ELSEVIER

Biochimica et Biophysica Acta 1185 (1994) 243–246

BBA Biochimica et Biophysica Acta

Short Sequence-Paper

Cloning and nucleotide sequence of the gene for NADH: FMN oxidoreductase from *Vibrio harveyi*

Yoshitaka Izumoto [a,1,*], Takashi Mori [a], Kazumi Yamamoto [b]

[a] *Bio-technology Research Laboratory, Chemical Research Institute, Sekisui Chemical Co., Ltd., Osaka, Japan*
[b] *Tsuruga Institute of Biotechnology, Toyobo Co., Ltd., Fukui, Japan*

(Received 21 December 1993)

Abstract

The gene encoding the enzyme NADH: FMN oxidoreductase (EC 1.6.99.3) from *Vibrio harveyi* has been isolated from a recombinant library of genomic DNA and sequenced. The deduced amino acid sequence, 237 amino acids long, shows 48% identity with *E. coli* NAD(P)H:flavin oxidoreductase and 40% identity with *Vibrio harveyi lux*G gene product.

Key words: Nucleotide sequence; Flavin reductase; NADH: FMN oxidoreductase; *lux*G; (*V. harveyi*)

NAD(P)H: FMN oxidoreductases were discovered in luminescent marine bacteria as components of a *lux* multiprotein system that catalyzes the light emitting reaction where the oxidoreductases, a luciferase, and a fatty acid reductase complex react [1]. In *Vibrio harveyi*, the NADH- and NADPH- specific oxidoreductases reducing FMN are discriminated and there exists an oxidoreductase specific for both cofactors [2]. Each of the NADH- and NADPH- specific oxidoreductases was obtained in an apparently homogeneous form. In solution they existed as a monomer with a molecular mass of 30 kDa, 40 kDa, respectively [3]. We report here the cloning and nucleotide sequence of the NADH: FMN oxidoreductase gene.

The NADH: FMN oxidoreductase was purified as described by Watanabe and Hastings [2] from *Vibrio harveyi* mutant strain MB-20, and subjected to amino acid sequencing: N-terminal sequence of 20 amino acids was obtained. The protein was digested with lysylendopeptidase and resulting two peptides were analyzed for N-terminal amino acid sequences. Two mixed polymerase chain reaction primers were constructed from this information. A sense oligonucleotide, 5'-GGGGATCCAC(A or C or G or T)AT(A or C or T)CA(A or G)TG(C or T)AA(A or G)GT-3', was based on the N-terminal amino acid sequence TIQCKV (equivalent to residues 1–6 in Fig. 2). An antisense oligonucleotide, 5'GGGGATCCGG(A or C or G or T)AC(A or G)AA(A or G)TG(A or G or C or T)AC(A or G)TT-3', was based on the internal amino acid sequence NVHFVP (equivalent to residues 164–169 in Fig. 2). Each primer had a restriction enzyme *Bam*HI recognition site linker, GGGGATCC, added to the 5' end. A part of the NADH: FMN oxidoreductase gene was amplified from *Vibrio harveyi* genomic DNA with these primers and the amplified 0.5 kb DNA fragment was successfully used as a probe to screen a λZAPII genomic library of *Vibrio harveyi*. A clone that contained a 5.4 kb *Eco*RI fragment of the *Vibrio harveyi* chromosome carried the NADH: FMN oxidoreductase gene. The DNA sequencing strategy employed is shown in Fig. 1. The nucleotide sequence of a 1431 bp *Ssp*I-*Stu*I region of the 5.4 kb *Eco*RI fragment was determined and shown in Fig. 2. This nucleotide sequence contains a 711 bp open reading frame, in which the determined N-terminal and two internal amino acid sequences could be located from the nucleotide sequence beginning with an ACC codon at position 1, ending with a TAA stop codon at position 709. Consis-

* Corresponding author (present address). Osaka University Medical School, Department of Medicine III, 2-2 Yamada-oka Suita-City, Osaka 565, Japan. Fax: +81 6 8793839.
[1] Present address: Osaka University Medical School, Department of Medicine III, Osaka, Japan.

The nucleotide sequence data reported in this paper have been submitted to the GenBank database under the accession number D14674.

SSDI 0005-2728(94)00011-S

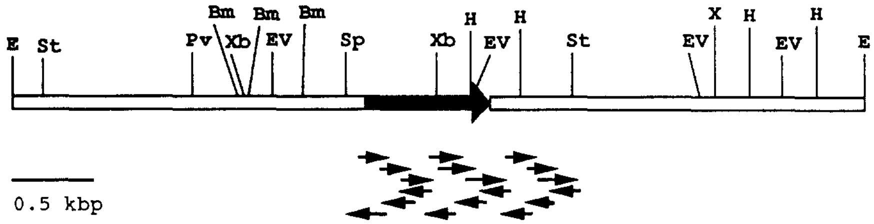

Fig. 1. Restriction map and sequencing strategy of the NADH:FMN oxidoreductase gene and flanking sequences. The position of the coding region is indicated by a thick arrow. The nucleotide sequence of the *SspI-StuI* fragment was obtained using nested deletions generated by the exonuclease III/mung bean nuclease method [12] and custum oligonucleotide primers. Thin arrows indicate the direction and extent of each sequencing run. All sequencing reactions were performed using the Sequenase (U.S. Biochemical) chain termination protocol. Abbreviations: E,*Eco*RI; St,*Stu*I; Pv,*Pvu*II; Bm,*Bam*HI; Xb,*Xba*I; EV,*Eco*RV; Sp,*Ssp*I; H,*Hind*III; X,*Xho*I.

tent with the N-terminal amino acid sequence is the ATG at position −3, which could serve as start codon. The coding region is preceded by a putative ribosome-binding site, GAGG, at nucleotide −7 to −10. Upstream 39 bp from the ATG is a potential promoter with AGTAAT(TATAAT) at −10 and TTGCCAA

```
                                                   Ssp I
-200                                               AATATTCGCTTCCCACATCGTTGCCAAGTTGGCGCTTGCGCC    -159
-158  ATGTGCATGTGTAAAAAGCTAACGGGCGAGGTGAGCTATCACCTCGAACCTATGCTAACCGAAAAAGAGCAACAGCAAGG    -79

                    -35                   -10                                RBS         M        -1
 -78  TTGGATATTCCCTTGCCAAGCCTATACAGAAAGTAATTTAGTGCTTACTTTTGACGAGTAAGCGAGAGGAACTCC ATG      -1

   1   T   I   Q   C   K   V   K   S   I   Q   P   L   A   C   N   T   Y   Q   I   L        20
   1  ACC ATC CAA TGT AAA GTA AAG TCT ATT CAG CCG TTA GCT TGT AAT ACT TAT CAA ATC CTT       60

  21   L   H   P   E   S   P   V   P   F   K   A   G   Q   Y   L   M   V   V   M   G        40
  61  CTT CAC CCA GAA TCA CCT GTA CCT TTT AAA GCA GGT CAG TAC CTC ATG GTT GTG ATG GGT      120

  41   E   K   D   K   R   P   F   S   I   A   S   S   P   C   R   H   E   G   E   L        60
 121  GAA AAA GAC AAA CGT CCT TTC TCG ATT GCG AGC AGT CCA TGT CGT CAT GAA GGT GAA CTT      180

  61   E   L   H   I   G   A   A   E   H   N   A   Y   A   L   E   V   V   E   A   M        80
 181  GAA CTG CAT ATC GGT GCG GCG GAA CAC AAC GCT TAT GCG CTA GAA GTC GTT GAA GCA ATG      240

  81   Q   A   A   L   E   T   D   G   H   I   E   I   D   A   P   H   G   D   A   W       100
 241  CAA GCG GCA TTA GAA ACA GAT GGT CAT ATC GAG ATT GAT GCT CCA CAT GGT GAT GCT TGG      300

 101   V   Q   E   E   S   E   R   P   L   L   L   I   A   G   G   T   G   F   S   Y       120
 301  GTT CAA GAA GAA AGC GAA CGC CCA CTA TTA TTG ATT GCT GGT GGT ACT GGT TTT AGT TAC      360

 121   V   R   S   I   L   D   H   C   V   A   Q   N   K   T   N   P   I   Y   L   Y       140
 361  GTG CGT TCA ATT CTA GAT CAC TGT GTT GCA CAG AAC AAA ACC AAC CCT ATC TAT CTA TAC      420

 141   W   G   A   R   D   N   C   Q   L   Y   A   K   E   E   L   V   E   I   A   D       160
 421  TGG GGG GCG CGT GAT AAC TGT CAG TTG TAC GCT AAA GAA GAG TTG GTC GAG ATT GCC GAC      480

 161   K   F   A   N   V   H   F   V   P   V   V   E   E   A   P   A   D   W   Q   G       180
 481  AAG TTT GCT AAT GTT CAC TTT GTG CCA GTA GTA GAA GAA GCG CCA GCA GAC TGG CAA GGT      540

 181   K   V   G   N   V   L   Q   A   V   S   E   D   F   E   S   L   E   N   Y   D       200
 541  AAA GTT GGT AAC GTG CTA CAA GCG GTG AGT GAA GAT TTC GAA AGC TTA GAA AAC TAC GAT      600

 201   I   Y   I   A   G   R   F   E   M   A   G   A   A   R   E   Q   F   T   Q   N       220
 601  ATC TAT ATT GCA GGT CGT TTC GAA ATG GCT GGC GCA GCA CGT GAA CAG TTC ACT CAG AAC      660

 221   K   K   A   K   S   E   R   M   F   A   D   A   Y   A   F   I   *                   236
 661  AAG AAA GCA AAG AGC GAA CGT ATG TTC GCA GAC GCG TAC GCA TTC ATT TAAATACAGCTTTTGA     724

 725  GGCAGAAAAAGAGGGTTTTTACCCTCTTTTTTGCTTTTTTGATGAAATAGTCATCGAACAGTTAGTTTTTTCCATTTTT    804
 805  TTTCAAAAAAGGGTTGCGAACGGATCTGAGTTCCCTATAATGCGCATCCACCGACACGGCAGACGCGATAAGGCTTCAGC    884
 885  AGGGTCGGAGAGGTGAAAAGCTTCTGAGAAAATAAATTGAAAAAAGTGTTTGACACTCTCAATTATCTCGTTAGAATGCA    964
 965  CCTCCGCTTTGAGAGAAAAACTTCTCGATAAGCAAGCTCTTTAACAATATAGACCTATCAATCTGTGTGGGCACTCGTTG   1044
1045  ATGATAATCCAATTAGATACTTCGGTATCAAATTAGGTTTCAATGAAACGAAGTGACCATTGAATCGAAAGATTCAGCAC   1124
1125  AGTCAATTCAAACATTACTTATGTAATGTTCAGTATTCATTGAGCCGAACAAAATCTTAAATTGAAGAGTTTGATCATGG   1204
                            Stu I
1205  CTCAGATTGAACGCTGGCGGCAGGCCT                                                        1231
```

Fig. 2. Nucleotide sequence of the NADH:FMN oxidoreductase gene and flanking regions. The complete nucleotide sequence of the 1431 bp *Ssp*I-*Stu*I fragment is listed along with the predicted amino acid sequence of the NADH:FMN oxidoreductase. The −10 and −35 regions of the putative promoter are indicated by underlines, as is the ribosome-binding site(RBS). Converging arrows indicate inverted repeats. Amino acid residues underlined were confirmed by protein sequencing.

(TTGACAT) at −35 (consensus sequences from reference [4]) separated by 12 nucleotides. At 18 bp downstream from the stop codon, there is a palindromic sequence between nucleotide 731 and 757. The free energy of this structure (−22.8 kcal/mol) would be expected to make it an efficient transcription terminator. The mature protein presumably consisting of 236 amino acid residues has a calculated molecular weight of 26350. This is slightly lower than the value of 30000 ± 2000 estimated with a calibrated Sephadex G-100 column on the purified enzyme[3]. The amino acid composition of the purified enzyme was determined. Table 1 compared the result with that deduced from the open reading frame. They are in close agreement, suggesting that we have cloned the gene for the *Vibrio harveyi* NADH:FMN oxidoreductase.

Protein database searches were performed and significant homology was found with NAD(P)H:flavin oxidoreductase of *E. coli* (48% identity) and *lux*G product of *Vibrio harveyi* (40% identity) (Fig. 3). The *E. coli* NAD(P)H:flavin oxidoreductase(flavin reductase enzyme, Fre) [5] was discovered as a component of a complex multiprotein system that catalyzes the transformation of an inactive form of ribonucleotide reductase into an active enzyme. The Fre produces reduced flavins, and as such these are capable of reducing the Fe(III) center of the ribonucleotide reductase to the Fe(II) [6]. The remarkable homology between the

Table 1
Amino acid composition of the NADH:FMN oxidoreductase

Amino acid	Residues/molecule	
	measured	predicted
Asx	21.8	20
Thr	7.4	6
Ser	11.4	11
Glx	36.7	36
Gly	16.6	14
Ala	27.7	29
Val	15.6	17
Met	4.8	5
Ile	13.3	14
Leu	17.8	17
Tyr	10.0	10
Phe	10.1	10
Lys	13.1	12
His	7.3	8
Arg	8.1	8
Pro	10.6	11
Cys	5.6	5
Trp	0.4	3
Total		236

A purified enzyme sample (0.6 nmol) was hydrolyzed in vacuo at 110°C for 22 h in 6M HCl containing 0.2% phenol (11). After hydrolysis the sample was evapolated to dryness and dissolved in 100 μl of water. The amino acid analysis was performed with a Hitachi model 835 amino acid analyzer. The number of residues from the DNA sequence was obtained from the predicted protein sequence shown in Fig. 2, except that the first methionine residue was excluded.

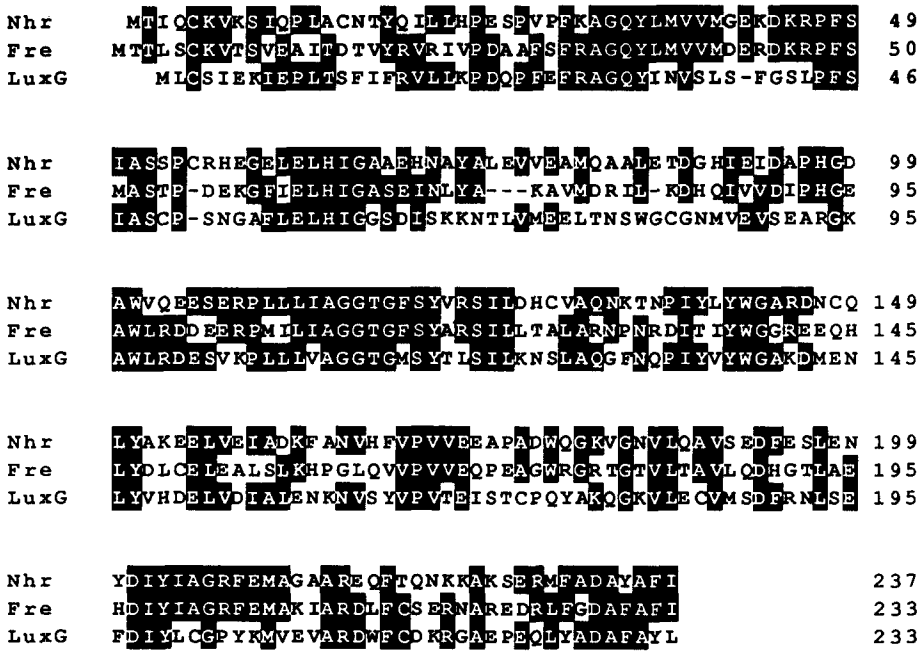

Fig. 3. Alignment of the deduced amino acid sequences of the *Vibrio harveyi* NADH:FMN oxidoreductase(Nhr), the *E. coli* NAD(P)H:FMN oxidoreductase(Fre), and the *Vibrio harveyi lux*G product(LuxG). Amino acid residues identical are indicated by black background. Homology analysis was carried out with the DNASIS program version 2.0(Hitachi).

NADH : FMN oxidoreductase and Fre suggests a similar or even identical physiological role for the two proteins. The *lux*G gene has so far been found in the *lux* operons of three different species of luminescent bacteria (*V. harveyi* [7], *V. fischeri* [8], and *P. leiognathi* [9]). The close relationship of the amino acid sequence and molecular weight of the LuxG protein with the NADH : FMN oxidoreductase possibly implicates it in producing $FMNH_2$ for the luminescent reaction. However, transposon mutagenesis of the *V. fischeri lux* system has shown that all transposon insertions that block luminescence were located within the two regulatory genes (*lux*R,I) or five structural genes (*lux*CDABE) [10].

References

[1] Meighen, E.A. (1993) FASEB J. 7, 1016–1022.
[2] Watanabe, H. and Hastings, J.W. (1982) Mol. Cell. Biochem. 44 181–187.
[3] Jablonski, E. and DeLuca, M. (1977) Biochemistry 16, 2932–2936.
[4] McClure,W.R. (1985) Annu. Rev. Biochem. 54, 172–204.
[5] Spyrou, G., Haggård-Ljungquist, E., Krook, M., Jörnvall, H., Nilsson, E. and Reichard, P. (1991) J. Bacteriol. 173, 3637–3679.
[6] Fontecave, M., Eliasson, R. and Reichard, P. (1987) J. Biol. Chem. 262, 12325–12331.
[7] Swartzman, E., Miyamoto, C., Graham, A. and Meighen, E. (1990) J. Biol. Chem. 265, 3513–3517.
[8] Swartzman, E., Kapoor, S., Graham, A.F. and Meighen, E.A. (1990) J. Bacteriology 172, 6797–6802.
[9] Lee, C.Y., Szittner, R.B. and Meighen, E.A. (1991) Eur. J. Biochem. 201, 161–167.
[10] Engebrecht, J., Nealson, K. and Silverman, M. (1983) Cell 32, 773–781.
[11] Lundell, D.J. and Howard, J.B. (1978) J. Biol. Chem. 253, 3422–3426.
[12] Henikof, S. (1984) Gene 28, 351–359.