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*, 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.

**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 pancreatic ribonuclease/ml. Ligation were performed in 66 mM-Tris/HCl buffer, pH 7.5, 6.6 mM-MgCl₂, 10 mM-dithiothreitol and 1 mM-ATP. DNA sequencing used [α-[^35]S]thio)dATP (Biggin et al., 1983) as the label (Amer- sham International) with deoxynucleoside and deoxydine nucleoside 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 EcoRI fragment (Cannon et al., 1979), which contains the *K. pneumoniae* *nifHDKY* operon. Primary subclones of this EcoRI fragment were made in M13 mp18 and mp19 by restriction with BamHI and HindIII (Fig. 1). The BamHI site delineates the 3' end of the partial *nifD* sequence previously determined (Scott et al., 1981), and the HindIII site lies in *nifK* (see also Holland et al., 1987). Subclones of the BamHI–HindIII 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*, 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...
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-
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-152 or Glu-93 or -92, Asp-162 or -163, and Glu- 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,
Sequence of *Klebsiella pneumoniae* nitrogenase \( \alpha \)-subunit

and 20\% \( \beta \)-sheet and 35\% \( \alpha \)-helix for the \( \alpha \)-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 \( \alpha \)-helix followed by \( \beta \)-sheeted regions through to position 125. Long \( \beta \)-sheeted areas ranging from position 176 through to position 230 were predicted for the \( \alpha \)-subunit.

Homology between \( \alpha \) - and \( \beta \)-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* \( \alpha \)- and \( \beta \)-subunits, which indicate a 2-fold relationship between the \( \alpha \) - and \( \beta \)-chains (Yamane *et al.*, 1982). Within the \( N \)-terminal of the *K. pneumoniae* \( \beta \)-subunit there exists sequence homology to the *K. pneumoniae* \( \alpha \)-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.

We thank Barry Smith, Ray Dixon, Martin Drummond and Mike Merrick for valuable discussions, John Postgate for his support and critical reading of this paper and Beryl Scott for typing the manuscript.

### REFERENCES


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

Conserved cysteine residues are boxed. Residues identical with amino acids present in the *K. pneumoniae* \( \beta \)-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: \( \alpha \), \( \alpha \)-helix; \( \beta \), \( \beta \)-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.


Received 26 November 1986/5 March 1987; accepted 25 June 1987

Vol. 247