287

Nucleotide sequence of the *Klebsiella pneumoniae* *nifD* gene and predicted amino acid sequence of the α-subunit of nitrogenase MoFe protein

Ioannis IOANNIDIS and Martin BUCK*
A.F.R.C. Unit of Nitrogen Fixation, University of Sussex, Brighton BN1 9RQ, U.K.

The nucleotide sequence of the *Klebsiella pneumoniae* *nifD* gene is presented and together with the accompanying paper [Holland, Zilberstein, Zamir & Sussman (1987) Biochem. J. 247, 277–285] completes the sequence of the *nifHDK* genes encoding the nitrogenase polypeptides. The *K. pneumoniae* *nifD* gene encodes the 483-amino acid-residue nitrogenase α-subunit polypeptide of M_r 54156. The α-subunit has five strongly conserved cysteine residues at positions 63, 89, 155, 184 and 275, some occurring in a region showing both primary sequence and potential structural homology to the *K. pneumoniae* nitrogenase β-subunit. A comparison with six other α-subunit amino acid sequences has been made, which indicates a number of potentially important domains within α-subunits.

INTRODUCTION

Biological nitrogen fixation is catalysed by nitrogenase (Gibson & Newton, 1981), an O_2-sensitive metalloenzyme that is composed of two separate proteins (Lowe et al., 1985). In *Klebsiella pneumoniae* and other genetically characterized diazotrophs the *nifH*, *nifD* and *nifK* genes encode the Fe protein and the α- and β-subunits of the MoFe protein respectively. The MoFe protein is a tetramer of the form $\alpha_2\beta_2$ and contains the site at which reducible substrates are bound (reviewed in Smith, 1977; Thorneley et al., 1978). Prosthetic groups associated with *K. pneumoniae* MoFe protein include four [4Fe-4S] clusters, two FeMo cofactors and one further class of Fe atoms (Orme-Johnson & Munck, 1980; Smith, 1983; Burgess, 1984). Analysis of *K. pneumoniae* *nifV* mutants suggests that FeMo cofactor is important in substrate specificity and reduction (Hawkes et al., 1984).

Recombinant DNA methods have permitted the identification and isolation of *nifH*, *nifD* and *nifK* genes from a number of diazotrophs, facilitated by the strong interspecies homology that exists between individual *nifH*, *nifD* and *nifK* genes and the availability of *nif* gene probes derived from the *K. pneumoniae* *nif* gene cluster (Mazur et al., 1980; Hennecke, 1981; Quiviger et al., 1982; Rice et al., 1982; Ruvkun et al., 1982). The present paper reports the completed *K. pneumoniae* *nifD* sequence.

MATERIALS AND METHODS

Restriction enzymes were obtained from Boehringer Mannheim or Northumbria Biologicals, and T4 DNA ligase was from Bethesda Research Laboratories. Digestions were performed in 33 mM-Tris/acetate buffer, pH 7.9, containing 66 mM-potassium acetate, 10 mM-magnesium acetate, 4 mM-spermidine and 5 μg of pan-creatic ribonuclease/ml. Ligations were performed in 66 mM-Tris/HCl buffer, pH 7.5, 6.6 mM-$MgCl_2$, 10 mM-dithiothreitol and 1 mM-rATP. DNA sequencing used [α-^{35}S]thio]dATP (Biggin et al., 1983) as the label (Amersham International) with deoxynucleoside and dideoxynucleoside triphosphates from P & L Laboratories. The 17 bp primer was purchased from Celltech.

The *nif* DNA to be sequenced was excised from pWF23 (Filler et al., 1986) as a 6.4 kb *Eco*RI fragment (Cannon et al., 1979), which contains the *K. pneumoniae* *nifHDKY* operon. Primary subclones of this *Eco*RI fragment were made in M13 mp18 and mp19 by restriction with *Bam*HI and *Hin*dIII (Fig. 1). The *Bam*HI site delineates the 3′ end of the partial *nifD* sequence previously determined (Scott et al., 1981), and the *Hin*dIII site lies in *nifK* (see also Holland et al., 1987). Subclones of the *Bam*HI–*Hin*dIII fragment for sequencing were made in M13 mp8, mp9, mp18 and mp19 (Messing, 1983) (see Fig. 1). DNA sequencing was carried out by using the chain termination method (Sanger et al., 1977). Secondary-structure predictions were made (Chou & Fasman, 1978) by using the Wisconsin PEPPLOT software package.

RESULTS AND DISCUSSION

Nucleotide sequence of the *nifD* gene

The nucleotide sequence of the *nifD* gene is shown in Fig. 2. Sequence was determined on both strands (Fig. 1), which with the data of Scott et al. (1981) (nucleotides 1–630 in Fig. 2) completes the 1449 bases of the *nifD* sequence and predicts a 483-amino acid-residue polypeptide of M_r 54156. Codon usage is typical of the *nifH* and *nifK* genes, with a preference for G or C in the third position (Scott et al., 1981; Holland et al., 1987). The *nifD* reading frame terminates with a single stop codon UGA, followed 55 bp downstream by the initiation codon for
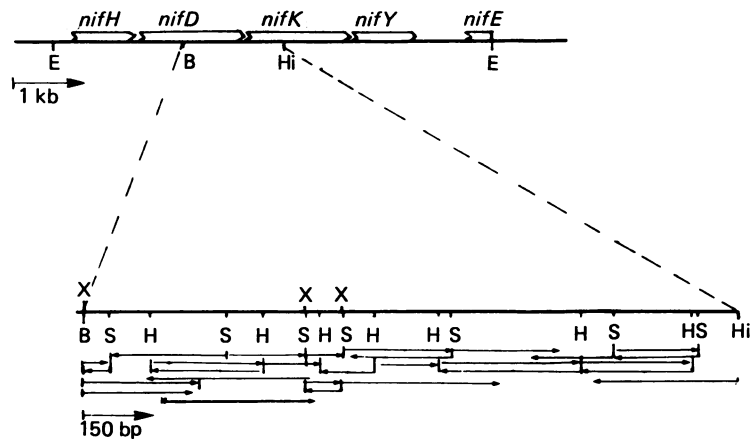
Fig. 1. Physical and genetic map of the *K. pneumoniae* nifD gene

Positions of restriction sites are given, and arrows indicate the direction and extent of each sequence obtained from subcloning into M13. A 23 bp synthetic oligonucleotide, complementary to nucleotides 827–849, was used to prime the sequencing indicated by the heavy arrow. The *Bam*HI site in the *nifD* gene was overlapped by sequencing from the *Nru*1 site (13 bp downstream of the *Bam*HI site) into the previously determined *nifD* gene sequence reported by Scott *et al.* (1981). Sequence through the *nifK* gene to the *Hind*III site was identical with that reported by Holland *et al.* (1987). B, *Bam*HI; S, *Sau*3A; H, *Hpa*II; Hi, *Hind*III; X, *Xho*II.

---

the *nifK* gene. The *nifD–nifK* intercistronic region shows homology to sequences upstream of the *nifH* gene translation start extending beyond the Shine & Dalgarno region (see Fig. 2). These sequences may have a role in ribosome binding and translation initiation during expression of the *nifH* and *nifK* genes (see, e.g., Stanssen *et al.*, 1986), but apparently not that of the *nifD* gene.

Amino acid sequence of nifD-gene product

Sequences for the *nifD*-gene products from *Anabaena* (Lammers & Haselkorn, 1983), *Bradyrhizobium japonicum* (Kaluza & Hennecke, 1984), *Rhizobium* sp. *Parasponia* (Weinman *et al.*, 1984), *Azotobacter vinelandii* (Brigle *et al.*, 1985), *Rhizobium* sp. cowpea (strain IRc78) (Yun & Szalay, 1984), *Clostridium pasteurianum* (Chen *et al.*, 1986) and *K. pneumoniae* are aligned in Fig. 3. The *C*- and *N*-termini are least well conserved, and account for the size variation amongst α-subunits. Homology of the *K. pneumoniae* α-subunit with other α-subunits is high, the greatest being with *A. vinelandii* at 72%, followed by *Rhizobium* sp. *Parasponium*, *Rhizobium* sp. cowpea (strain IRc78) and *B. japonicum* each at 70%, and then *Anabaena* at 67%. Least homology (44%) is found with *C. pasteurianum*.

Runs of two or more identical residues are marked with vertical lines and include up to eight identical amino acid residues (e.g. residues 229 through to 236; Fig. 3). Regions showing least homology are from residues 1–50, 210–218 and 390–396. These could be sites at which deletion and insertion events have occurred, as we predict in our alignment. As variation within proteins that have diverged from a common evolutionary origin is likely to occur on the protein surface, internal changes being constrained by the dense molecular packing of globular protein interiors, the rather divergent regions 210–220 and 390–396 of the α-subunits may be situated within looping-out strands of the overall tertiary protein structure.

Cysteine residues

Up to nine cysteine residues are found in α-subunits; six cysteine residues are present in the *K. pneumoniae* α-subunit and five (Cys-63, Cys-89, Cys-155, Cys-184 and Cys-275) are found in highly homologous regions of those α-subunits compared (boxed in Fig. 3). Cysteine residues may be ligands to the prosthetic groups of *K. pneumoniae* MoFe protein. Conserved Cys-63 is found in a hydrophobic region adjacent to amino acid residues (Gly-62 and Ala-64) with small side chains that are unlikely to hinder the binding of prosthetic groups should the peptide conformation at position 63 exist as a turn rather than an extended form. Secondary-structure predictions suggest that Cys-63, Cys-155 and Cys-184 occur at turns (indicated by T in Fig. 3). Cys-89 is situated in a hydrophobic region, and residues in its proximity are usually small and conserved (Gly-88, often Val-87, except for Ile-87 in *C. pasteurianum* and Ala-87 in *K. pneumoniae*, and Gly-90). In the *C. pasteurianum* α-subunit Ser-90 is present (rather than Gly-90), which has ligand potential. Prosthetic groups associated with Cys-89 in *C. pasteurianum* may therefore have different properties to those in, for example, *K. pneumoniae*. Cys-184 is conserved and found within a hydrophobic region flanked by the rather small residues Arg-183, Val-182 and Gly-186, which may favour prosthetic-group binding. Cys-275 is flanked by bulky residues (His-274 and Tyr-276), which could hinder [4Fe-4S]-group binding and create a different environment to, for example, Cys-63. It is possible that His-Cys-275-Tyr ligates FeMo cofactor. Cys-324 of *A. vinelandii* and Cys-298 and Cys-353 of *C. pasteurianum* are not conserved and are substituted by alanine, methionine, threonine, tyrosine or phenylalanine residues in other α-subunits. With the possible exception of tyrosine, these residues have a weaker potential than cysteine as ligands.

Amino acid sequences surrounding conserved cysteine residues show little conservation among themselves, in contrast with the cysteine residues of ferredoxins (Yas-
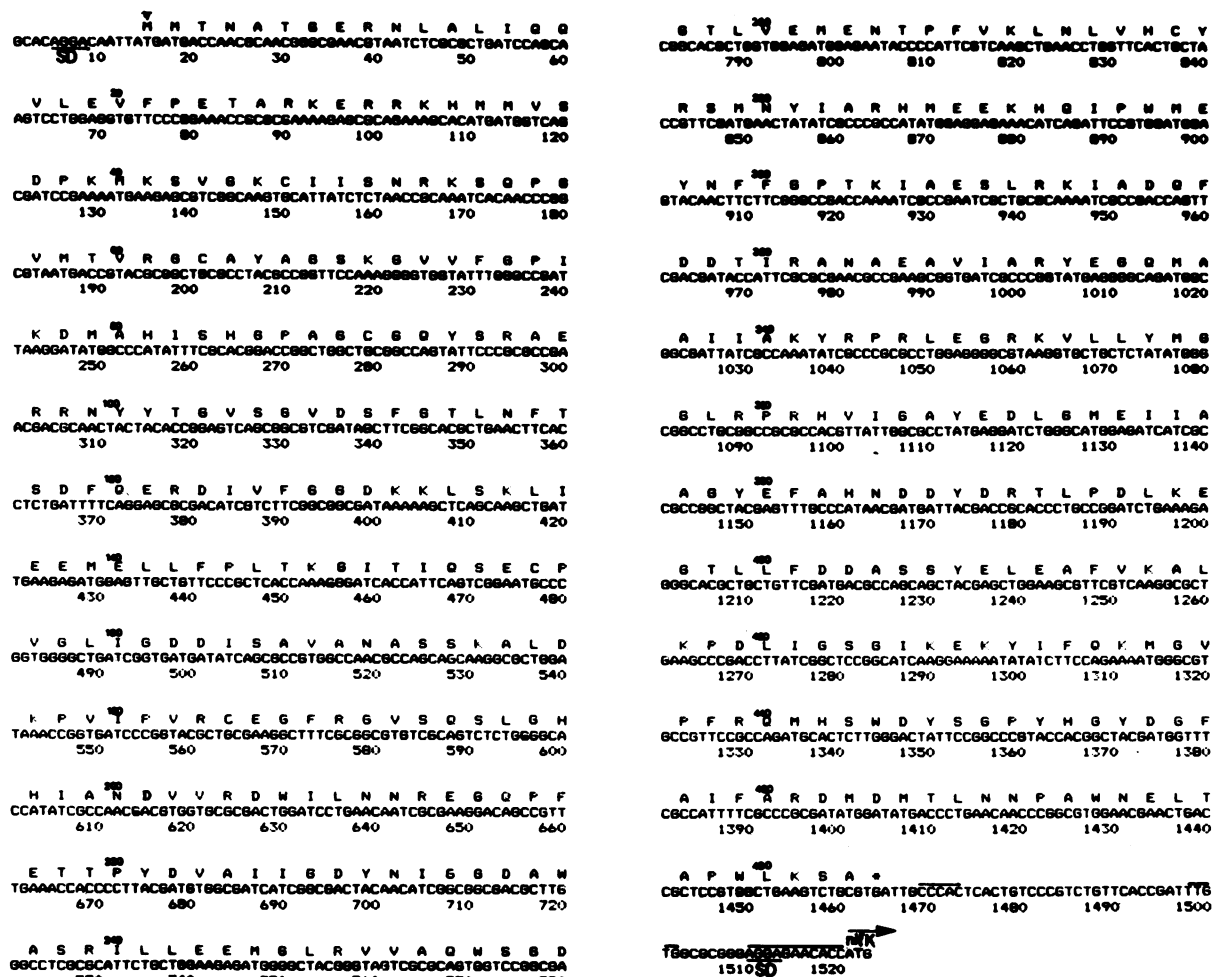
Fig. 2. Complete nucleotide sequence of the *nifD* gene of *K. pneumoniae* and its predicted translation product

Sequence up to the *BamHI* site (nucleotides 1–630) was determined previously (Scott *et al.*, 1981). The single stop codon of the *nifD* gene is marked with an asterisk. The overlined sequences upstream of the *nifK*-gene translation start are homologous to sequences upstream of the *nifH*-gene translation start. These are illustrated in the following sequence alignment:

nifH (upstream sequences): *--CCCAC---TTGTG---AGGAGAAGTCACC*

nifK (upstream sequences): *--CCCAC---TTGTG---AGGAGAA--CACC*

unobu & Tanaka, 1973). Cys-155 is the sole example within the α-subunits that is associated with an amino acid sequence similar to the sequence Glu-Cys-Pro-Val-Gly-Xaa-Ile (Xaa is a variant residue) believed to be important for iron–sulphur-centre binding in bacterial ferredoxins. The ferredoxin cysteine residue found within this sequence is contained within a type I hydrogen-bonded β-turn (Adman *et al.*, 1975). Secondary-structure analysis for the *Anabaena* (Lammers & Haselkorn, 1983) and *B. japonicum* (Kaluza & Hennecke, 1984) α-subunits in the Cys-155 region predicts a β-sheet before the Cys-155 residue followed by a turn and an α-helix. The same holds true for the *K. pneumoniae* α-subunit, indicating that Cys-155 is a [4Fe-4S]-centre ligand candidate.

Secondary-structure predictions around other conserved cysteine residues of the *K. pneumoniae* α-subunit indicate little similarity in structure to the ferredoxins. These cysteine residues do not conform to the sequence Cys-Xaa-Xaa-Cys typical of the [4Fe-4S]-cluster-binding site of ferredoxins. This may explain why the redox and

spectral properties of the [4Fe-4S] clusters of the MoFe protein are different to those of ferredoxins if the conserved cysteine residues are ligands. Some cysteine residues (Cys-63, Cys-89, Cys-155 and Cys-184) are flanked by potential oxygen-donor residues (Tyr-65, Tyr-or Gln-93 or -92, Asp-162 or -163, and Gln- or Tyr-152 respectively), which might modify the cluster's properties. Cys-275 has residues with amide and carboxylate functions surrounding it (Asn-271, Asn-280 and Glu-287) and, on the need for solvents with an amide function for extraction of the FeMo cofactor (Shah & Brill, 1977), and the binding of only one thiol molecule per molecule of cofactor (Burgess *et al.*, 1980), may be a FeMo cofactor ligand (Brigle *et al.*, 1985).

Comparison of the *K. pneumoniae* α- and β-subunits

The *K. pneumoniae* *nifK*-gene-product sequence was kindly provided by A. Zamir (Holland *et al.*, 1987). The α- and β-polypeptides have similar α-helix and β-sheet content (23 % β-sheet and 30 % α-helix for the α-subunit,

Kp:
Av:
RC:
RP:
Bj:
An:
Cp:

and 20% β-sheet and 35% α-helix for the β-subunit). Only in the N-terminal third of each subunit was structural similarity found. The first 45 amino acid residues are predicted to fold as an α-helix followed by β-sheeted regions through to position 125. Long β-sheeted areas ranging from position 176 through to position 230 were predicted for the α-subunit.

Homology between α- and β-subunits has been observed for *B. japonicum* (Thöny *et al.*, 1985) and *Anabaena* (Lammers & Haselkorn, 1983; see also Holland *et al.*, 1987). Structural homology is suggested by crystallographic studies of the *C. pasteurianum* α- and β-subunits, which indicate a 2-fold relationship between the α- and β-chains (Yamane *et al.*, 1982). Within the N-terminal of the *K. pneumoniae* β-subunit there exists sequence homology to the *K. pneumoniae* α-subunit. Residues with identity are marked with circles in Fig. 3. Little overall homology is observed except for two runs of four amino acid residues each within the third region described by Thöny *et al.* (1985). Some conserved cysteine residues are found in alike positions (Cys-63, Cys-89 and Cys-155). Those in the N-terminal regions (see above) may lie within a similar secondary structure. The significance of this remains to be shown.

REFERENCES

Adman, E., Watenpaugh, K. & Jensen, L. (1975) Proc. Natl. Acad. Sci. U.S.A. 77, 186–190

Biggin, M. D., Gibson, T. J. & Hony, G. F. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 3963–3965

Brigle, K. E., Newton, W. E. & Dean, D. R. (1985) Gene 37, 37–44

Burgess, B. K. (1984) in Advances in Nitrogen Fixation Research (Veeger, C. & Newton, W. E., eds.), pp. 103–114, Nijhoff/Junk, The Hague

---

Fig. 3. Comparison of the *K. pneumoniae* (Kp) α-subunit sequence with those from different organisms

Conserved cysteine residues are boxed. Residues identical with amino acids present in the *K. pneumoniae* β-subunit and which occur in a similar position are circled. Runs of two or more identical residues between subunits are indicated with vertical lines. Regions of low homology referred to in the text are underlined below with a broken line. To maximize homologies, packing residues indicated by dashes were introduced. Secondary structures are indicated by: α, α-helix; β, β-sheet; T, turn. Sequences compared are from *A. vinelandii* (Av), *Rhizobium* sp. cowpea (strain IRc78) (RC), *Rhizobium* sp. *Parasponium* (RP), *B. japonicum* (Bj), *Anabaena* (An) and *C. pasteurianum* (Cp). The nucleotide sequence of the *nifD* gene from *C. pasteurianum* (Chen *et al.*, 1986) predicts asparagine and arginine at positions 103 and 52 respectively rather than aspartate and lysine as deduced from protein sequencing.

Burgess, B. K., Stiefel, E. I. & Newton, W. E. (1980) J. Biol. Chem. 255, 353–356

Cannon, F. C., Riedel, G. E. & Ausubel, F. M. (1979) Mol. Gen. Genet. 174, 59–66

Chen, K. C.-K., Chen, J.-S. & Johnson, J. L. (1986) J. Bacteriol. 166, 162–172

Chou, P. Y. & Fasman, G. D. (1978) Annu. Rev. Biochem. 47, 251–276

Filler, W. A., Kemp, R. A., Ng, J. C., Hawkes, T. R., Dixon, R. A. & Smith, B. E. (1986) Eur. J. Biochem. 160, 371–377

Gibson, A. H. & Newton, W. E. (eds.) (1981) Current Perspectives in Nitrogen Fixation Research, Australian Academy of Sciences, Canberra

Hawkes, T. R., McLean, P. A. & Smith, B. E. (1984) Biochem. J. 217, 317–321

Hennecke, H. (1981) Nature (London) 291, 354–355

Holland, D., Zilberstein, A., Zamir, A. & Sussman, J. L. (1987) Biochem. J. 247, 277–285

Kaluza, K. & Hennecke, H. (1984) Mol. Gen. Genet. 196, 35–42

Lammers, P. J. & Haselkorn, R. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 4723–4727

Lowe, D. J., Thorneley, R. N. F. & Smith, B. E. (1985) in Metalloproteins (Harrison, P., ed.), vol. 1, pp. 207–249, Macmillan, London

Mazur, B. J., Rice, D. & Haselkorn, R. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 186–190

Messing, J. (1983) Methods Enzymol. 101, 20–78

Orme-Johnson, W. E. & Münck, E. (1980) in Molybdenum and Molybdenum-Containing Enzymes (Coughlin, M. P., ed.), pp. 427–438, Pergamon Press, New York

Quiviger, B., Franche, C., Lutfalla, G., Rice, D., Haselkorn, R. & Elmerich, C. (1982) Biochimie 64, 495–502

Rice, D., Mazur, B. J. & Haselkorn, R. (1982) J. Biol. Chem. 257, 13157–13163

Ruvkun, G. B., Sundaresan, V. & Ausubel, F. M. (1982) Cell 29, 551–559

Sanger, F., Nicklen, S. & Coulson, A. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463–5467

Scott, K. F., Rolfe, B. G. & Shine, J. (1981) J. Mol. Appl. Genet. 1, 71–81

Shah, V. K. & Brill, W. J. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 3249–3253

Smith, B. E. (1977) J. Less Common Metals 54, 465–475

Smith, B. E. (1983) in Nitrogen Fixation: The Chemical–Biochemical–Genetic Interface (Müller, A. & Newton, W. E., eds.), pp. 23–62, Plenum Press, New York and London

Stanssen, P., Remaut, E. & Fiers, W. (1986) Gene 36, 211–223

Thöny, B., Kaluza, K. & Hennecke, H. (1985) Mol. Gen. Genet. 198, 441–448

Thorneley, R. N. F., Eady, R. R. & Lowe, D. J. (1978) Nature (London) 272, 557–558

Weinman, J. J., Fellows, F. F., Gersshoff, M. P., Shine, J. & Scott, F. F. (1984) Nucleic Acids Res. 12, 8329–8344

Yamane, T., Weininger, M. S., Mortenson, L. E. & Rossmann, M. G. (1982) J. Biol. Chem. 257, 1221–1223

Yasunobu, K. & Tanaka, M. (1973) in Iron-Sulfur Proteins (Lovenberg, W., ed.), vol. 2, pp. 29–128, Academic Press, New York

Yun, A. C. & Szalay, A. A. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 7358–7362