Volume 255, number 1, 63-66 FEBS 07563 September 1989

Cloning and DNA sequence analysis of the glucose oxidase gene from *Aspergillus niger* NRRL-3

Marion Kriechbaum, Hans J. Heilmann, Franz J. Wientjes, Marina Hahn, Klaus-D. Jany, Hans G. Gassen, Fadel Sharif* and Gürdal Alaeddinoglu*

Institut für Biochemie, Technische Hochschule Darmstadt, Petersenstrasse 22, D-6100 Darmstadt, FRG and **Department of Biological Sciences, Middle East Technical University, Ankara, Turkey*

Received 17 July 1989

A cDNA library from *Aspergillus niger* strain NRRL-3 enriched in sequences encoding the glucose oxidase was constructed. An 800 bp cDNA clone isolated from this library was used to screen 12 000 recombinant phages from an EMBL3 genomic library. A 15 kbp DNA segment isolated from this library contained the 1815 bp structural gene for glucose oxidase as well as a short 5'- and a longer 3'-noncoding region. The deduced protein sequence was verified by partial peptide sequencing.

Glucose oxidase; DNA sequence; (*Aspergillus niger* NRRL-3)

1. INTRODUCTION

Glucose oxidase (β-D-glucose:oxygen 1-oxidoreductase, EC 1.1.3.4) from *Aspergillus niger* catalyzes the oxidation of β-D-glucose by molecular oxygen to give D-glucono-δ-lactone and hydrogen peroxide. The enzyme is a flavin-containing globular glycoprotein with a molecular mass of about 155 kDa [1-3]. Glucose oxidase is of considerable industrial importance. It is widely applied for the determination of glucose in body fluids and in removing residual glucose or oxygen from foods and beverages. Furthermore, glucose oxidase-procuding moulds such as *Aspergillus* and *Penicillium* species are used for the biological production of gluconic acid [4].

Correspondence address: H.G. Gassen, Institut für Biochemie, Technische Hochschule Darmstadt, Petersenstrasse 22, D-6100 Darmstadt, FRG

Abbreviations: bp, base pair; ds, double-stranded; GOD, glucose oxidase; HPLC, high-performance liquid chromatography

The nucleotide sequence presented here has been submitted to the EMBL/GenBank database under the accession no. X 16061

Although some data on properties of this important technical enzyme are now available, almost nothing is known about its biological function, the primary structure of the protein or the organization and regulation of the corresponding gene. Here, we describe the cloning and sequencing of the glucose oxidase gene of *A. niger* NRRL-3 including 5'- and 3'-flanking regions. We also present the DNA-derived amino acid sequence of glucose oxidase and show its identity with peptide sequences determined for parts of this protein.

2. MATERIALS AND METHODS

2.1. *Strains and culture conditions*

The source of fungal RNA and genomic DNA throughout this investigation was *A. niger* NRRL-3 (ATTC 9029). The mycelia were grown as described [5]. *Escherichia coli* strains DH5α [6] and XL1-Blue (Stratagene) were the host strains for the plasmids pUC9 [7] and pBluescript (Stratagene), respectively. *E. coli* strains NM 538 and NM 539 were host strains for propagation of phage λ EMBL3 and its derivatives [8]. Bacterial cells were grown in LB media under the appropriate selective conditions [9].

2.2. *Northern and Southern analyses*

DNAs and glyoxal-denatured RNAs were separated on

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50 © 1989 Federation of European Biochemical Societies

agarose gels, transferred to Hybond N membranes (Amersham) and hybridized either with ^{32}P-labeled oligonucleotides or nick-translated GOD cDNA fragments according to the manufacturers suggestions.

2.3. *Isolation of poly(A)$^+$ RNA and cDNA cloning*

Fungal total RNA was prepared from induced mycelia [5] with the guanidinium isothiocyanate procedure as described by Chirgwin et al. [10]. Poly(A)$^+$ RNA isolation, procedures for synthesis of cDNA and ds cDNA by M-MLV reverse transcriptase (BRL), insertion of dC-tailed ds cDNA into dG-tailed pUC9 (Pharmacia), transformation and the screening with ^{32}P-labeled oligonucleotide probes were performed as described [11].

2.4. *Isolation of chromosomal DNA and construction of a genomic library*

The preparation of chromosomal DNA was carried out as in [12]. The DNA was partially digested with *Mbo*I and 15–20-kbp fragments were isolated from a sucrose gradient [9]. DNA from bacteriophage EMBL3 was digested with *Bam*HI and *Eco*RI, followed by removal of the small linker fragments. Then about 80 ng *Aspergillus* DNA fragments were ligated to 500 ng vector arms. After in vitro packaging, recombinant phages were selected according to their ability to grow in cells of *E. coli* NM 539 carrying prophage P2 [8].

2.5. *DNA sequencing*

DNA sequences of both strands were determined by the dideoxynucleotide chain-termination method [14] using synthetic oligonucleotide primers.

2.6. *Protein analyses*

Cleavage of 2 mg purified GOD (Boehringer) with cyanogen bromide was performed as described [15]. Larger sized peptides (>25 residues) were subfragmented with trypsin [16] and purified as usual. Amino acid sequences were determined by automated Edman degradation on a gas-phase sequencer (Applied Biosystems, type 470 A), analysing the phenylthiohydantoin derivatives of amino acids on-line with a HPLC type 120 A apparatus from Applied Biosystems.

3. RESULTS AND DISCUSSION

3.1. *Analysis of cDNA coding for a part of the glucose oxidase gene*

A partial amino acid sequence of GOD (Boehringer) representing about 50% of the total primary structure was determined by automated Edman degradation. The sequence information of one of these fragments was used to construct a 20-mer oligonucleotide probe (MK7) complementary to the deduced mRNA sequences of the corresponding region. The third base degeneracies of amino acid codons were considered and thus the probe contained 64 different species of DNA sequences to ensure that the correct one was present (fig.1).

The GOD-specific oligoprobe MK7 was subsequently used to screen a cDNA library of *A. niger* NRRL-3. Among 3500 recombinant colonies only one positive clone harbouring a 800 bp cDNA insert was identified. DNA sequencing revealed an open reading frame of 728 nucleotides ranging from nucleotide 1127 to 1856 (fig.3). Comparison with data obtained later from the genomic clone shows perfect agreement with the cDNA sequence which comprises about 30% of the total glucose oxidase gene.

3.2. *Cloning and sequence analysis of the genomic glucose oxidase gene*

A library of the *A. niger* NRRL-3 genome was constructed in the phage λ substitution vector

```
amino acid sequence        E   Q   W   A   E   E   A

                         1229                      1248
possible codons      5' GAA CAA TGG GCT GAA GAA GC 3'
                        G   G       C   G   G
                                    A
                                    G

oligonucleotide MK7  3' CTT GTT ACC CGA CTT CTT CG 5'
                          C   C     G   C   C
                                    T
                                    C
```

Fig.1. Construction of the 20-mer oligonucleotide probe MK7 used for the isolation of the cDNA clone encoding glucose oxidase.

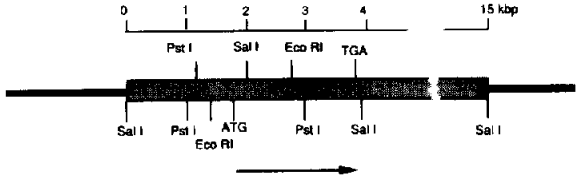

Fig.2. Location of the GOD gene on the 15 kbp insert (stippled box) of the recombinant EMBL3 clone. The arrow shows the size and orientation of the gene within the 15 kbp fragment. The start and stop codon as well as important restriction sites around the GOD gene are indicated.

Fig.3. Sequence of the glucose oxidase encoding gene. The DNA sequence and the derived amino acid sequence are given. Peptide fragments as determined by Edman degradation are underlined. Differences between the derived and the analyzed peptides are indicated by bold letters and the signal sequence is given in italics. Possible regulatory sequences such as the polyadenylation site are underlined. The cDNA clone (not shown) covered the region from nucleotide 1133 to 2028.

```
                                          M   Q   T   L   L   V   S   S   L   V   V   S   L   A   A   A   L
5' CAACCAGCCTTTCCTCTCTCATTCCCTCATCTGCCCATC ATG CAG ACT CTC CTT GTG AGC TCG CTT GTG GTC TCC CTC GCT GCG GCC CTG
                                                                  50
 P   H   Y   I   R   S   N   G   I   E   A   S   L   L   T   D   P   K   D   V   S   G   R   T   V   D   Y   I
CCA CAC TAC ATC AGG AGC AAT GGC ATT GAA GCC AGC CTC CTG ACT GAT CCC AAG GAT GTC TCC GGC CGC ACG GTC GAC TAC ATC
        100                                                                 150
 I   A   G   G   G   L   T   G   L   T   T   A   A   R   L   T   E   N   P   N   I   S   V   L   V   I   E   S
ATC GCT GGT GGA GGT CTG ACT GGA CTC ACC ACC GCT GCT CGT CTG ACG GAG AAC CCC AAC ATC AGT GTG CTC GTC ATC GAA AGT
                                200                                                                 250
 G   S   Y   E   S   D   R   G   P   I   I   E   D   L   N   A   Y   G   D   I   F   G   S   S   V   D   H   A
GGC TCC TAC GAG TCG GAC AGA GGT CCT ATC ATT GAG GAC CTG AAC GCC TAC GGC GAC ATC TTT GGC AGC AGT GTA GAC CAC GCC
                                                            300
 Y   E   T   V   E   L   A   T   N   N   Q   T   A   L   I   R   S   G   N   G   L   G   G   S   T   L   V   N
TAC GAG ACC GTG GAG CTC GCT ACC AAC AAT CAA ACC GCG CTG ATC CGC TCC GGA AAT GGT CTC GGT GGC TCT ACT CTA GTG AAT
    350                                                                 400
 G   G   T   W   T   R   P   H   K   A   Q   V   D   S   W   E   T   V   F   G   N   E   G   W   N   W   D   N
GGT GGC ACC TGG ACT CGC CCC CAC AAG GCA CAG GTT GAC TCT TGG GAG ACT GTC TTT GGA AAT GAG GGC TGG AAC TGG GAC AAT
                            450                                                                 500
 V   A   A   Y   S   L   Q   A   E   R   A   R   A   P   N   A   K   Q   I   A   A   G   H   Y   F   N   A   S
GTG GCC GCC TAC TCC CTC CAG GCT GAG CGT GCT CGC GCA CCA AAT GCC AAA CAG ATC GCT GCT GGC CAC TAC TTC AAC GCA TCC
                                                            550
 C   H   G   V   N   G   T   V   H   A   G   P   R   D   T   G   D   D   Y   S   P   I   V   K   A   L   M   S
TGC CAT GGT GTT AAT GGT ACT GTC CAT GCC GGA CCC CGC GAC ACC GGC GAT GAC TAT TCT CCC ATC GTC AAG GCT CTC ATG AGC
    600                                                                 650
 A   V   E   D   R   G   V   P   T   K   K   D   F   G   C   G   D   P   H   G   V   S   M   F   P   N   T   L
GCT GTC GAA GAC CGG GGC GTT CCC ACC AAG AAA GAC TTC GGA TGC GGT GAC CCC CAT GGT GTG TCC ATG TTC CCC AAC ACC TTG
                            700                                                                 750
 H   E   D   Q   V   R   S   D   A   A   R   E   W   L   L   P   N   Y   Q   R   P   N   L   Q   V   L   T   G
CAC GAA GAC CAA GTG CGC TCC GAT GCC GCT CGC GAA TGG CTA CTT CCC AAC TAC CAA CGT CCC AAC CTG CAA GTC CTG ACC GGA
                                                800
 Q   Y   V   G   K   V   L   L   S   Q   N   G   T   T   P   R   A   V   G   V   E   F   G   T   H   K   G   N
CAG TAT GTT GGT AAG GTG CTC CTT AGC CAG AAC GGC ACC ACC CCT CGT GCC GTT GGC GTG GAA TTC GGC ACC CAC AAG GGC AAC
    850                                                                 900
 T   H   N   V   Y   A   K   H   E   V   L   L   A   A   G   S   A   V   S   P   T   I   L   E   Y   S   G   I
ACC CAC AAC GTT TAC GCT AAG CAC GAG GTC CTC CTG GCC GCG GGC TCC GCT GTC TCT CCC ACA ATC CTC GAA TAT TCC GGT ATC
                            950                                                                1000
 G   M   K   S   I   L   E   P   L   G   I   D   T   V   V   D   L   P   V   G   L   N   L   Q   D   Q   T   T
GGA ATG AAG TCC ATC CTG GAG CCC CTT GGT ATC GAC ACC GTC GTT GAC CTG CCC GTC GGC TTG AAC CTG CAG GAC CAG ACC ACC
                                                1050
 A   T   V   R   S   R   I   T   S   A   G   A   G   Q   G   Q   A   A   W   F   A   T   F   N   E   T   F   G
GCT ACC GTC CGC TCC CGC ATC ACC TCT GCT GGT GCA GGA CAG GGA CAG GCC GCT TGG TTC GCC ACC TTC AAC GAG ACC TTT GGT
1100                                                                1150
 D   Y   S   E   K   A   H   E   L   L   N   T   K   L   E   Q   W   A   E   E   A   V   A   R   G   G   F   H
GAC TAT TCC GAA AAG GCA CAC GAG CTG CTC AAC ACC AAG CTG GAG CAG TGG GCC GAA GAG GCC GTC GCC CGT GGC GGA TTC CAC
                        1200                                                                1250
 N   T   T   A   L   L   I   Q   Y   E   N   Y   R   D   W   I   V   N   H   N   V   A   Y   S   E   L   F   L
AAC ACC ACC GCC TTG CTC ATC CAG TAC GAG AAC TAC CGC GAC TGG ATT GTC AAC CAC AAC GTC GCG TAC TCG GAA CTC TTC CTC
                                            1300                                                                1350
 D   T   A   G   V   A   S   F   D   V   W   D   L   L   P   F   T   R   G   Y   V   H   I   L   D   K   D   P
GAC ACT GCC GGA GTA GCC AGC TTC GAT GTG TGG GAC CTT CTG CCC TTC ACC CGA GGA TAC GTT CAC ATC CTC GAC AAG GAC CCC
                                                                1400
 Y   L   H   H   F   A   Y   D   P   Q   Y   F   L   N   E   L   D   L   L   G   Q   A   A   A   T   Q   L   A
TAC CTT CAC CAC TTC GCC TAC GAC CCT CAG TAC TTC CTC AAC GAG CTG GAC CTG CTC GGT CAG GCT GCC GCT ACT CAA CTG GCC
                1450                                                                1500
 R   N   I   S   N   S   G   A   M   Q   T   Y   F   A   G   E   T   I   P   G   D   N   L   A   Y   D   A   D
CGC AAC ATC TCC AAC TCC GGT GCC ATG CAG ACC TAC TTC GCT GGG GAG ACT ATC CCC GGT GAT AAC CTC GCG TAT GAT GCC GAT
                                    1550                                                                1600
 L   S   A   W   T   E   Y   I   P   Y   H   F   R   P   N   Y   H   G   V   G   T   C   S   M   M   P   K   E
TTG AGC GCC TGG ACT GAG TAC ATC CCG TAC CAC TTC CGT CCT AAC TAC CAT GGC GTG GGT ACT TGC TCC ATG ATG CCG AAG GAG
                                                                1650
 M   G   G   V   V   D   N   A   A   R   V   Y   G   V   Q   G   L   R   V   I   D   G   S   I   P   P   T   Q
ATG GGC GGT GTT GTT GAT AAT GCT GCC CGT GTG TAT GGT GTG CAG GGA CTG CGT GTC ATT GAT GGT TCT ATT CCT CCT ACG CAA
                1700                                                                1750
 M   S   S   H   V   M   T   V   F   Y   A   M   A   L   K   I   S   D   A   I   L   E   D   Y   A   S   M   Q
ATG TCG TCC CAT GTC ATG ACG GTG TTC TAT GCC ATG GCG CTA AAA ATT TCG GAT GCT ATC TTG GAA GAT TAT GCT TCC ATG CAG
                                    1800                                                                1850
TGAGTGGTATGATGGGGATATGAGTGAGGATATTAGGGGATGGTACTTAGATGCTGGGGAGGTATAATCATAGATTGGATAGAATTGGTAGGTTACATAGACAGCTTACATG
                                                1900                                                1950
AATAGACGTTCGTTATATGTGAGCAGACATTACTACCAAACAAGGGCATTGTTCAGTT
                                2000
```

EMBL3. From 12000 recombinant plaques screened with the nick-translated 800 bp cDNA fragment one hybridizing clone was isolated which contained an insert of 15 kbp. The phage DNA was cleaved with *Sal*I and the resulting fragments were subcloned into pBluescript SK(+). Using hybridisation techniques and the shotgun sequencing method we could identify an 1.8 kbp and a 2.0 kbp *Sal*I fragment containing the coding region of GOD as well as small 5′- and longer 3′-untranslated regions (fig.2). The overall DNA sequence of the gene is represented in fig.3.

12 bp upstream of the ATG start codon the 5′-noncoding region shows a CT-rich sequence. These regions are characteristic for highly expressed genes or genes lacking the TATA and CAAT boxes from *S. cerevisiae* and several filamentous fungi. Therefore, it is suggested that these CT regions may act as promoter elements. In filamentous fungi one transcription-initiation site often appears in or immediately downstream from these pyrimidine-rich sequences. Two sites (CATC) showing homology to known *A. niger* transcription-initiation sites were found within the region between the CT sequence and the start codon.

The structural gene consists of 1815 bp, encoding 605 amino acid residues. The mature protein contains 584 amino acids. The difference is due to 21 amino acids comprising the signal sequence which shows an unusual feature. The amino acid in position −1 is a basic one, namely arginine, which causes an unexpected cleavage at the arginine-serine bond. At present, we cannot identify introns in the GOD sequence. The deduced amino acid sequence of the coding region is almost identical to the known parts of the protein sequence, although a small number of amino acid exchanges were observed (fig.3). These may reflect minor differences between the GOD protein obtained from Boehringer Mannheim, and the corresponding enzyme of *A. niger* NRRL-3.

Furthermore, a 3′-untranslated region of 167 nucleotides including a potential polyadenylation signal (AATAGA) was found (fig.3).

Acknowledgements: This work was supported by the Bundesminister für Forschung und Technologie, Förderungskennzeichen 0318914A. We wish to thank Jürgen Stollwerk for the synthesis of oligonucleotides.

REFERENCES

[1] Pazur, J.H., Kleppe, K. and Cepure, A. (1965) Arch. Biochem. Biophys. 111, 351–357.
[2] O'Malley, J.J. and Weaver, J.L. (1972) Biochemistry 11, 3527–3532.
[3] Hayashi, S. and Nakamura, S. (1981) Biochim. Biophys. Acta 657, 40–51.
[4] Lockwood, L.B. (1975) in: The Filamentous Fungi (Smith, E.J. and Berry, D.R. eds) vol. 1, pp. 140–145, Edward Arnold, London.
[5] Mischak, H., Kubicek, C.P. and Röhr, M. (1985) Appl. Microbiol. Biotechnol. 21, 27–31.
[6] Hanahan, D. (1985) in: DNA Cloning (Glover, D.M. ed.) vol. I, pp. 109–135, IRL, Oxford.
[7] Yanish-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 109–119.
[8] Frischauf, A.-M., Lehrach, H., Poustka, A. and Murray, N. (1983) J. Mol. Biol. 170, 827–842.
[9] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) in: Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[10] Chirgwin, J.M., Przybula, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294–5299.
[11] Wirsel, S., Lachmund, A., Wildhardt, G. and Ruttkowski, E. (1989) Mol. Microbiol. 3, 3–14.
[12] Yelton, M.M., Hammer, J.E. and Timberlake, W.E. (1984) Proc. Natl. Acad. Sci. USA 81, 1470–1474.
[14] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[15] Fröschle, M., Ulmer, W. and Jany, K.-D. (1984) Eur. J. Biochem. 142, 533–540.
[16] Ulmer, W., Fröschle, M. and Jany, K.-D. (1983) Eur. J. Biochem. 136, 183–194.