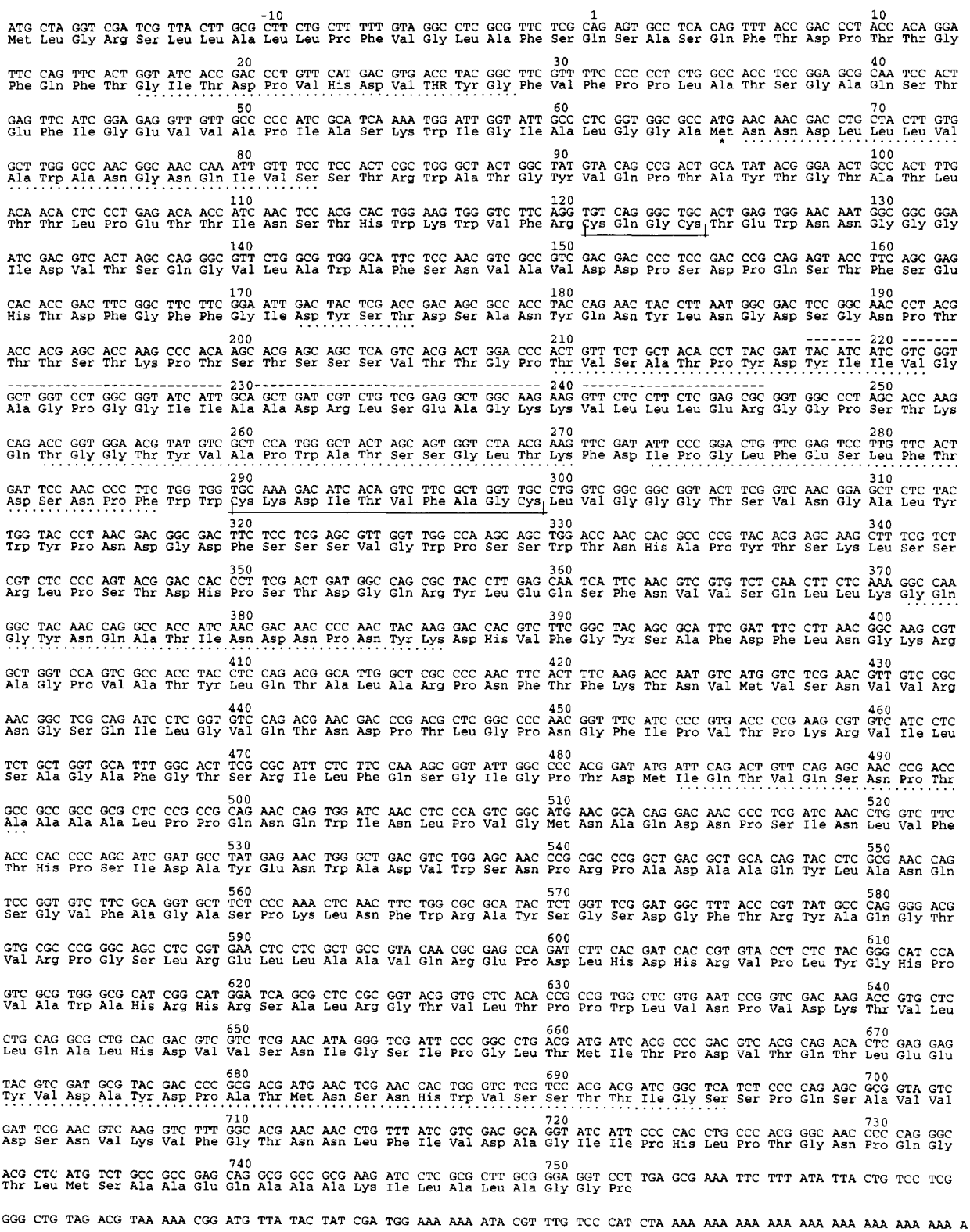

Fig. 1. The cDNA sequence of the CDH gene and the deduced amino acid sequence. Sequences corresponding to the peptides isolated from CDH are marked with (....). The FAD-binding motif is marked with (- - -), and the methionine residue chelating the heme iron [26] is marked with (*). Disulfide bridges are indicated by (LI). The first three peptides belong to the heme fragment, and the others from the FAD fragment.

a common sequence. In addition, some clones, positive in screening with oligo-nucleotides but negative in immuno-screening, were purified similarly.

2.5. Sequencing of cDNA

5 overlapping cDNA clones and one PCR product (the 5′ part) were sequenced in an automated DNA Sequencer (Applied Biosystems, USA), or using [α-^{35}S]dATP followed by autoradiography (Sequenase version 2.0 DNA polymerase kit, US Biochemical Corp., USA). Each region was sequenced at least 3 times, and the identity of each of the different clones could be confirmed.

2.6. Synthesis of 5′ end cDNA

Primers were purchased from Operon Inc. Technol., USA or obtained from the Department of Medical Genetics at Uppsala University, Sweden. Two primers were designed close to the 5′ end of the largest isolated cDNA clone. A biotin molecule was attached to the 5′ end of one of the primers. 2 μg of poly(A) RNA was used for the 5′ Amplifier RACE kit protocol (Clontech, USA). A DNA band of the expected size was extracted from the agarose gel (QUIAEX DNA extraction protocol), ligated to pT7blue(R) vector (Novagen, USA) and cloned in DH5α [18]. Plasmid purification was carried out using Wizard Maxiprep (Promega, USA).

2.7. Gene assembly

The 5′ (Ra-12) and the 3′ (CP3) cDNAs were amplified separately by PCR using Vent polymerase stock (2 U/μl; New England Biolabs, USA) and purified. To assemble the gene, the overlapping fragments were combined in a second PCR using the Ra-12 forward, and the CP3 reverse primers. The conditions were: 10 ng each of the Ra-12 and CP3 PCR products, 10 μl 10 mM dNTP, 0.5 μl 10 mg/ml bovine serum albumin, 28 μl H_2O, 5 μl 10× Vent Buffer, 1 μl (10 pmol) each of forward and reverse primer, 4 μl 100 mM $MgSO_4$ and 3 μl Vent polymerase stock, 94°C (1 min), 55°C (1 min), 72°C (3 min), 30 cycles. The product of expected size was isolated and cloned into the *Sma*I site of pUC 19 and pBluescript IISK. DH5α transformants were selected on LBA + IPTG and Xgal [18]. Correctly inserted CDH clones were identified through colony PCR using two sets of primers and 0.5 units of Taq Polymerase (Perkin Elmer) per set of reactions, 30 cycles each.

2.8. Deglycosylation

0.2 U PNGase F was added to a solution containing 20 μg CDH in 10 mM phosphate buffer, pH 5.5, and incubated at 37°C for 16 h. Furthermore, CDH was chemically deglycosylated using anhydrous trifluoromethane sulfonic acid (TFMS) with a GlycoFree Deglycosylation Kit (Oxford GlycoSystems, UK). Glycosylated and deglycosylated CDH was separated by SDS-PAGE and blotted onto a nitrocellulose filter. Presence of carbohydrate was analysed using Con A-Horseradish peroxidase conjungate (Sigma) (R. Garcia et al., submitted for publication) or the Schiff-base method [19].

3. Results and discussion

Screening of the cDNA-gene library enabled the isolation of more than 50 clones that were positive in immuno- or probe-screening. Restriction analysis, however, revealed that none of them corresponded to the full length cDNA that was expected from the single CDH mRNA species of 2.5 kb found by Northern blot analysis. Five of the cDNA clones were sequenced and found to overlap with identical sequences, indicating that they were synthesised from the same transcript, but lacking the 5′ end of the mRNA. To get the missing part, RACE PCR was performed using poly(A) RNA as template and two nestle primers designed from the 5′ end of the longest cDNA clone. One of the primers was biotinylated to prevent self-ligation of the cDNA by T4 RNA ligase. By this method, a PCR product (named Ra-12) was obtained whose size corresponded to the missing 5′ end. In the next step, the complete cDNA was restored by combining the 5′ and the 3′ parts, i.e. the PCR product (Ra-12) and the cDNA (CP3). The overlapping region, however, contained no unique restriction site that could be used for ligation. Instead, the full length cDNA was assembled by combining Ra-12 and CP3 in a recombinant PCR. The PCR product corresponded in size to the full length CDH cDNA, and its identity and correct assembly was verified by sequencing and comparison with the starting fragments.

The cDNA sequence and the deduced amino acid sequence of CDH are shown in Fig. 1. Sequences obtained from nine peptides of the FAD and heme fragments of CDH, obtained by Lys C and V8 proteolysis and cyanogen bromide cleavage, were identical or almost identical to the deduced sequence. The cDNA comprises 2449 nucleotides including the poly A tail, and contains 2310 translated bases with an open reading frame, coding for a protein of 770 amino acids. The sequence AAAAAA, 16 bp upstream of the poly(A) tail resembles the canonical eukaryotic polyadenylation signal AATAAA [20]. The codon usage for the coding region has a G + C content of 59%, similar to that reported for the total *P.c.* genomic DNA [21]. The proportion of G + C in degenerated codon positions is high, and the selectivity in codon usage is examplified by: (1) Asn: 42 AAC versus 4 AAT, (2) Glu: 14 GAG versus 1 GAA, and (3) Phe: 29 TTC versus 6 TTT.

CDH is a secreted enzyme and, thus, a signal peptide is expected at the N-terminus. A region with hydrophobic amino acid residues is found early in the open reading frame. The N-terminal region of the intact enzyme could not be sequenced successfully, indicating that the N-terminus is blocked and, hence, the position at which the signal peptide is cleaved off could not be determined. The algorithm developed by von Heijne [22] suggests that the residue denoted '1' in Fig. 1 should be the N-terminus of the mature enzyme (data not shown). The amino acid composition calculated from this N-terminal position correlates well with amino acid analysis results [7,11]

Table 1
Amino-acid composition of CDH

Amino acid	Amino acid analysis [7]	Deduced from cDNA
Aspartic acid	91[1]	40
Asparagine	nd[1]	46
Threonine	76	74
Serine	71	67
Glutamic acid	49[1]	15
Glutamine	nd[1]	34
Proline	53	55
Glycine	78	71
Alanine	66	66
Valine	51	56
Methionine	5	7
Isoleucine	33	33
Leucine	51	53
Tyrosine	24	24
Phenylalanine	33	33
Lysine	21	18
Histidine	11	15
Arginine	17	21
Tryptophan	18[2]	20
Cysteine	4	4
Sum	752	752

[1] Asparagine and glutamine are included in the corresponding carboxylic acids.
[2] Detected spectrophotometrically.

```
CDH 216  YDYIIVGAGPGGIIAADRLSEAG-KKVLLLERGGPSTK------QTGGTYVAPWATSSGLTKFDIPGLFESLFTDSNPF
CHD   3  FDYIIIGAGSAGNVLATRLTEDPNTSVLLLEAGGPDYRFDFRTQMPAALAFPLQGKRYNWAYETEPEPFMNNRRMEC--
GDH  65  YDFIVIGGGSAGSVVASRLSEVPQWKVLLIEAGGDE----PVGAQIPSMFLNFIGSDIDYRYNTEPEPMACLSSMEQRC
ADH   2  YDYIIVGAGSAGCVLANRLSADPSKRVCLLEAGPRD--TNPLIHMPLGIALLSNSKKLNWAFQTAPQQNLNGRSL----
GOX  20  VDYIIAGGGLTGLTTAARLTENPNISVLVIESGSYESDRGPIIE-DLNAYGDIFGSSVDHAYETVELATNNQTAL----
COX  12  VPALVIGSGYGGAVAALRLTQAG-IPTQIVEMGRSWDTPGSDGKIFCGMLNPDKRSMWLADKTDQPVSNFMGFGINKSI
         ========== 1 ==============

CDH 288  ---------WWCKDITVFAGCLVGGGTSVNGALYWYPNDGDFSSSVGWPS-----SWTNH--APYTSKLSSRLPSTDHP
CHD  80  ----------------GRGKGLGGSSLINGMCYIRGNALDL---DNWAQEPGLENWSYLDCLPYYRKAETR------D
GDH 142  ---------YW------PRGKVLGGTSVLNGMMYVRGNREDY---DDWAADGNP-GWAYNDVLPFFKKSEDN---LDLD
ADH  77  ---------FW------PRGKTLGGSSSINAMVYIRGHEDDY---HAWEQAA-GRYWQWYRALELFKRLECN----QRF
GOX  94  ----------------IRSGNGLGGSTLVNGGTWTRPHKAQV---DSWETVFGNEGWNWDNVAAYSLQAERARAPNAKQ
COX  90  DRYVGVLDSERFSGIKVYQGRGVGGGSLVNGGMAVTPKRNYF--EEILPSVDSNEMYNKYFPRANTGLGVNN-IDQAWF
                      ======== 2 =========

CDH 351  STDGQRYLEQSFNVVSQLLKGQGYNQATINDN-PNYK-----DHVFGYSAFDFL----------------NGKRAGPV
CHD 133  MGENDYH--GGDGPVS-------VTTSKPGVN-PLFEAMIEAGVQAGYPRTDDLNGYQQEGFGPMDRTVTPQGRRASTA
GDH 197  EVGTEYHAKGGLLPVG-----------KFPYNPPLSYAILKAGEELGFSVH-DLNGQNSTGFMIAQMTAR-NGIRYSSA
ADH 130  DKSEHHGVDGEL-AVSD---------LKY-IN-PLSKAFVQAGMEANINFNGDFNGEYQDGVGFYQVTQK-NGQRWSSA
GOX 154  IAAGHYFNASCHGVNGTVHAGPRDTGDDYS---PIVKALMSAVEDRGVPTKKDFGCGDPHGVSMFPNTLHEDQVRSDAA
COX 166  ESTEWYKFARTGRKTAQRSGFTTAFVPNVY---DFEYMKKEAAGQVTKSGLGGEVIYGNNA-----------GKKSLD
                                                                          =======

CDH 407  ATYLQTALARPNFTFKTNVMVSNVVR--NG--SQILGVQTNDPTLGP---NGFIPVTP-KRVILSAGAFGTSRILFQSG
CHD 202  RGYLDQAKSRPNLTIRTHAMTDHIIF--DG--KRAVGVEWL--EGDS---TIPTRATANKEVLLCAGAIASPQILQRSG
GDH 263  RAFLRPARMRNNLHILLNTTATKILI--HPHTKNVLGVEVSDQEGST---RKILVK---KEVVLSAGAVNSPHILLLSG
ADH 197  RAFLHGVLSRPNLDIITDAHASKILFE----DRKAVGVSYIKKNMHH---QVKTTSG--GEVLLSLGAVGTPHLLMLSG
GOX 230  REWLLPNYQRPNLQVLTGQYVGKVLLSQNGTTPRAVGVEFGTHKGNT---HNVYAK---HEVLLAAGSAVSPTILEYSG
COX 230  KTYLAQAAATGKLTITTLHRVTKVAPATGSGYS--VTMEQIDEQGNVVATKVVTADR----VFFAAGSVGTSKLLVSMK
         ========== 3 ==========                                  ======== 4 =======

CDH 478  IGPTDMIQTVQSNPTAAAALPPQNQWINLP-VGMNAQDNPSINLVFTHPSIDA-YENWADVWSNPRPADAAQYLANQ--
CHD 272  VGNAELL-----------AEFDIPLVHELPGVGENLQDHLEMYLQYECKEPVSLYPAL-QWWNQPKIGAEWLFGGTGVG
GDH 334  VGPKDEL-----------QQVNVRTVHNLPGVGKNLHNHVTYFTNFFIDDADT--APL-NWATAMEYLLFRDGLMSGTG
ADH 267  VGAAAEL-----------KEHGVSLVHDLPEVGKNLQDHLDITLMCAANSREPIGVAL-SFIPRGVSGLFSYVFKREGF
GOX 303  IGMKSIL-----------EPLGIDTVVDLP-VGLNLQDQTTATVRSRITSAGAGQGQA-AWFATFNETFGDYSEKAHEL
COX 303  AQGHLPNLSS-------------------Q-VGEGWGNNGNIMVGRANHMWDATGSKQ-ATIPTMGIDNWADPTAPIFA
         ======                  ===== 5 =====

CDH 553  -SGVF---------AGASPKLNFWRAYSGSDGFTRYAQGTVRPGSLRELLAAVQREPDLHDHRVPLYGHPVAWAHRHRS
CHD 339  ASNHF--------EAGGFIRSREEFA-WPNIQYHFLPVAINYNGSNAVKEHGFQCHVGS-------MRSPSRGHVRIKS
GDH 399  ISDVT---------AKLATRYADSPE-RPDLQLYFGG-YLASCARTGQVGELLSNNSRSIQIFPAVLNPRSRGFIGLRS
ADH 334  LTSNVA-------ESGGFVKSSPDRD-RPNLQFHFLPTYLKDHGRKIAGGYGYTLHICD-------LLPKSRGRIGLKS
GOX 369  LNTKLEQWAEEAVARGGFHNTTALLIQYENYRDWIVNHNVAYSELFLDTAGVASFDVWD-------LLPFTRGYVHILD
COX 361  EIAPLPAGLETY-----------------------------------VSLYLAITKNP--------ERARFQFNS

CDH 622  ALRGTVLTPPWLVNPVDKTVLLQALHDVVSNIGSIPGLTMITPDVTQTLEEYVDAYDPATMNSNHWSSTTIGSSP----
CHD 402  RDPHQHP-AILFNYMSHEQDWQEFRDAIRITREIMHQPALDQYRGREISPGVECQTDEQL------DEFVRNHAETAFH
GDH 467  ADPLEPP-RIVANYLTHERDVKTLVEGIKFVIRLSQTTPLKQYGMRLDKTVVKGCEAHAFGSDAYWECAVRQNTGPENH
ADH 398  ANPLQPP-LIDPNYLSDHEDIKTMIAGIKIGRAILQAPSMAKHFKHEVVPGQAVKTDDEII------EDIRRRAETIYH
GOX 441  KDPYLHHFAYDPQYFLNELDLLGQAAATQLARNISNSGAMQTYFAGETIPGDNLAYDADLSAWTE---YIPYHFRPNYH
COX 393  ---GTGKVDLTWAQSQNQKGIDMAKKVFDKINQKEGTIYRTDLFGVYYKTWGDD--------------------FTYH

CDH 698  -----------QS-AVVDSNVKVFGTNNLFIVDAGIIPHLPTGN-PQGT-LMSAAEQAAAKILALAGGP (752)
CHD 474  PCGTCKMGYDEMS--VVDGEGRVHGLEGLRVVDASIMPQIITGN-LNATTIMIGEKIADMI-RGQEALPRSTAGYFVAN
GDH 545  QAGSCKMGPSHDPMAVVNHELRVHGIRGLRVMDTSIMPKVSSGN-THAPAVMIAEKGAYLLKRAWGAKV (612)
ADH 470  PVGTCRMGKDPAS--VVDPCLKIRGLANIRVVDASIMPHLVAGN-TNAPTIMIAENAAEIIMRNLDVEALEASAEFARE
GOX 517  GVGTCSMMPKEMG-GVVDNAARVYGVQGLRVIDGSIPPTQMSSHVMTVFYAMALKISDAILEDYASMQ (583)
COX 448  PLGGVLLNK------ATDNFGRLPEYPGLYVVDGSLVPGNVGVNPFVTITALAERNMDKIISSDI (506)
                      ================== 6 ====================

CDH 752
CHD 549  GMPVRAKK (556)
GDH 612
ADH 546  GAELELAMIAVCM (558)
GOX 583
COX 506
```

Fig. 2. Alignment of the FAD fragment of the *P.c.* CDH sequence (residue 216–752) with four members of the GMC oxidoreductase family (the Swiss Prot database entries are given in soft brackets); CHD: choline dehydrogenase (*Escherichia coli*, BETA_ECOLI); GDH: glucose dehydrogenase (*Drosophila melanogaster*, DHGL_DROME); ADH: alcohol dehydrogenase (*Pseudomonas oleovorans*, ALKJ_PSEOL); GOX: glucose oxidase (*Aspergillus niger*, GOX_ASPNG); and COX: cholesterol oxidase (*Brevibacterium sterolicum*, CHOD_BREST). The regions that are considered similar are underlined and numbered 1–6 (*i.e.* regions that are likely to give rise to similar structure). Region 1 comprises the canonical ADP-binding motif [34]. An initial alignment was obtained with the Clustal algorithm in MEGALIGN included in the DNASTAR package of programs, and it was fine-tuned manually. Alignment of the sequences of GOX and COX was adjusted to optimally agree with the structural superposition of the FAD domains of the two enzymes (Protein Data Bank entries 1GAL and 3COX). The amino-acid sequence of cholesterol oxidase (COX) was taken from the Protein Data Bank entry 3COX where errors in the original sequence had been corrected.
←

(Table 1), although the calculated molecular weight is somewhat smaller than that estimated from SDS-PAGE [7]. Considering that CDH is a glycoprotein [8,23], the discrepancy in size may be due to glycosylation (there are eight potential *N*-glycosylation sites in the CDH sequence as defined by the sequon Asn-Xaa-Ser/Thr). This is supported by an observed reduction in molecular weight (not quantified) and sharpening of the protein band on PAGE gels after enzymatic or chemical removal of carbohydrates (see section 2.8 for details). Prior to removal of surface-attached carbohydrates, carbohydrate assays indicated the presence of carbohydrate, and specifically mannose. Since both *N*- and *O*-linked carbohydrate chains may contain mannose, these methods alone cannot discriminate between the two types of glycosylation. However, after treatment with PNGase F, which specifically hydrolyses the β-asparatyl-glucosamine bond between the peptide asparagine and the innermost *N*-acetyl glucosamine of most *N*-glycan trimannosyl cores, no carbohydrate could be detected on CDH. This indicates that CDH is mainly *N*-glycosylated, with little or no *O*-linked sugars.

The cysteine pairing scheme suggested in Fig. 1 is supported by the observations that the sequence of CDH contains four cysteines, that no free cysteines have been detected (i.e. no carboxymethyl cysteine was formed when denatured CDH was incubated with iodoacetic acid under non-reducing conditions), and that CDH can be cleaved by papain into two fragments [7], each containing two cysteines. Each domain must therefore contain one disulfide.

The first 190 residues of the mature protein show an amino acid composition similar to that of the heme fragment of CDH and the remaining part of the sequence has a composition similar to that of the FAD fragment [7,11]. This agrees well with the conclusion of Wood and Wood [24] that the heme domain is located at the N-terminus of the enzyme. Cox et al. [25] showed that the heme iron is chelated by one histidine and one methionine residue. The sequence corresponding to the heme fragment contains only one methionine (Fig. 1), Met-65, which is therefore likely to be one of the heme ligands.

There is a short region (residues 192–207) between the heme and FAD-binding domains which is surprisingly rich in threonines and serines. The region is susceptible to proteolytic cleavage by papain, and its presence prompts the question whether this stretch of amino acids may form a linker-region between the domains, similar to the glycosylated inter-domain spacer regions found in fungal cellulases [26,27]. In the case of cellulases, the spacer is ~30 residues long and separates the catalytic domain of the enzyme from its cellulose-binding domain to allow optimal operation of both domains on crystalline cellulose [26]. There is, however, no data suggesting that CDH would need a functional equivalent to the cellulase linkers. Furthermore, it has been suggested that the binding of CDH to cellulose may depend on a cellulose-binding domain [6,7], as is the case for most fungal cellulases. However, no sequence similarities between the fungal cellulose-binding domains [27,28] and CDH have been detected.

In order to investigate possible relationships between the primary structure of *P.c.* CDH and those of other proteins, protein sequence database resources were searched using the program package DNASTAR (DNASTAR Inc., USA). Separate homology searches were performed for the heme domain (residue 1–190) and the FAD domain (residue 208–752). No significantly homologous sequences were found for the heme domain. For the FAD domain, however, a distant relationship to a subfamily of flavoenzymes was discovered, namely to the GMC family of oxidoreductases [29]. Members of this subfamily include choline dehydrogenase (*Escherichia coli*), glucose dehydrogenase (*Drosophila melanogaster*), alcohol dehydrogenase (*Pseudomonas oleovorans*), methanol oxidase (*Hansenula polymorpha* and *Pichia pastoris*), glucose oxidase (*Aspergillus niger*) and cholesterol oxidase from *Brevibacterium sterolicum*. To date, three-dimensional structures have been reported for the two latter enzymes [30,31]. The GMC oxidoreductases all carry FAD as a cofactor, but in contrast to CDH, they do not contain a heme group. The polypeptide chains of glucose oxidase and cholesterol oxidase fold into two structurally distinct domains; a FAD-binding domain with high structural conservation between the two enzymes, and a substrate-binding domain with similar overall topology but low sequence identity. The FAD-binding domain contains the characteristic ADP-binding $\beta\alpha\beta$-motif frequently encountered in NAD and FAD dependent enzymes [32,33] (Figs. 1 and 2). The sequence similarity between *P.c.* CDH and the GMC oxidoreductases is essentially restricted to six regions (Fig. 2). When mapped to the three-dimensional crystal structures of glucose oxidase and cholesterol oxidase (Protein Data Bank [34] accession codes 1GAL and 3COX, respectively), these regions cluster exclusively in the FAD-binding domain. The less similar parts of the sequences map onto the substrate-binding domain. The diversity of specificity among these enzymes may account for the lack of amino-acid sequence conservation in their substrate-binding domains. Thus, assuming a structural relationship to glucose oxidase and cholesterol oxidase, the FAD fragment of CDH is likely to consist of at least two domains, one FAD-binding domain, similar to that of the GMC oxidoreductases, and a structurally distinct domain for binding of the substrate, cellobiose.

At present, there are only a few known examples of proteins that carry both heme and flavin as prosthetic groups, namely the flavocytochromes represented by flavocytochrome b_2 from yeast [35]. Although functional similarities may exist between the flavocytochromes and CDH, no sequence similarities were detected. Thus, it may well be that CDH represents a novel type

of flavocytochrome, different from the flavocytochrome b_2-type, and similar to the GMC family of oxidoreductases, though the relationship may not extend beyond the FAD-binding scaffold.

Acknowledgements: This work has been supported by the Swedish Agency for Research Co-operation with Developing Countries (SAREC) and the Swedish Natural Science Research Council (NFR). We thank Dr. Esteban Bontempi and Dr. Jan Henriksson for valuable comments and suggestions, Professor Ulf Pettersson for giving MR the possibility to work in his laboratory, Ann-Sofie Engström for help with the automatic DNA sequencing, and Bo Ek for help with the peptide sequencing.

References

[1] Eriksson, K.-E., Blanchete, R.A. and Ander, P. (1990) in: Microbial and enzymatic degradation of wood and wood components, Springer Verlag, Berlin, Germany.
[2] Westermark, U. and Eriksson, K.-E. (1974) Acta Chem. Scand. B28, 209–214.
[3] Bao, W. and Renganathan, V. (1991) FEBS Lett. 279, 30–32.
[4] Ander, P. (1994) FEMS Microb. Rev. 13, 297–312.
[5] Eriksson, K.-E., Habu, N. and Samejima, M. (1993) Enzym. Microbiol. Technol. 15, 1002–1008.
[6] Renganathan, V., Usha, S.N. and Lindenburg, F. (1990) Appl. Microbiol. Technol. 32, 609–613.
[7] Henriksson, G., Pettersson, G., Johansson, G., Ruiz, A. and Uzcategui, E. (1991) Eur. J. Biochem. 196, 101–106.
[8] Ayers, A., Ayers, S. and Eriksson, K.-E. (1978) Eur. J. Biochem. 90, 171–181.
[9] Bao, W., Usha, S.N. and Renganathan, V. (1993) Arch. Biochem. Biophys. 300, 705–713.
[10] Bao, W. and Renganathan, V. (1992) FEBS Lett. 302, 77–80.
[11] Henriksson, G. (1995) in: Structure, Function and Applications of Cellobiose Dehydrogenase from *Phanerochaete chrysosporium*, PhD Thesis, Uppsala University, Uppsala, Sweden. Almqvist & Wiksell International, Stockholm, Sweden.
[12] Allen, G. (1981) in: Sequencing of proteins and peptides, Elsevier Science, Oxford, UK.
[13] Maizel, J.V. (1969) in: Fundamental techniques in virology (Habel K. and Salzman, N. P., eds.) pp. 334–362, Academic press, NY.
[14] Hewick, R.M., Hunkapiller, M.W., Hood, L.E. and Dreyer, W.J. (1981) J. Biol. Chem. 256, 7990–7997.
[15] Uzcategui, E., Raices, M., Montesino, R., Johansson, G., Pettersson, G. and Eriksson, K.-E. (1991) Biotechnol. Appl. Biochem. 13, 323–334.
[16] Haylock, R. and Broda, P. (1988) Methods Enzymol. 161, 221–227.
[17] Board, P.G. and Webb, G.C. (1987) Proc. Natl. Acad. Sci. USA, 84, 2377–2381.
[18] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) in: Molecular cloning. A laboratory manual. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[19] Fairbanks, G., Steck, T.L. and Wallach, D.F.H. (1971) Biochemistry, 10, 2606–2617.
[20] Proudfoot, N.J. and Brownlee, G.C. (1976) Nature, 263, 211–214.
[21] Raeder, U. and Broda, P. (1984) Curr. Genetics 8, 499–506.
[22] von Heijne, G. (1986) Nucl. Acid Res. 14, 4683–4690.
[23] Morpeth, F. (1985) Biochem. J. 228, 557–564.
[24] Wood, J. and Wood, P. (1992) Biochem. Soc. Trans. 20, 109S.
[25] Cox, M., Rogers, M,. Cheesman, M., Jones, G., Thomson, A., Wilson, M. and Moore, G. (1992) FEBS Lett. 307, 233–236.
[26] Srisodsuk, M., Reinikainen, T., Penttilä, M. and Teeri, T. (1993) J. Biol. Chem. 268, 20756–20761.
[27] Knowles, J.K.C., Lehtovaara, P. and Teeri, T. (1987) Trends Biotechnol. 5, 255–261.
[28] van Tilbeurgh, H., Tomme, P., Claeyssens, M., Bhikhabhai, R. and Pettersson, L.G., (1986) FEBS Lett. 204, 223–227.
[29] Cavener, D.R. (1992) J. Mol. Biol. 223, 811–814.
[30] Hecht, H.J., Kalisz, H.M., Hendle, J,. Schmid, R.D. and Schomburg, D. (1993) J. Mol. Biol. 229, 153–172.
[31] Vrielink, A., Lloyd, L.F. and Blow, D.M. (1991) J. Mol. Biol. 219, 533–554.
[32] Wierenga, R.K., Drenth, J. and Schulz, G.E. (1983) J. Mol. Biol. 167, 725–739.
[33] Wierenga, R.K., Terpstra, P. and Hol, W.G.J. (1986) J. Mol. Biol. 187, 101–107.
[34] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice Jr., M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–542.
[35] Xia, Z.-X. and Mathews, F.S. (1990) J. Mol. Biol. 212, 837–863.