Volume 9 Number 21 1981 Nucleic Acids Research

# Nucleotide sequence of the R.meliloti nitrogenase reductase (nifH) gene

István Török[1] and Ádám Kondorosi[2]

[1]Institute of Biochemistry, and [2]Institute of Genetics, Biological Research Center, Hungarian Academy of Sciences, H-6701 Szeged, P.O. Box 521, Hungary

Received 24 August 1981

ABSTRACT

The nucleotide sequence of the structural gene (*nifH*) of nitrogenase reductase (Fe protein) from *R.meliloti* 41 with its flanking ends is reported. The amino acid sequence of nitrogenase reductase was deduced from the DNA sequence. The predicted *R.meliloti* nitrogenase reductase protein consists of 297 amino acid residues, has a molecular weight of 32,740 daltons and contains 5 cysteine residues. The codon usage in the *nifH* gene is presented. In the 5' flanking region, sequences resembling to consensus sequences of bacterial control regions were found. Comparison of the *R.meliloti nifH* nucleotide and amino acid sequences with those from different nitrogen-fixing organisms showed that the amino acid sequences are more conserved than the nucleotide sequences. This structural conservation of nitrogenase reductase may be related to its function and may explain the conservation of the *nifH* gene during evolution.

## INTRODUCTION

A wide variety of prokaryotic organisms have the ability to fix molecular nitrogen, in a process which is catalyzed by the nitrogenase enzyme complex. This complex consists of two components: nitrogenase (MoFe protein) and nitrogenase reductase (Fe protein). These proteins have been purified from several nitrogen-fixing bacteria and were shown to share many common properties (1, 2).

Recently, the structural conservation of the nitrogenase complex has been supported by the demonstration of interspecies homology of nucleotide sequences of genes coding for nitrogenase reductase (*nifH* gene) and for one of the nitrogenase subunits (*nifD*) (3, 4, 5). In these above studies a cloned DNA fragment containing the *nif* structural genes of the free-living nitrogen-fixing bacterium *Klebsiella pneumoniae* (pSA30, 6) was hybrid-

ized with DNA from either free-living or symbiotic nitrogen-fixing bacteria. On this basis, the *nif* structural genes have been isolated from three species, *Anabaena* 7120 (4), *Rhizobium meliloti* (5, 7) and *Rhizobium japonicum* (8). Using the cloned DNA fragments, the nucleotide sequences of the *nifH* gene from *Anabaena* 7120 (9) and from *K.pneumoniae* (10, 11) have been determined.

We report here the nucleotide sequence of the *R.meliloti* 41 *nifH* gene with its 5' flanking sequences. As a starting material, we used a 4.8 kb PstI fragment, carrying *R.meliloti nifH* and *nifD* genes, cloned previously into pBR322 (pID1), where a 2.1 kb HindIII fragment had been shown to contain the entire coding region of the *nifH* gene (7). The amino acid sequence of the nitrogenase reductase deduced from the nucleotide sequence is also presented. These sequences are compared with available nitrogenase reductase sequence data of four free-living nitrogen-fixing species, *K.pneumoniae*, *Anabaena* 7120, *Clostridium pasteurianum* and *Azotobacter vinelandii*, each belonging to different and distantly related families of the prokaryotes.

## MATERIALS AND METHODS

Isotopes and enzymes. $\gamma$-$^{32}$P-ATP (~5 000 Ci/mmol) was enzymatically synthetized according to Walseth and Johnson (12). Bacterial alkaline phosphatase was purchased from Worthington. Restriction endonucleases were purified according to standard procedures (13). $T_4$ polynucleotide kinase was kindly provided by Dr A. Simoncsits.

DNA isolation and restriction fragment preparation. Plasmid pID1 DNA (Fig. 1) was prepared as described previously (7). One mg of purified pID1 plasmid DNA was used to prepare the 750 bp SalI-SalI, 1050 bp SalI-SmaI and 720 bp XhoI-XhoI fragments. To prepare DNA fragments, digested DNA was run on 1,5% vertical agarose gel. The desired DNA fragments were excised after ethidium bromide staining and syringed into 10-fold volume of elution buffer (0,1 M NaCl, 10 mM Tris-HCl, pH 7,4, 1 mM EDTA) and incubated overnight at 4°C. After centrifugation at 20,000 rpm for 60 min in a JA-20 rotor the DNA was concentrated and purified on a 1 ml DEAE 52 cellulose column (14).

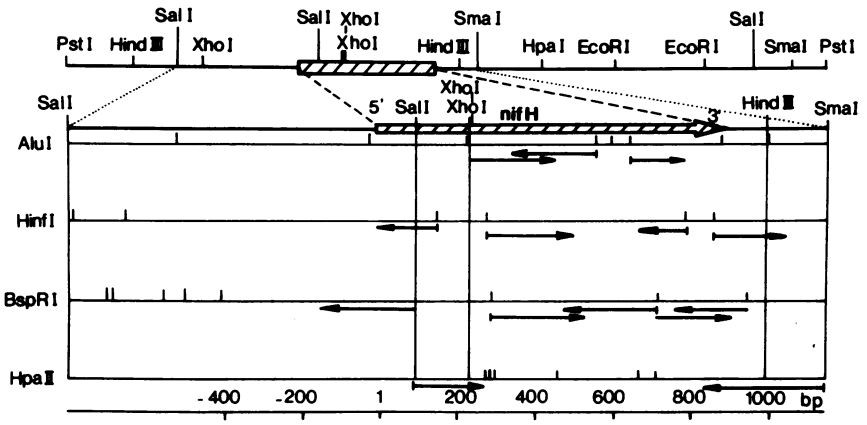


Fig. 1 Restriction map of the *R.meliloti nifH* region on pID1 and the scheme of the sequencing strategy. The restriction sites of the 4.8 kb PstI fragment of pID1 (upper part) were determined previously (7), restriction sites indicated in the lower part of the figure were mapped in this study. The extent and direction of sequence determination are indicated by the arrows.

Restriction mapping of the isolated DNA fragments. AluI, HinfI, BspRI and HpaII enzymes were used to construct a detailed restriction map of the isolated DNA fragments. 5' end labelled SalI-HindIII and XhoI-SalI fragments were partially digested with the enzymes mentioned above and mapped by the procedure of Smith and Birnstiel (14).

DNA sequencing. Restriction fragments were labelled at the 5' ends according to the method of Maxam and Gilbert (15) after dephosphorylation with bacterial alkaline phosphatase. The labelled fragments were restricted again, separated on a 7,5% polyacrylamide gel and eluted from the gel slices by electrophoresis into dialysis bags. In the case of the BspRI fragments labelled on both 5' ends, the strand separation was done according to the procedure of Maxam and Gilbert (15). For sequencing, the chemical method of Maxam and Gilbert (15) was used. Four reactions A > C, G > A, C + T and C were done. At the end of the procedure the samples were resuspended in 80% recrystallized formamide, 1 mM EDTA, 0,05% bromophenol blue

and 0,05% xylene cyanol and boiled for 1 min before loading onto 8% or 20% thin sequencing gels (16). 8% gels were used in four loadings to determine the sequence between 20-200 nucleotides. The first 20 bases were fractionated on 20% thin gels. For autoradiography Medifort RP X-ray films with Cronex Lighting Plus (DuPont) intensifying screen were used.

## RESULTS AND DISCUSSION

Sequence of the *nifH* gene. Primary sequence analysis was carried out on plasmid pID1. An approximately 1300 bp DNA segment was sequenced. The detailed restriction map of the *nifH* region and the strategy of sequencing is outlined in Fig. 1.

The isolated 739 bp SalI-SalI and 1050 bp SalI-SmaI fragments of pID1 clone (Fig. 1), selected on the basis of the previously demonstrated homology to a cloned *K.pneumoniae* DNA probe harbouring the *nifH* gene (7), were primarily used to generate small 5' end labelled DNA fragments for sequencing. A detailed restriction map was established for the 5' end labelled SalI-HindIII and XhoI-SalI fragments. When usable restriction sites were missing inside a fragment to be sequenced, single stranded 5' end labelled fragments were isolated. To eliminate the ambiguities, sequencing was done on both strands and with overlapping fragments.

The complete nucleotide sequence of the *R.meliloti nifH* gene is shown in Fig. 2. The sequence determined starts 160 nucleotides before the possible initiator AUG (position 1) and spans 280 bases beyond the 3' translational termination codon at position 892. This DNA region is exactly the same as the DNA fragment hybridizing with the *K.pneumoniae nifH* gene (7). Inspection of the three possible reading frames of the nucleotide sequence in both strands revealed that only one frame is large enough to encode the *nifH* polypeptide. This starts in the presented strand at ATG, position 1, and terminates at TGA, position 892. All the other reading frames are closed by termination codons at many places. The likely direction of transcription of the *nifH* gene was deduced by comparing its nucleotide and amino acid sequences with those of the *Anabaena* and *K.pneumoniae nifH* genes.

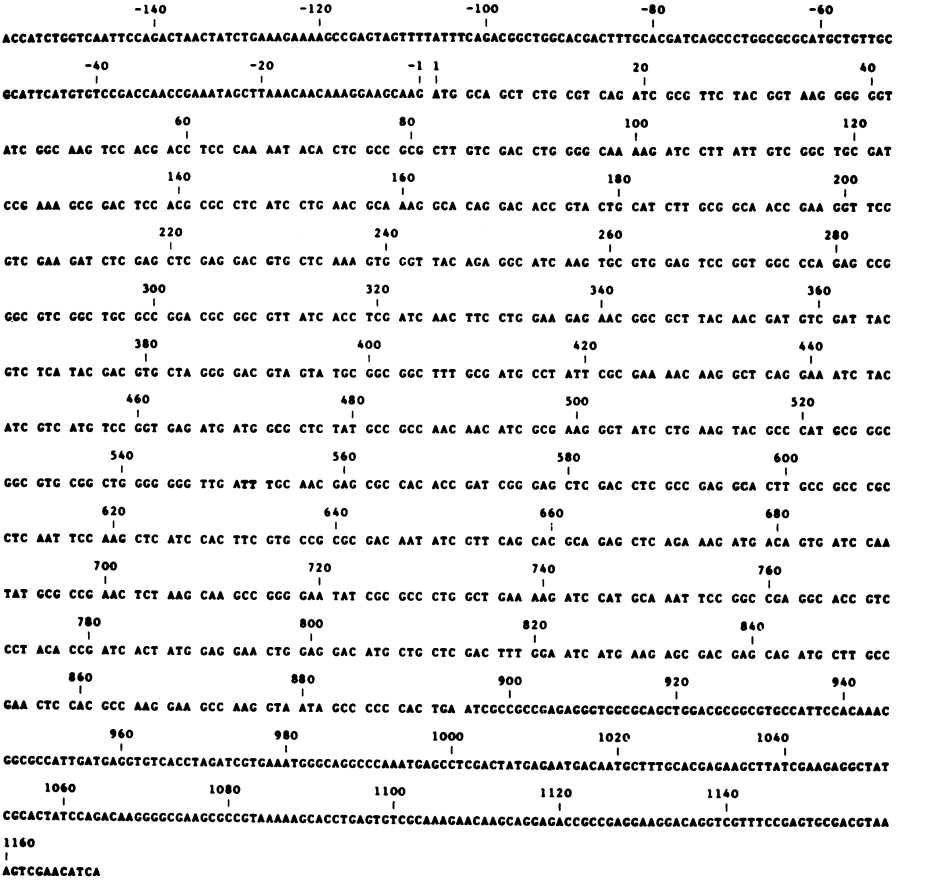


Fig. 2 Nucleotide sequence of the *R.meliloti nifH* gene with the two flanking regions. The sequence of the strand identical to the *nifH* mRNA is presented. The orientation of the *nifH* gene is the same as on the restriction map in Fig. 1. The triplets correspond to the nitrogenase reductase amino acid residues shown in Fig. 3. Since we do not have direct evidence for the transcription-initation site, numbering of nucleotide residues starts at the presumed translation-initiation codon.

It is well established that in *K.pneumoniae nifH* is the first gene in a polycistronic operon, followed by *nifD*, *nifK* and *nifY* (17, 18, 19). On the basis of hybridization of *Klebsiella nifH* and *nifD* genes with restriction fragments of pID1 carrying the *R.meliloti nifH* and *nifD* genes, a similar organi-

zation of the *R.meliloti nif* structural genes was suggested (7). Downstream from the TGA at position 892, presumed to be the termination codon for *nifH* gene, one ATG occurs at position 994 which is not in phase with any termination codons downstream in the sequenced region. This could be a potential initiator site for the *nifD* polypeptide.

Possible control regions in the *nifH* gene. No sequence data for *Rhizobium* DNA had been reported before, consequently the structures of control regions such as ribosome binding sequences, transcription-initiation and -termination sites, were not known either. In an attempt to identify these regions in the *R.meliloti nifH* gene, the region upstream from the presumptive initiation codon ATG at position 1 was inspected to find consensus sequences of bacterial control regions (20, 21, 22).

One sequence, at position -6 to -10, matching five bases (AAGGA) of the "Shine-Dalgarno" ribosome binding sequence (20) was found. It should be emphasized however, that the 3' end sequence of the *R.meliloti* 16S rRNA is not known yet.

Two other common sequences deduced from the analysis of a number of different promoters were also looked for. One of these sequences is the hexanucleotide Pribnow box, which occurs approximately ten nucleotides upstream from the starting point of transcription in *E.coli* and in some other organisms (21, 22). One candidate with only a partial homology to this consensus sequence was found at position -55 to -60. For the other consensus sequence, the so called -35 region, located about 16-18 nucleotides upstream from the Pribnow box in *E.coli* (22), a sequence at positions -79 to -85 (17 base pairs from the predicted Pribnow box) with a fairly good correspondance was found. A CAT sequence, often found at the transcription-initiation site in *E.coli* and located 5-9 base pairs downstream from the highly conservative T residue in the Pribnow box (22), is present at 5 to 7 base pairs downstream from the corresponding T residue of the predicted *R.meliloti nifH* hexanucleotide box. We must emphasize, however, that we do not have positive experimental data about the location of the transcription initiation site in the *R.meliloti nifH* gene.

In the 5' flanking region at positions -99 to -131 and -139 to -154 two sequences with a partial two-fold symmetry can form a bigger and a smaller stem-and-loop structure which might function as the transcription termination site for a gene preceding *nifH*.

Obviously, further studies are required to validate the significance of the sequences mentioned above.

The amino acid sequence of *R.meliloti* nitrogenase reductase

The amino acid sequence of the *nifH* gene product, deduced from the nucleotide sequence, is shown in Fig. 3. The predicted polypeptide consists of 297 amino acid residues and has a molecular weight of 32,740 daltons.

Due to extreme technical difficulties in handling the small nodules produced on alfalfa, *R.meliloti* nitrogenase reductase has not been purified yet. When the *R.meliloti nifH* gene was placed after a strong *E.coli* promoter, the size of the polypeptide synthesized from the *nifH* gene in *E.coli* minicells could be measured after gel electrophoresis. Preliminary data indicate that the *nifH* gene product has a molecular weight of about 33,000 daltons (Zora Svåb, personal communication).This value is in good agreement with that obtained from the sequence data.

The $NH_2$- and COOH-terminal sequences of the synthesized polypeptide have not been determined yet, but from the sequence data one can deduce a polypeptide with the molecular weight indicated above only with the termini presented in Fig. 3.

Codon utilization in the *R.meliloti nifH* gene. The codon utilization of the *R.meliloti nifH* gene is shown in Table 1. There are four codons which are not used for the amino acids present in nitrogenase reductase. In contrast to the *R.meliloti nifH*, in *Anabaena* 12 codons are not utilized (9). There are asymmetries in the codon usage for several amino acids, such as isoleucine,serine, arginine and cysteine. At present it is difficult to assess its significance; sequence data for further *Rhizobium* genes will help to answer this question. Table 1 shows that the predicted protein does not contain tryptophan. The absence of tryptophan is characteristic for nitrogenase reductases of different species (1, 2).

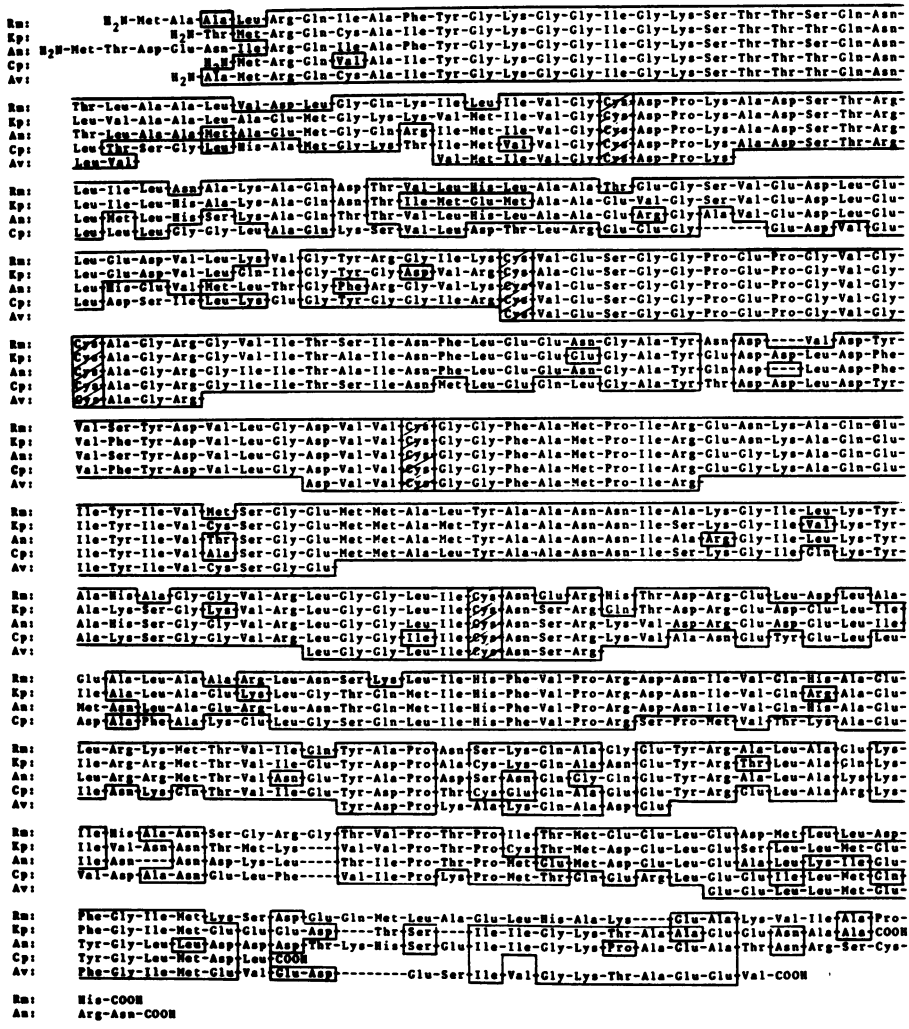

Fig. 3 Comparison of amino acid sequences of the nitrogenase reductase (Fe protein) from different nitrogen-fixing organisms. *R.meliloti*: Rm; *K.pneumoniae*: Kp (11); *Anabaena* 7120: An (9); *C.pasteurianum*: Cp (23); and *A.vinelandii*: Av(24). Conserved amino acid residues are enclosed in boxes. Those residues which occur more than once at a given position are also enclosed. The five highly conservative cysteine residues are hatched.

Table 1. Codon utilization in the *R.meliloti nifH* gene

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| UUU | Phe | 2 | UCU | Ser | 1 | UAU | Tyr | 3 | UGU | Cys | 0 |
| UUC | | 3 | UCC | | 7 | UAC | | 7 | UGC | | 5 |
| UUA | Leu | 0 | UCA | | 1 | UAA | ochre | 0 | UGA | opal | 1 |
| UUG | | 1 | UCG | | 2 | UAG | amber | 0 | UGG | Trp | 0 |
| CUU | Leu | 5 | CCU | Pro | 2 | CAU | His | 3 | CGU | Arg | 1 |
| CUC | | 13 | CCC | | 1 | CAC | | 5 | CGC | | 7 |
| CUA | | 1 | CCA | | 1 | CAA | Gln | 4 | CGA | | 1 |
| CUG | | 10 | CCG | | 5 | CAG | | 5 | CGG | | 2 |
| AUU | Ile | 3 | ACU | Thr | 1 | AAU | Asn | 4 | AGU | Ser | 0 |
| AUC | | 17 | ACC | | 6 | AAC | | 9 | AGC | | 1 |
| AUA | | 1 | ACA | | 3 | AAA | Lys | 2 | AGA | Arg | 2 |
| AUG | Met | 10 | ACG | | 2 | AAG | | 15 | AGG | | 0 |
| GUU | Val | 2 | GCU | Ala | 4 | GAU | Asp | 5 | GGU | Gly | 7 |
| GUC | | 8 | GCC | | 14 | GAC | | 11 | GGC | | 14 |
| GUA | | 4 | GCA | | 7 | GAA | Glu | 10 | GGA | | 2 |
| GUG | | 7 | GCG | | 9 | GAG | | 13 | GGG | | 6 |

Comparison of amino acid sequences of nitrogenase reductase and the nucleotide sequences of *nifH* gene from different nitrogen-fixing organisms. The comparison of the amino acid sequence of nitrogenase reductase of *R.meliloti* with the amino acid sequences from free-living nitrogen-fixing species, *Klebsiella pneumoniae* (11), *Anabaena* 7120 (9), *Clostridium pasteurianum* (23) and *Azotobacter vinelandii* (24) is shown in Fig. 3. The amino acid sequences of nitrogenase reductase of the first two species were deduced from nucleotide sequence data, while data for the second two species were obtained by amino acid sequencing of the isolated proteins. It is clear that those polypeptide regions which are highly conserved in the free-living bacteria are also strongly conserved in *R.meliloti* which can fix nitrogen only in symbiosis with alfalfa. *R.meliloti* nitrogenase reductase contains five cysteine residues, the *K.pneumoniae* protein has 9, *Anabaena* 6,

*Azotobacter* 7 and *C.pasteurianum* 6. All the five cysteine residues of *R.meliloti* are located in the highly conserved regions. As pointed out by Mevarech *et al.* (9), these strongly conserved regions probably have an important role in the maintenance of the enzyme structure and/or in its catalytic function.

The last 20 amino acid residues of the COOH end of the *R.meliloti* protein show very little homology with the COOH termini of the proteins from the other species suggesting that this region has little importance in the enzyme function. Table 2 shows the calculated homology percentages among nitrogenase reductase proteins from different sources. *R.meliloti*, *K.pneumoniae* and *Anabaena* 7120 share 67-71% homology. *C.pasteurianum* protein has somewhat lower homology with the others (58-67%). This is in line with several compartive biochemical data on the structure and catalytic activity of nitrogenase reductases from different species where the *C.pasteurianum* protein shows the least similarity to nitrogenase reductase of other species (1, 2).

If the base pair homologies among the *nifH* genes of *R.meliloti*, *K.pneumoniae* and *Anabaena* 7120 were calculated, relatively less homology (62-67%) was observed for each combination (considering that at random base pair distribution the value should be 25%). On the other hand, this degree of homology was enough to assure interspecies hybridization among *nifH* sequences (3, 4, 5). If the homology percentages were calculated from identities triplet codons at corresponding positions, these values were much lower (27-35%) (Table 2).

It is noticable that the *Anabaena* 7120 *nifH* gene has somewhat less base pair homology with *R.meliloti* and *K.pneumoniae* *nifH* genes than these latter ones have with each other. These homology differences are more apparent at triplet codon level and can be explained by the nonrandom codon utilization in *Anabaena* 7120 *nifH* gene. Differences in codon usage may indicate evolutionary distance or expressivity distinction at the translation level (25). In fact, *Rhizobium* and *Klebsiella* belong into the same taxonomical division *(Bacteria)* of prokaryotes and probably have closer evalutional relatedness than with *Anabaena* which is clarified as a member of the division *Cyano-*

Table 2. Homology among nitrogenase reductase proteins and *nifH* genes of different nitrogen-fixing organisms

| | Homology ( % ) with | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *R.meliloti* | | | *K.pneumoniae* | | | *Anabaena* 7120 | | | *C.pasteurianum* |
| | amino acids | base pairs | triplet codons | amino acids | base pairs | triplet codons | amino acids | base pairs | triplet codons | amino acids |
| *R.meliloti* | 100 | 100 | 100 | 67 | 66 | 34 | 70 | 63 | 27 | 60 |
| *K.pneumoniae* | 69 | 67 | 35 | 100 | 100 | 100 | 71 | 64 | 29 | 63 |
| *Anabaena* 7120 | 69 | 62 | 27 | 69 | 62 | 28 | 100 | 100 | 100 | 58 |
| *C.pasteurianum* | 65 | | | 67 | | | 64 | | | 100 |

Since the proteins and genes of various organisms have different length, homology was calculated for each species. The homologies, i.e. the percentages of amino acid residues, base pairs or triplet codons, identical at corresponding positions, were related to sequences of the species listed in the first column. Sequence data for *K.pneumoniae* was from Sundaresan and Ausubel (11), for *Anabaena* 7120 from Mevarech *et al.* (9), for *C.pasteurianum* from Tanaka *et al.* (23).



Nucleic Acids Research

*bacteria* (26).

Postgate (27) has suggested that the highly conserved structure of nitrogenase enzyme complex of different nitrogen-fixing organisms is the result of the late evolutionary emergence of nitrogen fixation. According to this hypothesis the nitrogen fixation genes evolved at a much later stage of evoluation than those species which are able to fix nitrogen at present, and the *nif* genes were laterally distributed via plasmids or transposable elements among a wide range of prokaryotes Sundaresan and Ausubel (11) proposed that the strict conservation of some regions in the nitrogenase reductase proteins from *K.pneumoniae, C.pasteurianum* and *A.vinelandii* might be related to their function and not merely to the late evolution of the *nif* genes. We show here that the amino acid sequence of nitrogenase reductase is more conserved than the corresponding triplet codons or the nucleotide sequence of the *nifH* gene. Moreover, out of the 17 *nif* genes of *K.pneumoniae* only two structural genes, *nifH* and *nifD*, were shown to hybridize with DNAs from other nitrogenfixing bacteria; remarkably, the third structural gene *nifK*, also did not show interspecies homology (5). Therefore, in agreement with Sundaresan and Ausubel (11) we conclude that in addition to the suggested late appearance of nitrogen fixation (27), the structural conservation of the nitrogenase components can be explained by supposing that the functioning of the nitrogenase enzyme complex requires strictly determined amino acid sequences of the *nifH* and (probably *nifD*) gene products.

ACKNOWLEDGEMENTS

We thank Angèla Németh for her valuable technical assistance and we are particularly indebted to Dr Pál Venetianer for support of this work.

REFERENCES

1. Eady, R.R. and Smith, B.E. (1979) in A Treatise on Dinitrogen Fixation, Hardy, R.W.F. Ed., pp. 299-491, John Wiley and Sons, New York.
2. Mortenson, L.E. and Thorneley, R.N.F. (1979) Annu. Rev. Biochem. 48, 387-418.

3. Nuti, M.P., Lepidi, A.A., Prakash, R.K., Schilperoort, R.A., Cannon, F.C. (1979) Nature 282, 533-535.
4. Mazur, B.J., Rice, D. and Haselkorn, R. (1980) Proc. Natl. Acad. Sci. USA 77, 186-190.
5. Ruvkun, G. and Ausubel, F.M. (1980) Proc. Natl. Acad. Sci. USA 77, 191-195.
6. Cannon, F.C., Riedel, G.E. and Ausubel, F.M. (1979) Molec. gen. Genet. 174, 59-66.
7. Bånfalvi, Z., Sakanyan, V., Koncz, C., Kiss, A., Dusha, I. and Kondorosi, A. (1981) Molec. gen. Genet. (in press).
8. Hennecke, H. (1981) Nature 291, 354-355.
9. Mevarech, M., Rice, D. and Haselkorn, R. (1980) Proc. Natl. Acad. Sci. USA 77, 6476-6480.
10. Scott, K.F., Rolfe, B.G., Shine, J., Sundaresan, V. and Ausubel, F.M. (1981) in Current Perspectives in Nitrogen Fixation, Gibson, A.H. and Newton, W.E. Eds. pp. 393-395. Australian Academy of Science, Canberra.
11. Sundaresan, V. and Ausubel, F.M. (1981) J. Biol. Chem. 256, 2808-2812.
12. Walseth, T.F. and Johnson, R.A. (1979) Biochim. Biophys. Acta 562, 11-31.
13. Roberts, R.J. (1980) Nucleic Acids Res. 8, r63-r80.
14. Smith, H.O. and Birnstiel, M.L. (1976) Nucleic Acids Res. 3, 2387-2398.
15. Maxam, A. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
16. Sanger, F. and Coulson, A.R. (1978) FEBS Letters 87, 107-110.
17. Dixon, R.A., Kennedy, C., Kondorosi, A., Krishnapillai, V. and Merrick, M. (1977) Molec. gen. Genet. 157, 189-198.
18. MacNeil, D., Roberts, G.P., Supiano, M.A. and Brill, W.J. (1978) J. Bacteriol. 136, 253-266.
19. Klipp, W. and Pühler, A. (1981) in Current Perspectives in Nitrogen Fixation, Gibson, A.H. and Newton, W.E. Eds. pp. 397. Australian Academy of Science, Canberra.
20. Shine, J. and Dalgarno, L. (1975) Nature 254, 34-38.
21. Rosenberg, M. and Court, D. (1979) Annu. Rev. Genet. 13, 319-353.
22. Siebenlist, U., Simpson, R.B. and Gilbert, W. (1980) Cell 20, 269-281.
23. Tanaka, M., Hanin, M., Yasunobu, K.T. and Mortenson, L.E. (1977) J. Biol. Chem. 252, 7013-7100.
24. Hausinger, R.P. and Howard, J.B. (1980) Proc. Natl. Acad. Sci. USA 77, 3826-3830.
25. Nichols, B.P., Blumenberg, M. and Yanofsky, C. (1981) Nucl. Acids Res. 9, 1743-1755.
26. Bergey's Manual of Determinative Bacteriology, VIII. Edition (1975) Buchanan, R.E., Gibbons, N.E., Ed., The Willams and Wilkins Company, Baltimore
27. Postgate, J.R. (1974) Symp. Soc. Gen. Microbiol. 24, 263-292.