JOURNAL OF BACTERIOLOGY, May 1989, p. 2894–2899
0021-9193/89/052894-06$02.00/0
Copyright © 1989, American Society for Microbiology
Vol. 171, No. 5

Analysis and Comparison of Nucleotide Sequences Encoding the Genes for [NiFe] and [NiFeSe] Hydrogenases from *Desulfovibrio gigas* and *Desulfovibrio baculatus*

GERRIT VOORDOUW,[1*] NANDA K. MENON,[2] JEAN LEGALL,[2] EUI-SUNG CHOI,[2] HARRY D. PECK, JR.,[2] AND ALAN E. PRZYBYLA[2]

Department of Biological Sciences, The University of Calgary, Calgary, Alberta T2N 1N4, Canada,[1] and Department of Biochemistry, School of Chemical Sciences, University of Georgia, Athens, Georgia 30602[2]

Received 2 December 1988/Accepted 5 February 1989

The nucleotide sequences encoding the [NiFe] hydrogenase from *Desulfovibrio gigas* and the [NiFeSe] hydrogenase from *Desulfovibrio baculatus* (N. K. Menon, H. D. Peck, Jr., J. LeGall, and A. E. Przybyla, J. Bacteriol. 169:5401–5407, 1987; C. Li, H. D. Peck, Jr., J. LeGall, and A. E. Przybyla, DNA 6:539–551, 1987) were analyzed by the codon usage method of Staden and McLachlan. The reported reading frames were found to contain regions of low codon probability which are matched by more probable sequences in other frames. Renewed nucleotide sequencing showed the probable frames to be correct. The corrected sequences of the two small and large subunits share a significant degree of sequence homology. The small subunit, which contains 10 conserved cysteine residues, is likely to coordinate at least 2 iron-sulfur clusters, while the finding of a selenocysteine codon (TGA) near the 3′ end of the [NiFeSe] large-subunit gene matched by a regular cysteine codon (TGC) in the [NiFe] large-subunit gene indicates the presence of some of the ligands to the active-site nickel in the large subunit.

Our knowledge of the structure and function of hydrogenases is increasing rapidly because of work on the molecular biology of the genes encoding these proteins. These investigations have provided the primary structure of the [Fe] hydrogenase from *Desulfovibrio vulgaris* subsp. *vulgaris* Hildenborough (12, 23, 24) and more recently the primary structures of the [NiFeSe] hydrogenase (9) and the [NiFe] hydrogenase (7) from *Desulfovibrio baculatus* and *Desulfovibrio gigas*, respectively. It has become evident from these reports that these three hydrogenases are representatives of three distinct classes encoded by separate genes. All three types of hydrogenase can be present in the same organism, as in the case of *D. vulgaris* (8).

The genetic organization of the hydrogenase operon was found to be the same for the [NiFe] and [NiFeSe] hydrogenases, with the gene for the small subunit (molecular masses, 26 and 29 kilodaltons, respectively) preceding that for the large subunit (molecular masses, 62 and 56 kilodaltons). In contrast, the gene for the large subunit (46 kilodaltons) precedes that for the small subunit (13.5 kilodaltons) in the [Fe] hydrogenase operon. While a very limited sequence homology has been reported between the [NiFe] and [NiFeSe] hydrogenases (7), there is no significant homology between either of these and [Fe] hydrogenase. A more elaborate analysis of the sequence data presented below indicates that the [NiFe] and [NiFeSe] hydrogenases are much more homologous than was previously thought.

Homology comparisons were carried out with the program DIAGON, as described by Staden (16). In addition, the published nucleotide sequences (7, 9, 23) were analyzed by the codon usage method of Staden and McLachlan (17) in order to define the most probable coding regions (e.g., see Fig. 1). Areas in the nucleotide sequences for the [NiFeSe] and [NiFe] hydrogenases where these most probable regions shifted frame were subjected to renewed dideoxy sequencing or gel reading on both strands by methods described before (9). The amino acid sequences of the small and large subunits derived from the revised nucleotide sequences are reported and compared below.

No homology could be detected when the nucleotide sequence of the [Fe] hydrogenase genes (23) was compared with that of either the [NiFe] hydrogenase genes (7) or the [NiFeSe] hydrogenase genes (9) by the DIAGON program. However, a faint but distinct diagonal could be observed if the gene sequences for the latter two hydrogenases were compared (not shown), suggesting a more extensive homology between [NiFeSe] and [NiFe] hydrogenases than expected (7). Analysis of the published nucleotide sequences by the codon usage method was undertaken next. The codon usage table of the gene for the large subunit of [Fe] hydrogenase from *D. vulgaris* (63% G+C) was used as the standard in initial calculations, which indicated that the gene for the large subunit of *D. baculatus* (57% G+C) and both genes for the small and large subunits of *D. gigas* (63% G+C) contained regions of low coding probability. These were matched by regions of high coding probability in other frames, indicating the possibility of frameshifts due to errors in the original nucleotide sequence data. Comparable results were obtained when the codon usage table for the small-subunit gene of the [NiFeSe] hydrogenase from *D. baculatus*, which displayed a high coding probability throughout its reading frame in these initial calculations, was used as the standard. Use of this table is preferred for calculations on the *D. baculatus* genes in view of the different G+C contents indicated above, and the results of the calculations on the previously published sequences (7, 9) are shown in Fig. 1A through F.

The large-subunit gene of *D. baculatus* hydrogenase, which contained three regions of codon improbability (Fig. 1C, I through III), was next resequenced, and this confirmed the correctness of the codon probability analysis. An erratum with the complete, corrected sequence has been pub-

* Corresponding author.

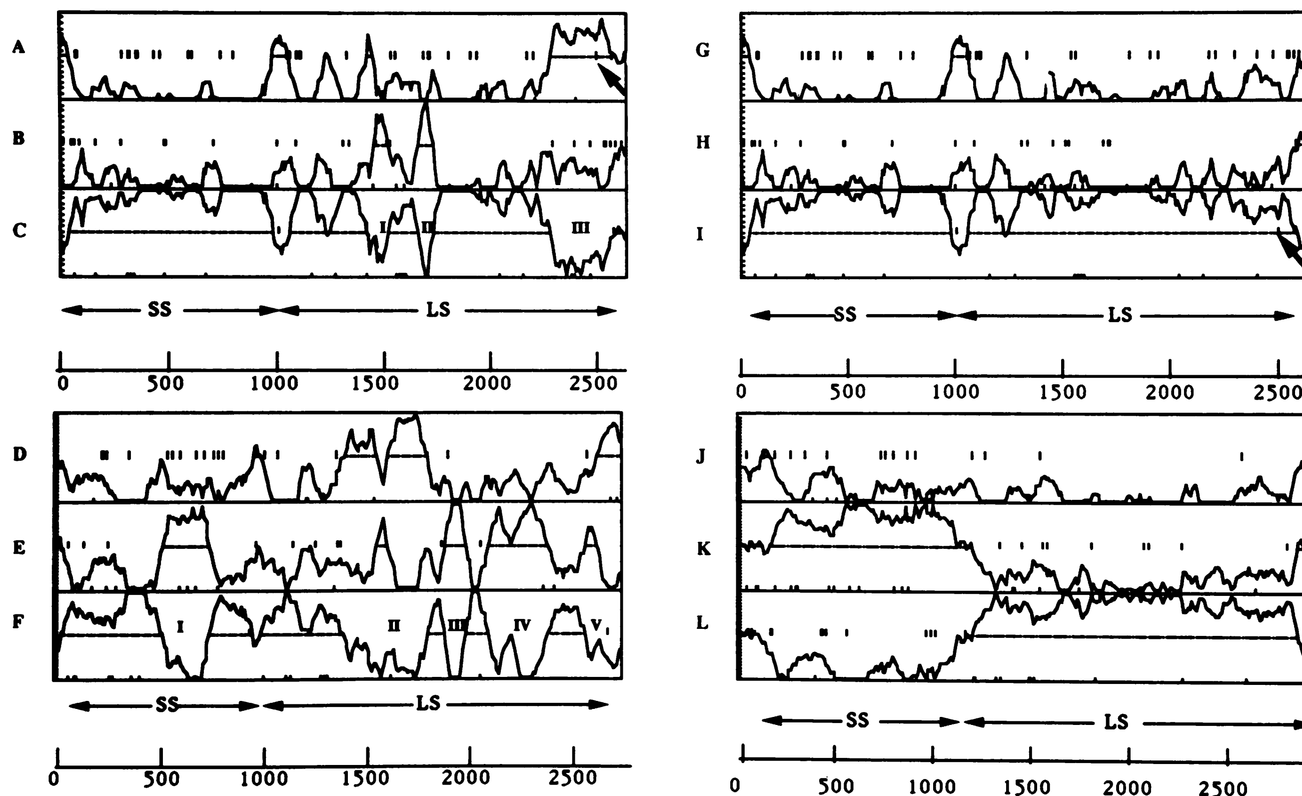

FIG. 1. Codon probability profiles for the nucleotide sequences encoding the [NiFeSe] hydrogenase of *D. baculatus* and the [NiFe] hydrogenase of *D. gigas*. The probability that a stretch of sequence (25 bases) is coding is plotted for the three reading frames of every sequence. Methionine (ATG) codons are indicated on the base line of each frame, while stop (TAA, TAG, and TGA) codons are indicated at half level in each frame. The proposed positions of the small-subunit (SS) and large-subunit (LS) genes and a scale (in base pairs) are shown for every sequence. (A through C) Nucleotide sequences of [NiFeSe] hydrogenase genes as published by Menon et al. (9). The published coding regions for both small-subunit and large-subunit genes (with three improbable regions, I through III) are in frame C. A TGA stop codon, now thought to encode selenocysteine, is indicated (↖) in frame A. (D through F) Nucleotide sequences of [NiFe] hydrogenase genes as published by Li et al. (7). The published coding regions for both small-subunit genes (with improbable region I) and large-subunit genes (with improbable regions II through V) are in frame F. (G through I) Corrected nucleotide sequence of the [NiFeSe] hydrogenase genes (9). The coding regions for small-subunit and large-subunit genes are in frame I. The selenocysteine codon near the 3′ end of the large subunit gene is indicated (↖). (J through L) Corrected nucleotide sequence of the [NiFe] hydrogenase genes (Fig. 2). The coding regions for small-subunit and large-subunit genes translated in Fig. 2 are present in frames K and L, respectively.

lished (9). The new sequence has a high coding probability throughout the large subunit gene (Fig. 1I). An unusual feature of the revised sequence is a TGA stop codon in the large subunit gene which was uncovered by codon probability calculations of the originally published sequence (Fig. 1A, III), as will be discussed in more detail below.

Because of sequencing errors in the published nucleotide sequence of the [NiFe] hydrogenase from *D. gigas*, the codon usage method indicated a large number of frameshifts. These have been labeled I through V in Fig. 1F. A single region of low coding probability (region I) was found in the published sequence for the small subunit gene and was matched by a region of high coding probability in frame E. Even before resequencing of the gene was started, it could be concluded that frame E contained the correct sequence of this region of the small subunit, since it was 40% identical to the corresponding sequence of the small subunit of the [NiFeSe] hydrogenase which was found to be devoid of errors (Fig. 1C). The large subunit gene contains several regions of low coding probability (II through V), and it was clear that the amino acid sequence derived from this gene had to be largely wrong. The entire [NiFe] hydrogenase operon of *D. gigas* was resequenced, and the corrected sequence is shown in Fig. 2. The small- and large-subunit genes have been translated into protein, and regions I through V, corresponding to new protein sequences, found to be correctly predicted by the codon probability method, have been indicated by overlining. These regions amount to 35% of the small-subunit and 42% of the large-subunit amino acid sequences. The following criteria indicate the correctness of the nucleotide sequence in Fig. 2. (i) Both the small and large subunits are encoded by highly probable frames (Fig. 1K and L, respectively). (ii) The amino acid compositions of the separate small and large subunits and of the hydrogenase (which is a 1:1 complex of these two subunits) derived from the revised nucleotide sequence are in good agreement with data (5, 11) determined for the protein (not shown). (iii) The amino acid sequences derived for the small and large subunits of the [NiFe] hydrogenase from *D. gigas* show 70% identity with those obtained from the sequence of the gene for [NiFe] hydrogenase from *D. vulgaris* (A. E. Przybyla et al., unpublished data). (iv) These sequences show lower but still significant and very interesting homology to the sequences of the corresponding subunits from the [NiFeSe] hydrogenase from *D. baculatus*.

The amino acid sequence for the small subunit of the

```
ATGCATTATCAATGCCTTCAATGCACTTCAAGCTGCCTGGACCTGCCCGATGACGACGCGCAGCGCTTGGTATAGAGTGTGGAACTCCAGCCGTTACTGTCGCATGTTCCCAATGCGCGC
       10        20        30        40        50        60        70        80        90       100       110       120

                                                                         rbs     M  K  C  Y  I  G  R  G  K  D  Q  V  E  E
AAGCAGGCCCTGTTTGATCCTAACACTGGGTGTGTGTGTCCGAACGGACCTTAGAACTAGAGAGGACGGAGGGTACGAAATGAAGTGTTACATTGGACGTGGCAAGGACCAGGTCGAGGA
      130       140       150       160       170       180       190       200       210       220       230       240

 R  L  E  R  R  G  V  S  R  R  D  F  M  K  F  C  T  A  V  A  V  A  M  G  M  G  P  A  F  A  P  K  V  A  E  A  L  T  A  K
ACGCCTGGAGCGCCGGGGCGTCTCGCCGCGACTTCATGAAGTTCTGCACGGCCGTGGCCGTCGCCATGGGTATGGGTCCCGCCTTTGCCCCCAAGGTTGCCGAGGCGTTGACCGCCAA
      250       260       270       280       290       300       310       320       330       340       350       360

 K  R  P  S  V  V  Y  L  H  N  A  E  C  T  G  C  S  E  S  L  L  R  T  V  D  P  Y  V  D  E  L  I  L  D  V  I  S  M  D  Y
GAAGCGCCCCAGCGTGGTGTATCTGCACAATGCAGAATGCACCGGCTGCAGCGAATCTCTGCTGCGGACCGTTGACCCGTACGTTGACGAACTCATTCTTGATGTGATTTCCATGGATTA
      370       380       390       400       410       420       430       440       450       460       470       480

 H  E  T  L  M  A  G  A  G  H  A  V  E  E  A  L  H  E  A  I  K  G  D  F  V  C  V  I  E  G  G  I  P  M  G  D  G  G  Y  W
TCACGAAACCCTCATGGCTGGCGCCGGTCATGCCGTGGAAGAAGCCCTGCACGAAGCCATCAAGGGCGATTTCGTCTGCGTCATTGAAGGCGGCATTCCCATGGGCGACGGCGGCTACTG
      490       500       510       520       530       540       550       560       570       580       590       600

 G  K  V  G  R  R  N  M  Y  D  I  C  A  E  V  A  P  K  A  K  A  V  I  A  I  G  T  C  A  T  Y  G  G  V  Q  A  A  K  P  N
GGGCAAGGTCGGCCGCCGCAACATGTACGACATCTGCGCCGAAGTGGCTCCCAAGGCCAAGGCAGTCATCGCCATCGGCACCTGCGCCACCTACGGTGGCGTGCAGGCTGCCAAGCCCAA
      610       620       630       640       650       660       670       680       690       700       710       720
                                                              I
 P  T  G  T  V  G  V  N  E  A  L  G  K  L  G  V  K  A  I  N  I  A  G  C  P  P  N  P  M  N  F  V  G  T  V  V  H  L  L  T
CCCCACGGGCACCGTGGGTGTGAACGAAGCCCTGGGCAAACTGGGCGTGAAGGCTATCAACATCGCCGGCTGCCCGCCGAACCCCATGAACTTTGTGGGCACCGTGGTGCATCTGCTCAC
      730       740       750       760       770       780       790       800       810       820       830       840

 K  G  M  P  E  L  D  K  Q  G  R  P  V  M  F  F  G  E  T  V  H  D  N  C  P  R  L  K  H  F  E  A  G  E  F  A  T  S  F  G
CAAGGGCATGCCCGAGCTGGACAAGCAAGGCCGCCCGGTGATGTTCTTCGGCGAAACCGTGCACGACAACTGCCCCCGTCTGAAGCACTTCGAAGCGGGCGAGTTTGCCACCTCCTTCGG
      850       860       870       880       890       900       910       920       930       940       950       960

 S  P  E  A  K  K  G  Y  C  L  Y  E  L  G  C  K  G  P  D  T  Y  N  N  C  P  K  Q  L  F  N  Q  V  N  W  P  V  Q  A  G  H
CTCCCCTGAAGCCAAGAAGGGCTACTGCCTCTATGAGCTGGGTTGCAAGGGTCCTGATACGTACAACAACTGCCCCAAGCAGCTCTTCAACCAGGTCAACTGGCCGGTCCAGGCCGGGCA
      970       980       990      1000      1010      1020      1030      1040      1050      1060      1070      1080

 P  C  I  A  C  S  E  P  N  F  W  D  L  Y  S  P  F  Y  S  A  *                                                   rbs
CCCCTGCATCGCCTGCAGCGAGCCCAATTTCTGGGATCTCTATTCGCCGTTCTACAGCGCCTAGACGACCTCGCCAAGCTCATCGAGCACTAGCCAGGAACTTCGTTTCGCACAGGAGAA
     1090      1100      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200

       M  S  E  M  Q  G  N  K  I  V  V  D  P  I  T  R  I  E  G  H  L  R  I  E  V  E  V  E  G  G  K  I  K  N  A  W  S  M
GACCCGATGTCTGAAATGCAAGGCAACAAGATCGTCGTCGATCCCATCACCCGGATCGAGGGCCATCTTCGCATTGAAGTGGAAGTGGAAGGCGGCAAGATCAAGAACGCCTGGAGCATG
     1210      1220      1230      1240      1250      1260      1270      1280      1290      1300      1310      1320

 S  T  L  F  R  G  L  E  M  I  L  K  G  R  D  P  R  D  A  Q  H  F  T  Q  R  A  C  G  V  C  T  Y  V  H  A  L  A  S  V  R
TCCACGCTGTTCCGCGGCCTGGAAATGATCCTCAAGGGCCGCGACCCGCGCGACGCCCAGCACTTCACCCAGCGCGCCTGCGGCGTGTGCACCTATGTGCATGCCCTGGCCTCCGTCCGC
     1330      1340      1350      1360      1370      1380      1390      1400      1410      1420      1430      1440

 A  V  D  N  C  V  G  V  K  I  P  E  N  A  T  L  M  R  N  L  T  M  G  A  Q  Y  M  H  D  H  L  V  H  F  Y  H  L  H  A  L
GCCGTGGACAACTGCGTGGGCGTGAAGATTCCCGAAAACGCCACTCTCATGCGCAACCTCACCATGGGCGCGCAATACATGCACGACCACTTGGTGCACTTCTACCACCTGCATGCCCTT
     1450      1460      1470      1480      1490      1500      1510      1520      1530      1540      1550      1560

 D  W  V  N  V  A  N  A  L  N  A  D  P  A  K  A  A  R  L  A  N  D  L  S  P  R  K  T  T  T  E  S  L  K  A  V  Q  A  K  V
GACTGGGTGAACGTGGCCAACGCCCTGAACGCCGATCCGGCCAAGGCTGCCCGCCTGGCCAACGATCTCTCCCCCCGCAAAACCACCACGGAAAGCCTCAAGGCCGTGCAGGCCAAGGTC
     1570      1580      1590      1600      1610      1620      1630      1640      1650      1660      1670      1680

 K  A  L  V  E  S  G  Q  L  G  I  F  T  N  A  Y  F  L  G  G  H  P  A  Y  V  L  P  A  E  V  D  L  I  A  T  A  H  Y  L  E
AAGGCCCTGGTGGAAAGCGGCCAGCTGGGCATCTTCACCAACGCCTACTTCCTGGGCGGCCACCCCGCCTATGTGCTGCCGGCGGAAGTGGACCTCATCGCCACCGCCCACTACCTGGAA
     1690      1700      1710      1720      1730      1740      1750      1760      1770      1780      1790      1800
                                                             II
 A  L  R  V  Q  V  K  A  A  R  A  M  A  I  F  G  A  K  N  P  H  T  Q  F  T  V  V  G  G  C  T  N  Y  D  S  L  R  P  E  R
GCCCTGCGCGTGCAGGTGAAGGCCGCCCGGGCCATGGCCATCTTCGGCGCCAAGAACCCCCACACCCAGTTCACCGTGGTGGGTGGTTGCACCAACTACGACTCCCTGCGTCCCGAACGC
     1810      1820      1830      1840      1850      1860      1870      1880      1890      1900      1910      1920

 I  A  E  F  R  K  L  Y  K  E  V  R  E  F  I  E  Q  V  Y  I  T  D  L  L  A  V  A  G  F  Y  K  N  W  A  G  I  G  K  T  S
ATCGCCGAATTCCGCAAGCTGTACAAGGAAGTCCGCGAGTTCATCGAGCAGGTGTACATCACGGACCTGCTGGCCGTGGCCGGGTTCTACAAGAACTGGGCCGGCATCGGCAAAACGTCC
     1930      1940      1950      1960      1970      1980      1990      2000      2010      2020      2030      2040
                                                                                                               III
 N  F  L  T  C  G  E  F  P  T  D  E  Y  D  L  N  S  R  Y  T  P  Q  G  V  I  W  G  N  D  L  S  K  V  D  D  F  N  P  D  L
AACTTCCTCACCTGCGGCGAATTCCCCACCGACGAATACGATCTGAACTCCCGCTACACCCCGCAGGGCGTGATCTGGGGCAACGACCTCAGCAAGGTGGACGACTTCAACCCCGACCTC
     2050      2060      2070      2080      2090      2100      2110      2120      2130      2140      2150      2160

 I  E  E  H  V  K  Y  S  W  Y  E  G  A  D  A  H  H  P  Y  K  G  V  T  K  P  K  W  T  E  F  H  G  E  D  R  Y  S  W  M  K
ATCGAGGAGCACGTCAAGTACTCCTGGTACGAAGGCGCGGACGCCCACCATCCGTACAAGGGCGTCACCAAGCCCAAGTGGACCGAATTCCACGGCGAAGATCGCTACTCCTGGATGAAG
     2170      2180      2190      2200      2210      2220      2230      2240      2250      2260      2270      2280

 A  P  R  Y  K  G  E  A  F  E  V  G  P  L  A  S  V  L  V  A  Y  A  K  K  H  E  P  T  V  K  A  V  D  L  V  L  K  T  L  G
GCCCCCCGGTACAAGGGCGAAGCCTTCGAAGTGGGGCCGTTGGCCTCCGTGCTCGTGGCTTACGCCAAGAAGCACGAACCCACCGTCAAGGCTGTGGATCTTGTCCTCAAGACCCTGGGC
     2290      2300      2310      2320      2330      2340      2350      2360      2370      2380      2390      2400

 V  G  P  E  A  L  F  S  T  L  G  R  T  A  A  R  G  I  Q  C  L  T  A  A  Q  E  V  E  V  W  L  D  K  L  E  A  N  V  K  A
GTGGGCCCCGAAGCCCTGTTCTCCACCCTGGGCCGCACCGCCGCCCGCGGCATTCAGTGCCTCACCGCCGCTCAGGAAGTGGAAGTGTGGCTGGACAAGCTGGAAGCCAACGTCAAGGCT
     2410      2420      2430      2440      2450      2460      2470      2480      2490      2500      2510      2520
                                                       IV
 G  K  D  D  L  Y  T  D  W  Q  Y  P  T  E  S  Q  G  V  G  F  V  N  A  P  R  G  M  L  S  H  W  I  V  Q  R  G  G  K  I  E
GGCAAGGACGATCTCTACACCGATTGGCAGTACCCCACCGAGTCCCAGGGCGTGGGCTTTGTTAACGCGCCCCGCGGCATGCTCTCCCACTGGATTGTCCAACGTGGCGGCAAGATCGAG
     2530      2540      2550      2560      2570      2580      2590      2600      2610      2620      2630      2640

 N  F  Q  H  V  V  P  S  T  W  N  L  G  P  R  C  A  E  R  K  L  S  A  V  E  Q  A  L  I  G  T  P  I  A  D  P  K  R  P  V
AACTTCCAGCATGTGGTGCCCTCCACCTGGAACCTCGGCCCCCGCTGCGCCGAACGGAAACTCTCCGCTGTGGAACAGGCCCTCATCGGTACGCCCATCGCCGATCCCAAGCGTCCTGTG
     2650      2660      2670      2680      2690      2700      2710      2720      2730      2740      2750      2760
                                                                V
 E  I  L  R  T  V  H  S  Y  D  P  C  I  A  C  G  V  H  V  I  D  P  E  S  N  Q  V  H  K  F  R  I  L  *
GAAATCCTGCGCACCGTGCACTCCTACGACCCCTGCATCGCCTGCGGCGTGCACGTGATCGATCCGGAGTCCAACCAGGTGCACAAGTTCCGCATCCTGTAGTTCCGGACGCCAGACCCA
     2770      2780      2790      2800      2810      2820      2830      2840      2850      2860      2870      2880

ACGCAACAACCAGGAGCCCGGCCTTG
     2890      2900
```

FIG. 2. Corrected nucleotide sequence for the [NiFe] hydrogenase operon from *D. gigas*. The coding regions for the small subunit (nucleotides 200 through 1141) and the large subunit (nucleotides 1207 through 2859) have been translated into protein with the one-letter amino acid code. All cysteine residues have been circled. The mature small subunit (NH_2-terminal sequence LTAK, etc.) is preceded by a complex signal sequence of 50 amino acid residues. Both genes are preceded by a purine-rich sequence that could function in translation initiation (rbs). Those parts of the amino acid sequence that differ from a previously published sequence (7) are overlined. The new sequences, designated I through V, were correctly predicted by the codon probability analysis and correspond to low-codon-probability areas I through V in Fig. 1F.

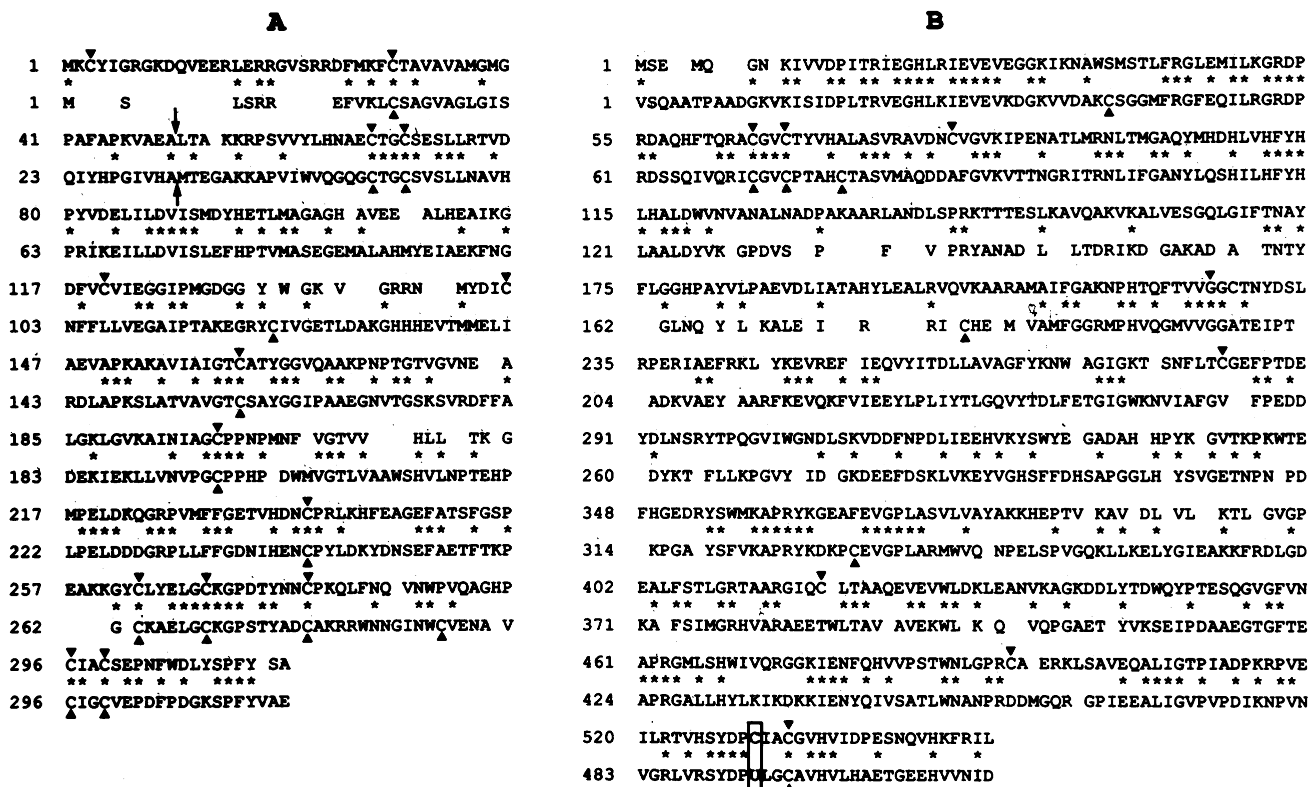

FIG. 3. Comparison of the amino acid sequences of the small (A) and large (B) subunits of the *D. gigas* [NiFe] hydrogenase (top line) and the *D. baculatus* [NiFeSe] hydrogenase (bottom line). The comparison was made with the Beckman Microgenie program. The signal peptide processing site is indicated in both small subunits (↑, ↓). The locations of cysteine residues (▲, ▼) and conserved cysteine residues (⧗) are shown in both subunits. The cysteine-selenocysteine (U) homology at the COOH-terminal end of the large subunit sequence is highlighted by a box.

[NiFeSe] hydrogenase from *D. baculatus* is compared with that of the [NiFe] hydrogenase from *D. gigas* in Fig. 3A, while the two large-subunit sequences are compared in Fig. 3B. Both mature small-subunit sequences are preceded by an NH_2-terminal signal sequence of 32 or 50 residues, respectively. These complex sequences must function in the export of these hydrogenases or in their binding to the membrane. The [NiFe] hydrogenase of *D. gigas* has been firmly established as a periplasmic enzyme. Both large subunits lack an NH_2-terminal signal sequence and do not appear to be processed beyond the removal of the *N*-formyl methionine. The sequences for the mature small subunits share approximately 38% of sequence identity. Both contain 12 cysteine residues, of which 10 are found to be conserved. The amino acid residues flanking these cysteines are also conserved (Fig. 3A). Contrary to expectations, the sequences of the large subunits, which are 34% identical, contain fewer cysteines (seven and nine residues, respectively), of which only three are conserved. In addition, a fourth cysteine codon (TGC) near the 3′ end of the large-subunit gene for [NiFe] hydrogenase is replaced by a TGA codon, thought to encode selenocysteine, in the homologous region of the large-subunit gene for [NiFeSe] hydrogenase (Fig. 4).

The [NiFe] and [NiFeSe] hydrogenases are thus related, and the modest degree of homology between the two sequences allows a number of potentially essential residues to be defined. Both are very different from [Fe] hydrogenase. The mature small subunit of [Fe] hydrogenase lacks cysteine residues, while an arrangement of eight cysteines as in 8Fe-8S ferredoxin has been found at the NH_2 terminus of the large subunit. It has been proposed that these eight cysteine residues coordinate two of the three iron sulfur clusters that are present in [Fe] hydrogenase (23). In contrast, 10 conserved cysteines are found in the small subunit of [NiFe] and [NiFeSe] hydrogenases and are sufficient for the coordination of two 4Fe-4S clusters. The spacing of these cysteines is very different from that found at the NH_2 terminus of the large subunit of [Fe] hydrogenase and is not homologous to 8Fe-8S ferredoxin.

The [NiFe] and [NiFeSe] hydrogenases are different with respect to the metal content and composition of their nonheme iron centers (7, 13). The 35% sequence homology is

(a)	Formate Dehydrogenase (*M. formicicum*)	CAC	TGT	GCA	CGA	CTC	TGC	CAC	GGC	CCA	ACT
(b)	Formate Dehydrogenase (*E. coli*)	TGC	TGC	GCT	CGT	GTC	TGA	CAC	GGC	CCA	TCG
(c)	[NiFeSe]Hydrogenase (*D. baculatus*)	CGC	TCC	TAC	GAC	CCG	TGA	CTG	GGC	TGT	GCC
(d)	[NiFe]Hydrogenase (*D. gigas*)	CAC	TCC	TAC	GAC	CCC	TGC	ATC	GCC	TGC	GGC
(e)	[NiFe]Hydrogenase (*D. vulgaris*)	CAC	TCC	TTC	GAC	CCG	TGC	ATA	GCC	TGT	GGC

FIG. 4. Comparison of homologous nucleotide sequences coding for cysteine-selenocysteine in (a) the formate dehydrogenase from *M. formicicum* (14), (b) the formate dehydrogenase of *E. coli* (26), (c) the [NiFeSe] hydrogenase of *D. baculatus* (9), and (d and e) the [NiFe] hydrogenases of *D. vulgaris* (unpublished data) and *D. gigas* (Fig. 2, nucleotides 2779 through 2808).

insufficient for cross-hybridization on genomic Southern blots (N. Menon, unpublished data), while in Western blots, antibodies directed against the purified proteins fail to cross-react (13). The homology within the [NiFe] class of hydrogenases is much higher and allowed the rapid cloning of the [NiFe] hydrogenase gene from *D. vulgaris* with a probe from *D. gigas*. It appears that the two enzymes are 70% homologous at the amino acid sequence level.

The biochemical role of selenium in the [NiFeSe] hydrogenases is under active investigation. Selenium has been reported to be incorporated into polypeptide chains by a nonspecific mechanism as selenomethionine (15) and by a specific mechanism (18) which involves the modification of a seryl tRNA to a selenocysteinyl tRNA (6) followed by cotranslational insertion of a selenocysteine. Inorganic selenium has been reported in carbon monoxide oxidase (10). Selenium has been shown to serve as a replacement for sulfur in some nonheme iron proteins (1, 22).

It is most relevant that selenium, present as selenocysteine in formate dehydrogenase from *Escherichia coli*, is encoded by a TGA codon (26). Moreover, this TGA codon is replaced by a TGC (cysteine) codon in the formate dehydrogenase from *Methanobacterium formicicum*, which lacks selenium (14). This is analogous to the results of the present study, and the relevant nucleic acid sequences are compared in Fig. 4. The occurrence of selenocysteine in *Methanococcus vannielii* hydrogenase has been demonstrated elsewhere (25), and the presence of a TGA codon for selenocysteine has also been reported in mammalian glutathione peroxidases (3, 19).

Nickel is thought to be involved in the activation of hydrogen by the nickel-containing hydrogenases (4, 20, 21). The higher H_2/HD ratios observed with the [NiFeSe] hydrogenases suggest that selenium modifies the reactivity of this active-site nickel (2). The possible explanation that selenium as selenocysteine serves as a ligand to the nickel has been recently substantiated by the observed broadening of the nickel electron paramagnetic resonance signals of [NiFeSe] hydrogenase enriched with selenium-77 (6a) and by extended X-ray absorption fine-structure spectroscopy studies which indicate that selenium is adjacent (0.246 nm) to the nickel (3a). Together these spectroscopic data demonstrate that selenium, as selenocysteine, serves as a ligand to nickel in [NiFeSe] hydrogenase, while the data presented in Fig. 3B point to the cysteine in the large subunit that is likely to have an equivalent function in [NiFe] hydrogenases.

The statistical analyses were done during a short visit by G.V. to the Laboratory of Molecular Biology, MRC Centre, Cambridge. G.V. acknowledges fruitful discussions with R. Staden, S. G. Mayhew, D. H. Boxer, and G. Yates.

G.V. was supported by the Alberta Heritage Foundation of Medical Research and the National Engineering and Science Research Council of Canada. This research was also supported in part by the U.S. Department of Energy under contract DE-AS09-80 ER 10466 to H.D.P. and by grant DMB 8602789 from the National Science Foundation and Public Health Service grant GM34903 from the National Institutes of Health to J.L., A.E.P., and H.D.P.

LITERATURE CITED

1. **Aunc, P., J. Gaillard, J. Meyer, and J.-M. Moulis.** 1987. Analysis of the high spin states of the $2[4Fe\text{-}4S]^{+}$ ferredoxin from *Clostridium pasteurianum* by Mossbauer spectroscopy. Biochem. J. **242:**525–530.
2. **Berlier, Y., G. D. Fauque, J. LeGall, E. S. Choi, H. D. Peck, Jr., and P. A. Lespinat.** 1987. Inhibition studies of three classes of *Desulfovibrio* hydrogenase: application to the further characterization of the multiple hydrogenases found in *Desulfovibrio vulgaris* (Hildenborough). Biochem. Biophys. Res. Commun. **146:**147–153.
3. **Chambers, I., J. Frampton, P. Goldfarb, N. Affara, W. McBain, and P. R. Harrison.** 1986. The structure of the mouse glutathione peroxidase gene: the selenocysteine in the active site is encoded by the "termination" codon, TGA. EMBO J. **5:** 1221–1227.

3a. **Eidsness, M. K., R. A. Scott, B. Pickrill, D. V. DerVartanian, J. LeGall, I. Moura, and H. D. Peck, Jr.** 1989. Evidence for selenocysteine coordination to the active site nickel in the [NiFeSe] hydrogenase from *Desulfovibrio baculatus*. Proc. Natl. Acad. Sci. USA **86:**147–151.

4. **Fernandez, V. M., E. C. Hatchikian, D. S. Patil, and R. Cammack.** 1986. ESR detectable nickel and iron-sulphur centers in relation to the reversible activation of *Desulfovibrio gigas* hydrogenase. Biochim. Biophys. Acta **883:**145–154.
5. **Hatchikian, E. C., M. Bruschi, and J. LeGall.** 1978. Characterization of the periplasmic hydrogenase from *Desulfovibrio gigas*. Biochem. Biophys. Res. Commun. **82:**451–461.
6. **Leinfelder, W., E. Zehelein, M.-A. Mandrand-Berthelot, and A. Bock.** 1988. Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. Nature (London) **331:**723–725.

6a. **He, S.-H., M. Texeira, J. LeGall, D. S. Patil, D. V. DerVartanian, B. H. Huyn, and H. D. Peck, Jr.** 1989. EPR studies with ^{77}Se enriched [NiFeSe] hydrogenase of *Desulfovibrio baculatus*. Evidence for a selenium ligand to the active-site nickel. J. Biol. Chem. **264:**2678–2682.

7. **Li, C., H. D. Peck, Jr., J. LeGall, and A. E. Przybyla.** 1987. Cloning, characterization and sequencing of the genes encoding the large and small subunits of the periplasmic [NiFe] hydrogenase of *Desulfovibrio gigas*. DNA **6:**539–551.
8. **Lissolo, T., E. S. Choi, J. LeGall, and H. D. Peck, Jr.** 1986. The presence of multiple intrinsic membrane nickel containing hydrogenase in *Desulfovibrio vulgaris* (Hildenborough). Biochem. Biophys. Res. Commun. **139:**701–708.
9. **Menon, N. K., H. D. Peck, Jr., J. LeGall, and A. E. Przybyla.** 1987. Cloning and sequencing of the genes encoding the large and small subunits of the periplasmic (NiFeSe) hydrogenase of *Desulfovibrio baculatus*. J. Bacteriol. **169:**5401–5407. (Erratum, **170:**4429, 1988.)
10. **Meyer, O., and K. V. Rajagopalan.** 1984. Selenite binding to carbon monoxide oxidase from *Pseudomonas carboxydovorans*. Selenium binds covalently to the protein and activates specifically the CO methylene blue reaction. J. Biol. Chem. **259:**5612–5617.
11. **Nivière, V., N. Forget, G. Bovier-Lapierre, J. Bonicel, and C. Hatchikian.** 1988. Isolation, amino acid analysis and N-terminal sequence determination of the two subunits of the nickel-containing hydrogenase of *Desulfovibrio gigas*. Biochimie **70:** 267–271.
12. **Prickril, B. C., M. H. Czechowski, A. E. Przybyla, H. D. Peck, Jr., and J. LeGall.** 1986. Putative signal peptide on the small subunit of the periplasmic hydrogenase from *Desulfovibrio vulgaris*. J. Bacteriol. **167:**722–725.
13. **Prickril, B. C., S.-H. He, C. Li, N. Menon, E. S. Choi, A. E. Przybyla, D. V. DerVartanian, H. D. Peck, Jr., G. Fauque, J. LeGall, M. Texeira, I. Moura, J. J. G. Moura, D. Patil, and B. J. Huyn.** 1987. Identification of three distinct classes of hydrogenase in the genus *Desulfovibrio*. Biochem. Biophys. Res. Commun. **149:**369–377.
14. **Shuber, A. P., E. C. Orr, M. A. Recny, P. F. Schendel, H. D. May, N. L. Schauer, and J. G. Ferry.** 1986. Cloning, expression and nucleotide sequence of the formate dehydrogenase genes from *Methanobacterium formicicum*. J. Biol. Chem. **261:**12942–12947.
15. **Sliwkowski, M. X., and T. C. Stadtman.** 1985. Incorporation and distribution of selenium into thiolase from *Clostridium kluyveri*. J. Biol. Chem. **260:**3140–3144.
16. **Staden, R.** 1982. An interactive graphics program for comparing and aligning nucleic acid or amino acid sequences. Nucleic Acids Res. **10:**2951–2961.
17. **Staden, R., and A. D. McLachlan.** 1982. Codon preference and

its use in identifying protein coding regions in long DNA sequences. Nucleic Acids Res. **10**:141–156.
18. **Stadtman, T. C.** 1987. Specific occurrence of selenium in enzymes and amino acid tRNA's. FASEB J. **1**:375–379.
19. **Sukenaga, Y., K. Ishida, T. Takeda, and K. Takagi.** 1987. cDNA sequence coding for human glutathione peroxidase. Nucleic Acids Res. **15**:7178.
20. **Texeira, M., G. Fauque, I. Moura, P. A. Lespinat, Y. Berlier, B. Prickril, H. D. Peck, Jr., A. V. Xavier, J. LeGall, and J. J. G. Moura.** 1987. Nickel-[iron-sulfur]-selenium containing hydrogenases from *Desulfovibrio baculatus* (DSM 1743). Redox centers and catalytic properties. Eur. J. Biochem. **167**:47–58.
21. **Texeira, M., I. Moura, A. V. Xavier, B. H. Huyn, D. V. DerVartanian, H. D. Peck, Jr., J. LeGall, and J. J. G. Moura.** 1985. Electron paramagnetic resonance studies on the mechanism of activation and the catalytic cycle of the nickel-containing hydrogenase from *Desulfovibrio gigas*. J. Biol. Chem. **260**:8942–8950.
22. **Tsibris, J. C. M., M. J. Namvedt, and I. C. Gonsalus.** 1968. Selenium as an acid labile sulfur replacement in putidaredoxin. Biochem. Biophys. Res. Commun. **30**:323–327.
23. **Voordouw, G., and S. Brenner.** 1985. Nucleotide sequence of the gene encoding the hydrogenase from *Desulfovibrio vulgaris* (Hildenborough). Eur. J. Biochem. **148**:515–520.
24. **Voordouw, G., J. E. Walker, and S. Brenner.** 1985. Cloning of the gene encoding the hydrogenase from *Desulfovibrio vulgaris* (Hildenborough) and determination of the NH_2-terminal sequence. Eur. J. Biochem. **148**:509–514.
25. **Yamazaki, S.** 1982. A selenium-containing hydrogenase from *Methanococcus vannielii*. Identification of the selenium moiety as a selenocysteine residue. J. Biol. Chem. **257**:7926–7929.
26. **Zinoni, F., A. Birkman, T. C. Stadtman, and A. Bock.** 1986. Nucleotide sequence and expression of the selenocysteine containing polypeptide of formate dehydrogenase (formate-hydrogen-lyase linked) from *Escherichia coli*. Proc. Natl. Acad. Sci. USA **83**:4650–4654.