Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 4225–4229, May 1991
Biochemistry

cDNA-derived amino acid sequence of the NADH-binding 51-kDa subunit of the bovine respiratory NADH dehydrogenase reveals striking similarities to a bacterial NAD^+-reducing hydrogenase

(*Alcaligenes eutrophus*/mRNA levels in tissues/human genomic sequences)

SALIL D. PATEL*, RUEDI AEBERSOLD†, AND GIUSEPPE ATTARDI

Division of Biology, California Institute of Technology, Pasadena, CA 91125

Contributed by Giuseppe Attardi, January 22, 1991

ABSTRACT **A λgt10 bovine brain and a λgt11 bovine heart cDNA library were screened with oligonucleotide probes corresponding to partial protein sequences directly determined from the isolated 51-kDa subunit of the bovine respiratory-chain NADH dehydrogenase. Clones were isolated that encode a protein of 464 amino acids containing all the 11 partial tryptic peptide sequences determined from the 51-kDa subunit. The size and amino acid composition of this protein agree with those determined for the purified 51-kDa subunit. Furthermore, this protein contains a putative NADH-binding domain, a possible FMN-binding site, and a putative binding site for an iron-sulfur cluster. The above evidence indicates that the cloned protein is the 51-kDa subunit or its precursor. A search for sequence similarity with proteins in the Protein Identification Resource data base has revealed that the 51-kDa subunit has 32% amino acid sequence identity with a major portion of the α subunit of the soluble NAD^+-reducing hydrogenase from *Alcaligenes eutrophus*. In particular, there are three segments of high sequence similarity (70–88%) between the two proteins which correspond to the three ligand-binding sites.**

The NADH:ubiquinone oxidoreductase or rotenone-sensitive NADH dehydrogenase (EC 1.6.5.3) is the first enzyme of the mitochondrial respiratory chain (for a review see ref. 1). The molecular genetic analysis of this enzyme complex (complex I) started a few years ago with the discovery that 7 of the ≈25 subunits of the human enzyme (2, 3) and 6 of the ≈22 subunits of the *Neurospora crassa* enzyme (4) are encoded in mitochondrial DNA. Recent efforts have been directed toward the cloning and structural analysis of nuclear-encoded subunits of the flavoprotein and iron-protein fragments of the enzyme. Of the 3 polypeptides that constitute the bovine flavoprotein fragment—i.e., the 51-kDa, 24-kDa, and 9-kDa subunits (1, 5)—the cDNA for the 24-kDa subunit has already been cloned (6, 7). Similarly, of the 6 polypeptides identified in the iron-protein fragment (1), the cDNAs for the 49-kDa (8) and the 75-kDa (9) subunit have been cloned.

In the present work, the cDNA for the 51-kDa subunit of the flavoprotein fragment has been cloned by screening cDNA libraries with oligonucleotides corresponding to partial protein sequences directly determined from the isolated subunit.‡ A search for sequence similarities with proteins in the Protein Identification Resource data base has unexpectedly revealed the existence in the 51-kDa subunit of regions strikingly similar to segments of the soluble NAD^+-reducing hydrogenase of *Alcaligenes eutrophus* (10).

MATERIALS AND METHODS

Isolation and Sequence Analysis of the 51-kDa Subunit. Complex I was isolated from bovine heart mitochondria according to Hatefi *et al.* (11) and resolved by perchlorate into the flavoprotein fragment, the iron-protein fragment, and the hydrophobic protein fragment, as described (12). Since preliminary tests indicated that the NH_2 terminus of the 51-kDa subunit was blocked, the sequence of internal peptides was analyzed. Two hundred micrograms of the flavoprotein fragment was electrophoresed in an SDS/polyacrylamide gel, and the separated proteins were electroblotted onto nitrocellulose (Schleicher & Schuell; 0.45 μm), with 20% methanol/0.195 M glycine/25 mM Tris, pH 8.3, as transfer buffer. The nitrocellulose filter was stained with amido black, the region where the 51-kDa subunit was bound was sliced out, and the bound protein was digested on the matrix with trypsin (13). The resulting peptides were separated by HPLC (13) on a Vydac C_4 column (150 × 2.1 mm) in a Waters peptide analyzer, using a trifluoroacetic acid/acetonitrile buffer system (14). Prominent peptides were sequenced on an Applied Biosystems model 477A sequenator, using standard protocols. Table 1 shows the sequences of the peptides thus determined.

Oligonucleotide Synthesis. Some of the determined peptide sequences of the 51-kDa protein, which had a relatively low codon degeneracy, were used to design mixtures of oligonucleotides for library screening (Table 1). The minimum dissociation temperatures were estimated as described (15). Inosinic acid (I) was substituted, whenever possible, for deoxyguanylic acid, deoxyadenylic acid, and thymidylic acid, since I can base-pair with C, T, and A (16).

Isolation of Clones. Two libraries, a λgt11 bovine heart (Clontech) and a λgt10 bovine brain library, were screened with a gel-purified 256-fold degenerate 26-mer, corresponding to the peptide sequence K (Table 1), labeled with [γ-^{32}P]ATP (Amersham) and polynucleotide kinase (Boehringer Mannheim). Duplicate nitrocellulose filter (Schleicher & Schuell) replicas from plates were treated according to standard protocols and then hybridized with the synthetic oligonucleotides (0.6 pmol/ml) for 16 hr at 56°C, as described (7). Seventeen positive clones thus isolated from the λgt10 library and five positive clones isolated from the λgt11 library were rescreened with the oligonucleotide mixtures corresponding to peptide sequences G and J in Table 1, and, respectively, the oligonucleotide mixtures corresponding to peptide sequences A and B, at temperatures 5°C below the minimum dissociation temperature of the mixtures. Phages were purified and the recombinant λ DNA was isolated as described (17). The inserts were excised by digestion with *Eco*RI, purified by electrophoresing them directly onto

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*Present address: Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305.
†Present address: Biomedical Research Center, University of British Columbia, Vancouver VGT 1W5, BC, Canada.
‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. M63009).

Table 1. Sequences of tryptic peptides from the 51-kDa subunit and corresponding oligonucleotides

Peptide	Peptide sequence*	Oligonucleotide mixture (5′ → 3′)†
A	41 IFTNLYG(R)	CCRTAOARITTOGTRAAOAT (64, −)
B	76 ILGEVQ	ATOCTOGGOGAIGTOCA (16, +) ATOTTIGGOGAIGTOCA (8, +)
	. . . . K	
C	90 GAGFPTGLKPGT	
	. . . . . WSF	
D	153 AAYIYIR	
E	160 GEFYNEASNL	
F	175 EAYEAGLIGK	
G	226 KPPFPADVGVFG	CCIAAOACOCCOACITCOGCO-GGIAAOGGOGGYTT (256, −)
H	370 LIEFYK	
I	387 TGVDWMNK	
	E	
J	407 AEIDSL(WEIS)	AIOGAITCIATYTCOGC (8, −) AIICTITCIATYTCOGC (4, −)
K	451 QQFAQQHQA	CARCARTTYGCOCARCAR-CAYCARGC (256, +)

*The position in the cDNA-derived protein sequence of the first amino acid of each peptide is indicated. The cDNA-derived sequence, if different from the directly determined peptide sequence, is shown underneath the latter, with identical residues indicated by dots. Amino acids in parentheses were not used to derive oligonucleotide sequences.

†Degeneracy of the oligonucleotide mixture and the strand polarity of the sequence are given in parentheses. R = A or G; Y = C or T; O = I or C.

DEAE-nitrocellulose paper, and then subcloned into the pBluescript KS(+) vector (Stratagene).

DNA Sequence Analysis. One λgt10-derived cDNA clone (pBND51-1) and one λgt11-derived clone (pBND51-2) were sequenced by the dideoxy chain-termination method (18) as modified for double-stranded DNA plasmid sequencing (19). Both strands of each clone were fully sequenced using 25 synthetic primers, Sequenase (version 1.0, United States Biochemical), and [5′-[α-^{35}S]thio]dATP. In some cases dITP and ddITP were used instead of dGTP and ddGTP, respectively, to reduce compression in the sequencing gels.

RNA and DNA Analysis. RNA was isolated from bovine tissues, fractionated by electrophoresis through 1.5% agarose/2.2 M formaldehyde gels, transferred from the gels onto Zeta-Probe nylon membranes (Bio-Rad) by electroblotting, and hybridized with the insert of pBND51-1, ^{32}P-labeled by random priming, essentially as described (7). HeLa cell DNA was digested with several restriction endonucleases, electrophoresed in a 0.7% agarose gel, denatured, transferred by capillarity onto a Zeta-Probe nylon membrane, and hybridized with the ^{32}P-labeled insert of pBND51-2, as described (7), except that hybridization was carried out at 68°C in 5× standard saline citrate.

Data Analysis. DNA sequences were analyzed, and the protein sequence of the 51-kDa subunit was compared with those in the Protein Identification Resource data base (National Biomedical Research Foundation, release 26) by the computer programs in DNA-MASTER (written by Tim Hunkapiller). Hydrophobicity analysis was done using the program SOAP (20).

RESULTS

Isolation and Sequencing of the cDNA Clones Encoding the 51-kDa Subunit. Screening of 400,000 plaques from a λgt10 bovine brain library and 200,000 plaques from a λgt11 bovine heart library with an oligonucleotide mixture corresponding to peptide sequence K (Table 1) yielded 22 putative positive clones (17 from the λgt10 library and 5 from the λgt11 library). These clones were purified to homogeneity, and the λgt10 clones were rescreened with oligonucleotide mixtures corresponding to peptides G and J.

Five λgt10 library clones hybridized strongly with all three oligonucleotide mixtures. Two of them, which contained inserts of identical size [≈1300 base pairs (bp)], were further analyzed. The 1300-bp insert was subcloned into the pBluescript KS(+) vector (to give pBND51-1) and then sequenced completely on both strands by use of synthetic oligonucleotide primers (Fig. 1). The sequence revealed a 1325-bp insert containing a poly(A)-addition signal (21) at position 1288, followed by a 19-residue poly(A) tail 17 bp downstream from the signal. The insert contained a reading frame of 415 amino acids, which included all the peptide sequences obtained by direct sequencing of the 51-kDa subunit, except peptide sequence A (IFTNLYG).

The above data suggested that pBND51-1 lacked a 5′ segment of the desired cDNA. Thus, oligonucleotide mixtures corresponding to the most NH_2-terminal peptide sequence (B) and to the missing peptide sequence (A) were used to rescreen the five positive λgt11 clones. One clone hybridized with both oligonucleotide mixtures, and its insert was isolated, subcloned into the pBluescript KS(+) vector (to give pBND51-2), and sequenced as described above. pBND51-2 was found to contain an insert of 1555 bp, encompassing a reading frame of 490 amino acids. This reading frame contains the missing peptide sequence A at amino acid positions 67–74; furthermore, a potential initiation codon exists at amino acid position 27 and a termination codon at position 490. The nucleotide sequence of the insert of

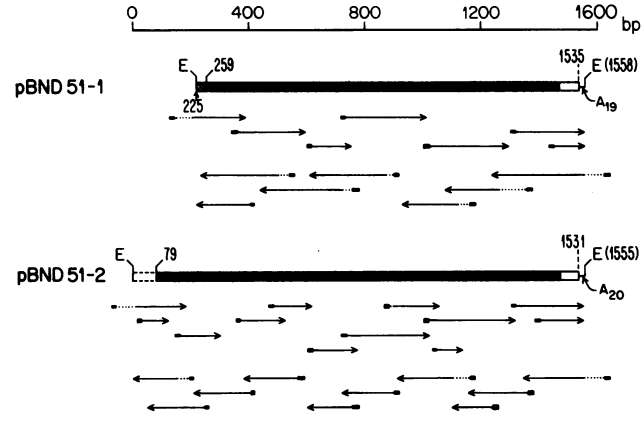

FIG. 1. Sequencing strategy for the inserts of the 51-kDa-subunit cDNA clones. The inserts of pBND51-1 and pBND51-2 have been aligned on the basis of the cDNA-derived protein sequences. The scale above the cDNA inserts indicates the position in base pairs (bp) relative to the 5′ end of the pBND51-2 insert. In the two inserts, the black bar represents the protein-coding region; the open bar, a 3′ untranslated region; the hatched bar, a sequence unrelated to the 51-kDa cDNA; the open bar with dashed contour, a segment of reading frame of uncertain significance; and the lines, portions of the clones pertaining to the poly(A) tails. The arrows represent the direction and extent of sequencing from the indicated sites. The black blocks at the origins of the arrows represent the positions of hybridization of the synthetic oligonucleotide primers used for sequencing. E, *Eco*RI.

pBND51-2 from nucleotide 258 to the 3′ end is identical to that of the insert of pBND51-1 from nucleotide 34 to the 3′ end, except for an extra 4 nucleotides in the nontranslated region immediately upstream of the poly(A) tail and an extra adenine residue in this tail. Correspondingly, the reading frame of pBND51-2 from amino acid residue 87 to residue 490 matches perfectly that of the reading frame of pBND51-1 from residue 12 to residue 415. Upstream of nucleotide 258, the sequence of the insert of pBND51-2 diverges completely from that of pBND51-1. The latter contains an extra 33 bases encoding 11 amino acids that have no similarity to the corresponding region of the reading frame of pBND51-2. It seems very likely that this segment was incorporated into pBND51-1 as a result of a cloning artifact. In Fig. 2, the complete sequence of the insert of pBND51-2 is shown, and the divergent sequences present in pBND51-1 are indicated. In numbering the amino acids of the reading frame, the methionine at position 27 in pBND51-2 was chosen as NH_2-terminal amino acid (see *Discussion*).

Identification of the 51-kDa-Subunit mRNA in Various Bovine Tissues. RNA transfer hybridization analysis of total RNA from bovine tissues, utilizing the insert from pBND51-1 as a probe, revealed in each case a single hybridizing RNA species (Fig. 3) with an estimated size of ≈1650 nucleotides. The intensities of the bands indicated a widely varying level of expression of the 51-kDa gene in different tissues. From a densitometric analysis of the autoradiograms it was estimated that the concentration of 51-kDa mRNA in heart RNA is 15-fold, in kidney RNA ≈4-fold, and in brain RNA ≈2-fold higher than in liver RNA.

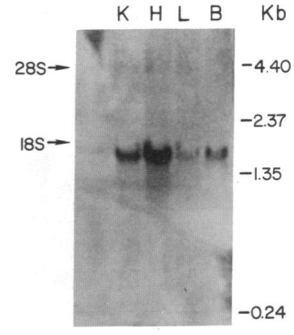

FIG. 3. RNA transfer hybridization analysis of bovine tissue RNAs. Samples of total RNA (20 μg) from various tissues were electrophoresed, electroblotted, and hybridized with the ^{32}P-labeled insert of pBND51-1. The positions of migration of 28S and 18S rRNA and commercial RNA markers (Kb, kilobases) are indicated. The estimated size of the 51-kDa mRNA is 1650 nt. Lanes: K, kidney; H, heart; L, liver; B, brain.

Southern Blot Analysis of Human Genomic DNA. A Southern blot analysis of human genomic DNA digested with seven different restriction enzymes, utilizing the ^{32}P-labeled insert of pBND51-2 as a probe, revealed simple patterns that consisted of one band (*Eco*RV and *Ssp* I), or two bands (*Eco*RI, *Bgl* II, *Kpn* I, and *Pvu* II), or three bands (*Hin*dIII)

```
   1  AATTCGCGTCTGAGCCTCAGAGCTGTGAAGGTGACAGCGGGCGGTCACTGGCTTGCCTTTCCTCCCCCCATCCCCGCGATGCTGGCGGCACGGCGGCTGCTCGGCGGGTCGCTCCCCGCG   120
 -26  N S R L S L R A V K V T A G G H W L A F P P P I P A M L A A R R L L G G S L P A    14

                                                                                                   pBND51-1    AATTCCGGCGGGGCTG
                                                                                                                I P A G L
 121  CGGGTGTCTGTGCGATTCAGCGGCGACACGACAGCGCCCAAGAAAACCTCATTTGGCTCGCTGAAGGACGAAGACCGGATCTTCACCAACCTGTATGGCCGCCATGACTGGAGGCTTAAA   240
  15  R V S V R F S G D T T A P K K T S F G S L K D E D R I F T N L Y G R H D W R L K    54

      GTGTCCGTCTGTCTGGAT......
       V S V C L D
 241  GGTGCCCAGAGTCGAGGTGACTGGTACAAGACGAAGGAGATTCTGCTGAAGGGGCCTGACTGGATCCTGGGTGAGGTCAAGACGTCGGGCTTGCGGGGCCGTGGAGGTGCTGGCTTCCCC   360
  55  G A Q S R G D W Y K T K E I L L K G P D W I L G E V K T S G L R G R G G A G F P    94

 361  ACTGGCCTTAAATGGAGCTTCATGAATAAGCCCTCAGATGGCAGGCCCAAGTATCTGGTGGTGAACGCGGATGAGGGGGAGCCAGGCACGTGCAAGGACCGAGAGATCATCCGCCACGAC   480
  95  T G L K W S F M N K P S D G R P K Y L V V N A D E G E P G T C K D R E I I R H D    134

 481  CCCCACAAGCTGGTGGAAGGCTGCCTAGTAGGGGGCCGGGCCATGGGCGCTCGTGCCGCCTACATCTACATCCGCGGGGAGTTCTACAACGAGGCCTCCAATCTGCAGGTAGCCATCCGA   600
 135  P H K L V E G C L V G G R A M G A R A A Y I Y I R G E F Y N E A S N L Q V A I R    174

 601  GAGGCCTACGAGGCTGGTCTGATTGGCAAGAATGCCTGCGGCTCCGGCTACGATTTCGACGTGTTTGTGGTGCGTGGGGCCGGGGCCTACATCTGCGGGGAGGAGACGGCGCTCATCGAG   720
 175  E A Y E A G L I G K N A C G S G Y D F D V F V V R G A G A Y I C G E E T A L I E    214

 721  TCCATCGAGGGCAAACAGGGCAAGCCCCGCCTGAAGCCGCCTTTCCCTGCAGACGTGGGAGTGTTTGGCTGCCCTACAACTGTGGCCAATGTGGAGACTGTGGCCGTGTCCCCTACCATC   840
 215  S I E G K Q G K P R L K P P F P A D V G V F G C P T T V A N V E T V A V S P T I    254

 841  TGCCGCCGTGGGGGTGCCTGGTTTGCCAGCTTCGGCCGAGAGCGCAACTCAGGCACCAAACTGTTCAACATCTCTGGCCACGTCAACAACCCCTGCACCGTGGAGGAGGAGATGTCTGTG   960
 255  C R R G G A W F A S F G R E R N S G T K L F N I S G H V N N P C T V E E E M S V    294

 961  CCGCTGAAGGAGCTGATTGAAAAGCACGCCGGGGGTGTCACGGGTGGCTGGGACAACCTCCTTGCTGTGATCCCTGGCGGCTCGTCCACCCCGCTGATCCCCAAGTCCGTGTGTGAGACG  1080
 295  P L K E L I E K H A G G V T G G W D N L L A V I P G G S S T P L I P K S V C E T    334

1081  GTGCTGATGGACTTCGACGCGCTGATCCAGGCGCAGACGGGGCTGGGCACGGCTGCCGTGATCGTCATGGACCGCTCGACGGACATTGTGAAAGCCATCGCCCGCCTCATCGAGTTCTAC  1200
 335  V L M D F D A L I Q A Q T G L G T A A V I V M D R S T D I V K A I A R L I E F Y    374

1201  AAGCATGAGAGCTGTGGCCAGTGTACCCCGTGCCGCGAGGGTGTGGACTGGATGAACAAGGTGATGGCCCGCTTTGTGAGGGGAGATGCCCGGCCGGCAGAGATCGACTCTCTGTGGGAG  1320
 375  K H E S C G Q C T P C R E G V D W M N K V M A R F V R G D A R P A E I D S L W E    414

1321  ATCAGCAAGCAGATCGAGGGCCACACCATCTGCGCCCTGGGCGATGGGGCTGCGTGGCCCGTGCAGGGCCTGATCCGACACTTCCGGCCGGAGCTCGAGGAGCGGATGCAGCAGTTTGCC  1440
 415  I S K Q I E G H T I C A L G D G A A W P V Q G L I R H F R P E L E E R M Q Q F A    454

                                                                                  pBND51-1    .....ACTC-.....

1441  CAGCAGCACCAGGCCAGGCAAGCAGCTTTCTGAGCCCACGGGGCTGGTGTCCGTCTGTCTGGATGGCTGGACAATAAAGCTAGCGATACCC----AAAAAAAAAAAAAAAAAAAACCCG  1555
 455  Q Q H Q A R Q A A F *    464
```

FIG. 2. Nucleotide sequence and derived amino acid sequence of the cDNAs encoding the 51-kDa subunit of NADH dehydrogenase. The complete sequence derived from the insert of pBND51-2 is shown, and the divergent sequences present in the insert of pBND51-1 are indicated. The amino acid sequence is numbered starting from the methionine residue at position 27 (arrow). The putative initiator codon ATG, the terminator codon TGA, and the poly(A)-addition signal are underlined in the nucleotide sequence; the directly determined peptide sequences identical to the cDNA-derived protein sequences are underlined in the amino acid sequence. Cysteine residues forming the putative binding site for the iron-sulfur cluster are circled. See text for details.

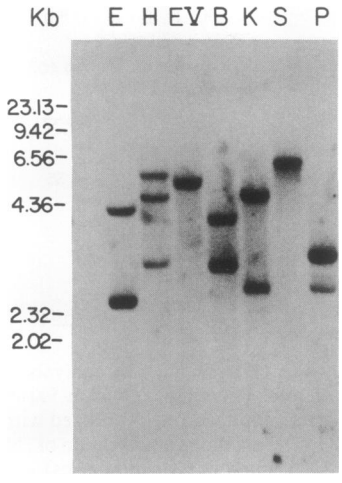

FIG. 4. Southern blot analysis of human genomic DNA. Samples (10 μg) of HeLa cell DNA were digested with one of seven restriction enzymes, blotted, and hybridized with the ^{32}P-labeled insert of pBND51-2. The migration and size (Kb, kilobases) of the *Hin*dIII restriction fragments of λ DNA are indicated at the left of the autoradiogram. Lanes: E, *Eco*RI; H, *Hin*dIII; EV, *Eco*RV; B, *Bgl* II; K, *Kpn* I; S, *Ssp* I; P, *Pvu* II.

(Fig. 4). These results pointed to the existence of a single gene for the 51-kDa subunit in the human genome.

DISCUSSION

The protein encoded in the reading frame identified in the pBND51 clones contains all the partial protein sequences determined for 11 internal tryptic peptides of the isolated 51-kDa subunit, and it is therefore very probably the 51-kDa subunit itself or its precursor. The single amino acid divergences at the last position of peptide B and at the first position of peptide I are probably explained by ambiguity in amino acid sequencing. The divergence at the last three positions of peptide C cannot be easily accounted for by DNA or protein sequencing mistakes and is therefore likely to reflect a true gene polymorphism. The whole reading frame contained in the pBND51 clones encompasses 490 amino acids. However, it is likely that the methionine at the 27th position is the NH_2-terminal amino acid. Several observations point to this conclusion. The molecular mass of the protein from the first methionine residue, taken as residue 1, to the termination codon after residue 464 is 50,749 daltons. If no cleavable presequence existed in this protein, the estimated mass of the mature protein would be in agreement with that derived from the electrophoretic mobility of the 51-kDa subunit in SDS/polyacrylamide gels (5). The stretch corresponding to the first 20 amino acids after the methionine residue has structural features consistent with its being an import signal sequence (22). In fact, it is relatively rich in positively charged amino acids (arginines) and hydroxylated amino acids and devoid of acidic amino acids or extended hydrophobic stretches. When plotted on a "helical wheel" (23), most of the positive charges are clustered on one side, whereas uncharged and hydrophobic residues are clustered on the opposite side, forming a reasonable amphipathic helix (24). No information has been obtained in the present work as to whether this potential targeting sequence is cleaved to generate the mature protein. However, if cleavage occurred after the phenylalanine residue at position 20, in agreement with the reported preferential occurrence of an arginine residue at position −2 relative to the cleavage site (22), the mature protein would have a molecular mass of 48,599 daltons. Although this mass is lower than expected from the observed electrophoretic mobility of the 51-kDa subunit in SDS/polyacrylamide gels, it should be noted that in a different gel system (urea/SDS/polyacrylamide), this subunit has been found to migrate very similarly to the 49-kDa subunit of the iron-protein fragment of the bovine NADH dehydrogenase (25), which has a molecular mass of 49,175 Da (8). The calculated amino acid composition of the cDNA-encoded protein is in very good agreement with that directly determined for the 51-kDa subunit (5), independently of whether the methionine residue or the serine at position 21 is considered as the NH_2-terminal amino acid of the mature protein (data not shown). Therefore, on the basis of the data available here, the nature of the NH_2 terminus of the 51-kDa subunit remains undecided.

A hydropathy plot of the 51-kDa subunit sequence failed to reveal a hydrophobic stretch of sufficient length to be folded into a membrane-spanning α-helix (Fig. 5). This observation is consistent with the evidence from radiolabeling and crosslinking experiments (26) and from immunochemical experiments (27) that indicates that this subunit is not transmembranous but is only partially embedded in the membrane, being exposed largely on the matrix side of the membrane.

The NADH-binding site of the bovine respiratory-chain NADH dehydrogenase has been localized to the 51-kDa subunit by photoaffinity labeling (28). An examination of the sequence of the pBND51 cDNA-encoded protein (Fig. 2) reveals indeed a cluster of glycine residues with a sequence consistent with the motif GXGXXGXXXG (residues 87–96). This motif has been identified as a highly conserved sequence within a βαβ dinucleotide-binding fold common to the NAD(H)-binding domains of many enzymes (29). The 51-kDa subunit has also been reported to carry one [4Fe-4S] cluster (30). An examination of the 51-kDa cDNA-encoded protein sequence reveals 12 cysteine residues. In particular, there is a cluster of cysteine residues with the sequence CXXCXXC (residues 379–385), a motif that has been shown to occur in [4Fe-4S] ferredoxins (31).

A search of the Protein Identification Resource data base for protein sequences having similarity to the 51-kDa protein revealed a striking 32% amino acid identity of this protein to a major portion of the α subunit of the soluble NAD^+-reducing hydrogenase from *A. eutrophus* H16 (10) (Fig. 6). This enzyme, which produces energy from hydrogen oxidation, consists of four subunits, of which the largest (α subunit), of 66.8 kDa, contains an NAD^+-binding site, FMN, and one [2Fe-2S] cluster (10). A sequence comparison of this subunit with the bovine 51-kDa subunit shows several stretches of high similarity. Among these, a stretch encompassing 17 amino acids with 88% sequence identity (residues 83–99; Fig. 6) corresponds to the NADH-binding domain of the 51-kDa subunit that was mentioned above. Another stretch, encompassing 31 amino acids (residues 200–230) with 81% sequence identity, two of the six amino acid substitutions being conservative (Fig. 6), contains a group of

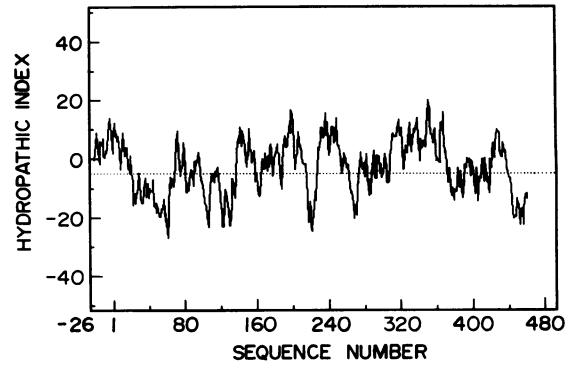

FIG. 5. Hydropathy plot of the 51-kDa subunit of the bovine NADH dehydrogenase (from residue −27 to residue 464). A window of 11 residues was used.

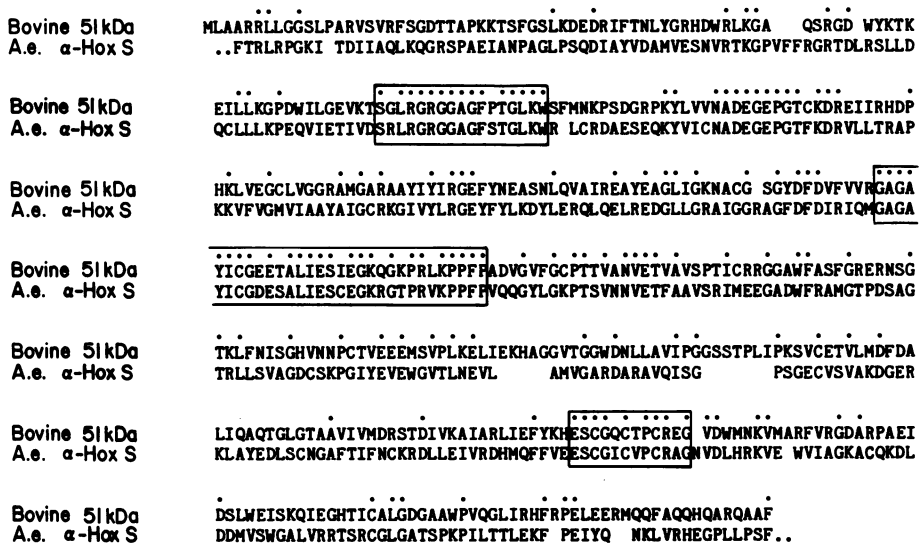

FIG. 6. Sequence comparison between the 51-kDa subunit of the bovine NADH dehydrogenase and the α subunit of the *A. eutrophus* soluble NAD^+-reducing hydrogenase (A.e. α-Hox S). The sequences corresponding to the putative $NADH/NAD^+$-binding domain, the possible FMN-binding site, and the putative binding site for an iron-sulfur cluster are boxed. Identical residues are indicated by dots above the letters. The first 132 residues and the last 21 residues of the α subunit are not shown.

conserved glycine residues and may represent the FMN-binding site (32). Still another stretch, of 12 amino acids (residues 377–388) with 70% sequence identity, contains the above-mentioned CXXCXXC sequence, which may be the binding site for an iron-sulfur cluster. The striking sequence similarity between the 51-kDa subunit of the bovine NADH dehydrogenase and the α subunit of *A. eutrophus* soluble NAD^+-reducing hydrogenase strongly suggests that the two proteins had a common ancestor. It has recently been shown that the 50-kDa subunit of the *Paracoccus denitrificans* NADH dehydrogenase, which is the NADH-binding subunit of the complex, crossreacts with antibodies raised against the 51-kDa subunit of the bovine mitochondrial enzyme and has similar amino acid composition to the latter subunit (33).

The fact that 51-kDa cDNAs identical in nucleotide sequence were isolated from the bovine heart and bovine brain libraries and the observation that apparently a single mRNA species for this protein is expressed in the four tissues examined are consistent with the possibility that there is a single gene for the 51-kDa subunit in the bovine genome. The simple patterns observed in the Southern blot analysis of genomic human DNA are in keeping with this possibility. A wide variation in the steady-state levels of 51-kDa mRNA, which parallels that reported for the 24-kDa subunit of the flavoprotein fragment of the bovine NADH dehydrogenase (7), has been observed in different bovine tissues.

We are greatly indebted to C. I. Ragan for providing us with a preparation of complex I, to R. Miake-Lye and M. Simon for the λgt10 bovine brain library, to Anne Chomyn for gifts of RNA, and to L. Simpson (University of California, Los Angeles) for the hydropathy plot. We also want to express our appreciation to Anne Chomyn for her help in the early phase of the work and useful discussions and to Russell Doolittle (University of California, San Diego) for valuable comments on the manuscript. The technical assistance of Ms. S. L. Lai, Ms. A. Drew, Ms. L. Tefo, and T. Bures is gratefully acknowledged. These investigations were supported by National Institutes of Health Grant GM11726 to G.A., Medical Research Council of Canada Grant MT-10868 and scholarship to R.A., and a Procter & Gamble fellowship to S.P.

1. Ragan, C. I. (1987) *Curr. Top. Bioenerg.* **15**, 1–36.
2. Chomyn, A., Mariottini, P., Cleeter, M. W. J., Ragan, C. I., Matsuno-Yagi, A., Hatefi, Y., Doolittle, R. F. & Attardi, G. (1985) *Nature (London)* **314**, 592–597.
3. Chomyn, A., Cleeter, M. W. J., Ragan, C. I., Riley, M., Doolittle, R. F. & Attardi, G. (1986) *Science* **234**, 614–618.
4. Ise, W., Haiker, H. & Weiss, H. (1985) *EMBO J.* **4**, 2075–2080.
5. Galante, Y. M. & Hatefi, Y. (1979) *Arch. Biochem. Biophys.* **192**, 559–568.
6. Pilkington, S. & Walker, J. E. (1989) *Biochemistry* **28**, 3257–3264.
7. Chomyn, A. & Tsai Lai, S. S.-A. (1989) *Curr. Genet.* **16**, 117–125.
8. Fearnley, I. M., Runswick, M. J. & Walker, J. E. (1989) *EMBO J.* **8**, 665–672.
9. Runswick, M. J., Gennis, R. B., Fearnley, I. M. & Walker, J. E. (1989) *Biochemistry* **28**, 9452–9459.
10. Tran-Betcke, A., Warnecke, U., Böcker, C., Zaborosch, C. & Friedrich, B. (1990) *J. Bacteriol.* **172**, 2920–2929.
11. Hatefi, Y., Haavik, A. G. & Griffith, D. E. (1962) *J. Biol. Chem.* **237**, 1676–1680.
12. Hatefi, Y. & Stempel, K. E. (1967) *J. Biol. Chem.* **244**, 2350–2357.
13. Aebersold, R., Leavitt, J., Saavedra, R. A., Hood, L. E. & Kent, S. B. H. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6970–6974.
14. Aebersold, R. (1989) in *A Practical Guide to Peptide and Protein Purification for Microsequencing*, ed. Matsudaira, P. (Academic, New York), pp. 73–85.
15. Wallace, R. B., Johnson, M. J., Hirose, T., Miyake, T., Kawashima, E. H. & Itakura, K. (1981) *Nucleic Acids Res.* **9**, 879–894.
16. Ohtsuka, E., Matsuki, T., Ikehara, M., Yoosuke, T. & Matsubara, K. (1985) *J. Biol. Chem.* **260**, 2605–2608.
17. Davis, R. W., Botstein, D. & Roth, J. R. (1980) *Advanced Bacterial Genetics* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
18. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
19. Hattori, M. & Sakaki, Y. (1986) *Anal. Biochem.* **152**, 232–238.
20. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
21. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214.
22. Hartl, F.-U., Pfanner, N., Nicholson, D. W. & Neupert, W. (1989) *Biochim. Biophys. Acta* **988**, 1–45.
23. Schiffer, M. & Edmundson, A. B. (1967) *Biophys. J.* **7**, 121–135.
24. Roise, D., Horvath, S. J., Tomich, J. M., Richards, J. H. & Schatz, G. (1986) *EMBO J.* **5**, 1327–1334.
25. Chomyn, A., Patel, S. D., Cleeter, M. W. J., Ragan, C. I. & Attardi, G. (1988) *J. Biol. Chem.* **263**, 16395–16400.
26. Patel, S. D., Cleeter, M. W. J. & Ragan, C. I. (1988) *Biochem. J.* **256**, 529–535.
27. Han, A.-L., Yagi, T. & Hatefi, Y. (1988) *Arch. Biochem. Biophys.* **267**, 490–496.
28. Chen, S. & Guillory, R. J. (1981) *J. Biol. Chem.* **256**, 8318–8323.
29. Scrutton, N. S., Berry, A. & Perham, R. N. (1990) *Nature (London)* **343**, 38–43.
30. Ohnishi, T., Ragan, C. I. & Hatefi, Y. (1985) *J. Biol. Chem.* **260**, 2782–2788.
31. Yasunobu, K. & Tanaka, M. (1980) *Methods Enzymol.* **69**, 228–238.
32. Rossmann, M. G., Moras, D. & Olsen, K. W. (1974) *Nature (London)* **250**, 194–199.
33. Yagi, T. & Dinh, T. M. (1990) *Biochemistry* **29**, 5515–5520.