Biochem. J. (1987) 247, 277–285 (Printed in Great Britain)

A quantitative approach to sequence comparisons of nitrogenase MoFe protein α- and β-subunits including the newly sequenced *nifK* gene from *Klebsiella pneumoniae*

Doron HOLLAND,* Aviah ZILBERSTEIN,*‡ Ada ZAMIR*§ and Joel L. SUSSMAN†

*Department of Biochemistry and †Department of Structural Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel

The nucleotide sequence was determined for part of the *Klebsiella pneumoniae nif* gene cluster containing the 3′ end of the *nifD* gene and the entire length of the *nifK* gene (encoding the α- and β-subunits of the nitrogenase MoFe protein respectively), as well as the putative start of the *nifY* gene, a gene of as yet unknown function. A broad-based comparison of a number of MoFe protein α-subunits, β-subunits and α- versus β-subunits was carried out by the use of a computer program that simultaneously aligns three protein sequences according to the mutation data matrix of Dayhoff. A new kind of quantitative statistical measure of the similarity between the aligned sequences was obtained by calculating and plotting standardized similarity scores for overlapping segments along the aligned proteins. This calculation determines if a test sequence is similar to the consensus sequence of two other proteins that are known to be related to each other. The different β-subunits compared were found to be significantly similar along most of their sequence, with the exception of two relatively short regions centred around residues 225 and 300, which contain insertions/deletions. The overall pattern of similarity between different α-subunits exhibits resemblance to the overall pattern of similarity between different β-subunits, including regions of low similarity centred around residues 225 and 340. Comparison of α-subunits with β-subunits showed that a region of significant similarity between the two types of subunits was located approximately between residues 120 and 180 in both subunits, but other parts of the proteins were only marginally similar. These results provide insights into likely tertiary structural features of the MoFe protein subunits.

INTRODUCTION

Biological nitrogen fixation is catalysed by nitrogenase, a complex of the MoFe protein (component I) and the Fe protein (component II). Component I, thought to be responsible for substrate binding and reduction, is a heterotetramer consisting of two pairs of α- and β-subunits, encoded by *nifD* and *nifK* genes respectively (Cannon *et al.*, 1985). The apoprotein is associated with several iron–sulphur clusters, and two copies of a dissociable FeMo cofactor (Mortenson & Thorneley, 1979; Burgess, 1984). Comparison of available gene nucleotide sequences and amino acid sequences of MoFe protein subunits from different organisms indicates that different α-subunits or β-subunits are highly conserved (Lundell & Howard, 1981; Mazur & Chui, 1982; Lammers & Haselkorn, 1983; Hase *et al.*, 1984; Weinman *et al.*, 1984; Yun & Szalay, 1984; Kaluza & Hennecke, 1984; Thony *et al.*, 1985). Five conserved cysteine residues in α-subunits and three conserved cysteine residues in β-subunits were proposed to function as ligands for metal–sulphur clusters (Lammers & Haselkorn, 1983; Thony *et al.*, 1985). Structural symmetry between α- and β-subunits was suggested by low-resolution X-ray-diffraction analysis of component I from *Clostridium pasteurianum* (Yamane *et al.*, 1982), and, more recently, limited sequence homology between α- and β-subunits was proposed for three regions located in the *N*-terminal third of the proteins (Thony *et al.*, 1985).

All sequence comparisons of MoFe protein subunits conducted to date relied on manual alignments and mostly stressed sequence conservation. However, with the accumulation of sequence data, a more objective and quantitative approach to sequence comparison is desirable.

In the present study, we determined the nucleotide sequence of the *nifK* gene from *Klebsiella pneumoniae*. The sequences of nitrogenase-encoding genes from *K. pneumoniae* are of particular interest in view of the wealth of mutants available (MacNeil *et al.*, 1978; Merrick *et al.*, 1980) and the relative ease of genetic manipulation of *K. pneumoniae*. These features of *K. pneumoniae* make it an ideal model system for site-directed mutagenesis of the *nifD* and *nifK* genes. The amino acid sequence deduced for the *K. pneumoniae nifK*-gene product was included in a computer-assisted comparison of different MoFe protein subunit sequences. The program used (Murata *et al.*, 1985) simultaneously aligns three protein sequences according to amino acid identities as well as similarities. Standardized similarity scores, calculated for segments of the aligned sequences, serve to measure quantitatively the similarity between the different parts of the compared proteins. Applying

‡ Permanent address: Department of Botany, The George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel.
§ To whom reprint requests should be sent.
These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00315.

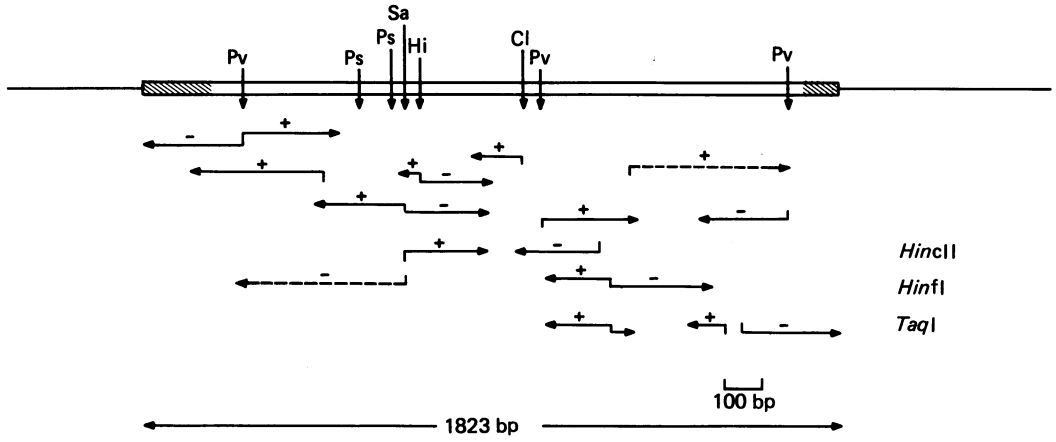

Fig. 1. Restriction map and sequencing strategy for the *nifK* gene

Restriction map of the sequenced region of DNA: open bar, *nifK* coding sequence; hatched bar, *nifK* flanking regions. Restriction sites: PV, *Pvu*; Ps, *PstI*; Sa, *SalI*; Hi, *Hin*dIII; Cl, *ClaI*. Lines, sequenced fragments generated by digestion with the enzymes indicated; arrow heads, direction of sequencing; dashed arrows, site and direction of sequence obtained by using the dideoxy chain-termination sequencing method. Strands are denoted as + and −.

this analysis to the sequences of *nifK*-gene and *nifD*-gene products provided a detailed quantitative representation of the relationships between different α- or β-subunit sequences. The analysis also defined a region of significant similarity between the two subunits that was difficult to discern unambiguously by a more conventional approach using pairwise sequence comparisons.

EXPERIMENTAL

DNA

A cloned 6.3 kb *Eco*RI fragment containing the *K. pneumoniae nifHDKY* operon (Cannon *et al.*, 1979), served as the source of DNA for sequence analysis.

DNA sequence analysis

The analysis was performed essentially in accordance with Maxam & Gilbert with some modifications (Smith & Calvo, 1980). Sequence data were also obtained by using the dideoxy chain-termination sequencing method of Sanger *et al.* (1977) as described in Amersham's sequencing handbook. Four 17-mer synthetic primers complementary to different *nifK* regions in the orientation subcloned into M13mp13 vector were used.

Computer procedures

Simultaneous three-sequence comparisons were carried out by using a program developed by Murata *et al.* (1985), an extension of the two-sequence alignment algorithm of Needleman & Wunsch (1970), using the Dayhoff & Schwartz similarity matrix MDM_{78} (Dayhoff *et al.*, 1978). The score for each amino acid position was obtained by summing up the three pairwise similarity scores, with a gap penalty of −6 for each deletion irrespective of its length. Owing to limitations of virtual memory, even on a large mainframe IBM 3081 computer (16 Mbytes), each run of the three-sequence comparison program included no more than a length of 194 amino acid residues per protein. The scanning was done on a window of this length, and then moving over 30 amino acid residues. The extensive overlaps between consecutive runs (164 amino acid residues) allowed the unambiguous joining of all the separate alignments. This method assumes a co-linear relation between the three compared sequences, and without modification would not be suitable for the analysis of proteins with circularly permuted sequences. The degree of similarity along the aligned chains was estimated by calculating standardized similarity scores for a window of 60 amino acid residues, with one sequence as a frame of reference, and moving the window 20 amino acid residues at a time. For graphic presentation, standardized scores were plotted as a function of the middle position of the 60-amino acid-residue window.

Standardized similarity score = (similarity score − mean of 100 random scores)/standard deviation of random scores.

Similarity score: sum of scores of three pairwise matches summed over each position in the window; when there was a gap of any length in one or two of the sequences, with respect to the reference sequence, a gap penalty of 6 was subtracted.

Random score: similarity scores obtained for each window, when one of the tested sequences was randomized, while the amino acid composition was kept constant. When a test sequence was being compared against two sequences already known to be related, then the test sequence was also the one chosen to be randomized. If three similar sequences were being compared, then arbitrarily one was chosen to be randomized. Assuming a normal distribution of scores for random sequences, the random probability of any window having a standardized similarity score of 3 standard deviations, or larger, is less than 1%.

The VAX 11/780 VMS and IBM 3081 VM/CMS computers were used in this analysis.

RESULTS

Sequence of the *nifK* gene and adjoining regions

The entire *nifHDKY* operon in *K. pneumoniae* is situated within a 6.3 kb *Eco*RI fragment (Cannon *et al.*, 1979). Plasmid clones of this fragment were used for DNA sequence analysis and for the generation of

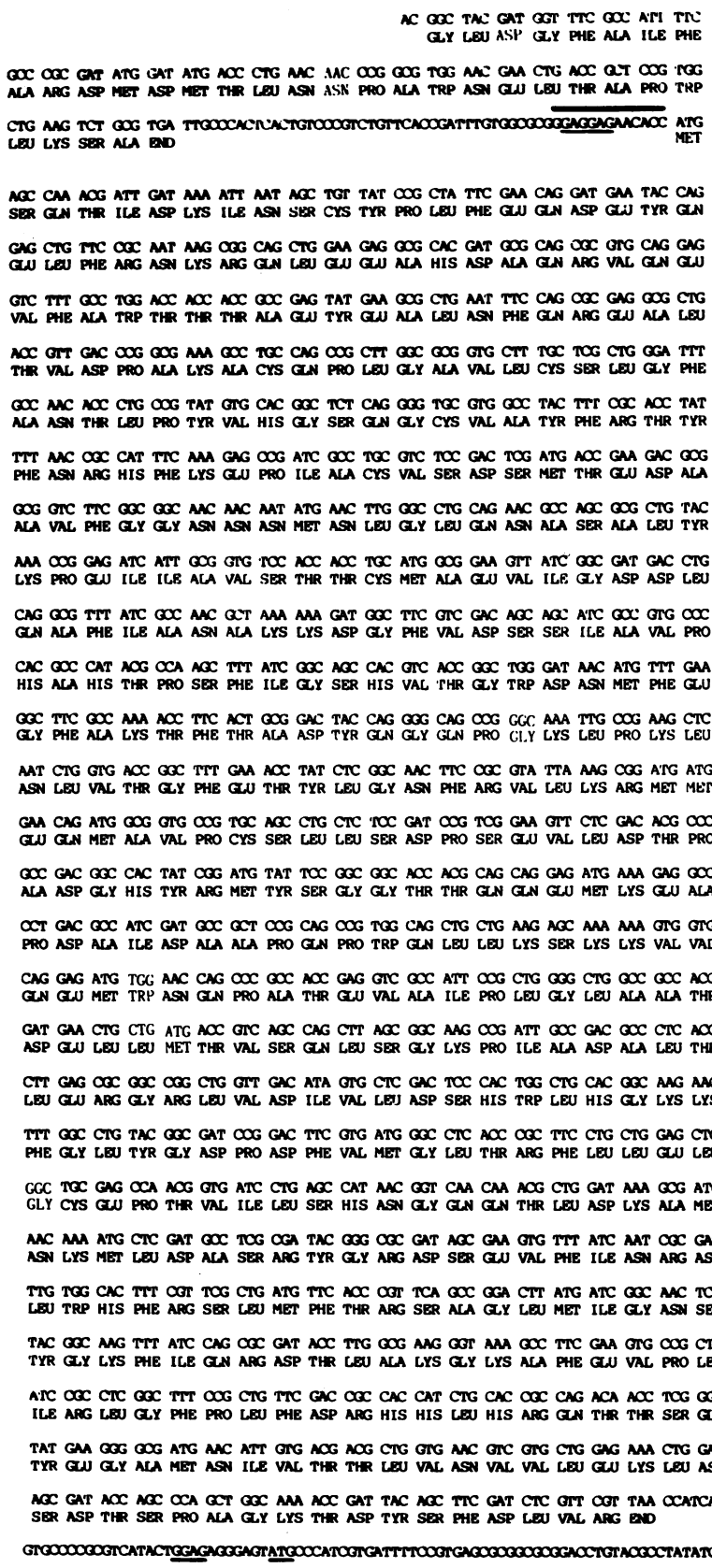

Fig. 2. Nucleotide sequence of the *nifK* gene and part of the *nifD* gene and the deduced amino acid sequence

The part of the open reading frame on top corresponds to the 3′ end of the *nifD* gene. Underlines, putative SD sequence of the *nifK* gene and SD and ATG sequences of the *nifY* gene; heavy overline, sequence homology to the sequence upstream to the *nifH* coding sequence (Scott *et al.*, 1981):

nifH: AGGAGAAGTCACC
nifK: AGGAGAA--CACC

Fig. 3. Comparison of β-subunit sequences from different organisms

(*a*) The sequences were computer-aligned as described in the Experimental section, with the top sequence serving as the frame of reference. Kp, *K. pneumoniae*; An, *Anabaena* 7120; PR, *Rhizobium* sp. *Parasponia*. Underlines, identical residues in the three sequences; asterisks, conserved cysteine residues. Because of computing limitations, the sequences shown are no longer than 499 amino acid residues, and do not include the *C*-terminal residues. (*b*) Amino acid sequence alignment around Cys-112. Av, *A. vinelandii*; Bj, *B. japonicum*.

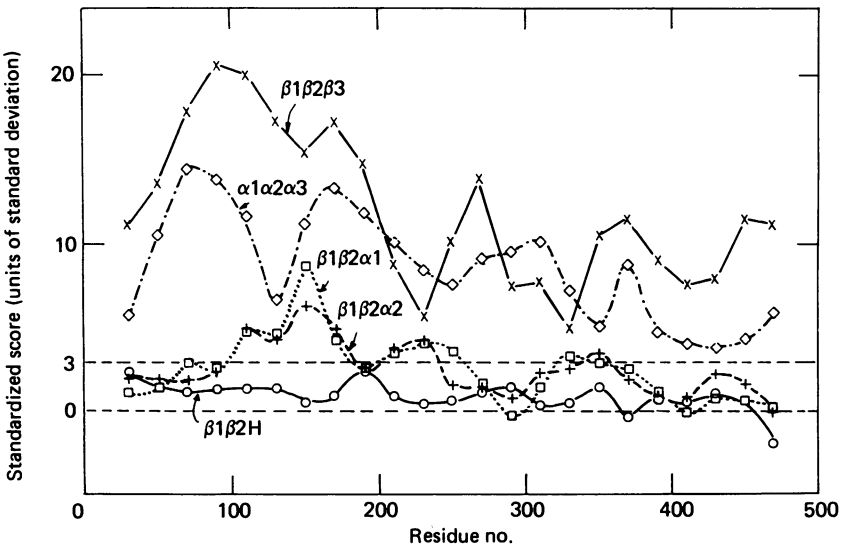

Fig. 4. Statistical analysis of sequence similarities between different MoFe protein subunits

The statistical significance of the similarity between the aligned sequences of β-subunits (Fig. 3), α-subunits (Fig. 5) and α- versus β-subunits (Fig. 6) was determined and plotted as described in the Experimental section. To align the two β-subunits (the same as shown in Fig. 6) with the mouse transplantation antigen H-2D^b, the latter, 338-amino acid-residue long, sequence was extended by adding a repeat of the 161 amino acid residues from the *N*-terminal on the protein. The resulting 499-amino acid-residue-long sequence was aligned with and compared with the β-subunit sequences. In all cases, the third sequence was the one randomized. β-Subunits: β1, *K. pneumoniae*; β2, *Anabaena* 7210; β3, *Rhizobium* sp. *Parasponia*. α-Subunits: α1, *Rhizobium* sp. *Parasponia*; α2, *Anabaena* 7120; α3, *C. pasteurianum*. H, mouse transplantation antigen H-2D^b. Horizontal broken line at 3 standard deviation units describes the limit of statistical significance (see the Experimental section). Residue numbers refer to the sequence serving, in each comparison, as the frame of reference.

subclones. From the partial sequence of the *nifD* gene from *K. pneumoniae* (Scott *et al.*, 1981) and the estimated M_r of the *nifD*-gene product (Roberts *et al.*, 1978) it was possible to deduce the approximate location of the *nifK* gene. This region was subjected to DNA sequence analysis. The restriction map of the sequenced DNA and the strategy used to determine the 1823-nucleotide sequence are summarized in Fig. 1, and the sequence is shown in Fig. 2. Homology to the 3′ end of the *nifD* gene from *Rhizobium* sp. cowpea (Yun & Szalay, 1984) provided the basis for identification of the 3′ end of the *nifD* gene, and the *nifK* coding region was identified by its homology to other *nifK* gene sequences (Lundell & Howard, 1981; Mazur & Chui, 1982; Hase *et al.*, 1984). The sequence AGGAG, found in the *nifD–nifK* intercistronic region at positions −7 to −11 with respect to the *nifK* coding sequence, is the most likely candidate for the *nifK* gene Shine & Dalgarno (SD) sequence. Interestingly, this sequence is part of a longer sequence (from position −1 to −11) that is homologous to the sequence between −1 to −13 relative to the *K. pneumoniae nifH* coding sequence (Scott *et al.*, 1981). This sequence is different from the corresponding region of the *nifD* gene (Scott *et al.*, 1981).

At 38 bp downstream to the *nifK* gene, an ATG preceded by a putative SD sequence is likely to represent the start of the *nifY* gene (Puhler & Klipp, 1981).

Deduced amino acid sequence of *nifK*-gene product and comparison with other β-subunit sequences

As deduced from the nucleotide sequence, the β-subunit of the MoFe protein from *K. pneumoniae* is 518 amino acid residues long and of M_r 57751.

The sequence of the *nifK*-gene product from *K. pneumoniae* was simultaneously compared with two other *nifK*-gene-product sequences from *Rhizobium* sp. *Parasponia* (Weinman *et al.*, 1984) and *Anabaena* 7120 (Mazur & Chui, 1982). These three organisms belong to phylogenetically separated organisms (Hennecke *et al.*, 1985). The alignment (Fig. 3*a*) shows that conserved amino acid residues occur in 1–12-residue clusters along the entire length of the proteins and account for approx. 40% of the total amino acid residues. In addition to the three conserved cysteine residues (69, 94 and 152, numbering as in the sequence from *K. pneumoniae*) noted in previous comparisons (Mazur & Chui, 1982; Weinman *et al.*, 1984), the present alignment indicates the presence of a fourth conserved cysteine residue, at position 112. A corresponding cysteine residue is also found in β-subunits from *Azotobacter vinelandii* (Lundell & Howard, 1981) and *Bradyrhizobium japonicum* (Thony *et al.*, 1985) (Fig. 3*b*). Examination of the alignment of all five sequences shows that, except for the sequence from *Anabaena* 7120, all the cysteine residues corresponding to Cys-112 in the *K. pneumoniae* sequence can be matched without introduction of gaps, and are flanked by conserved residues on one side and by mostly conservatively replaced amino acid residues on the other side.

The three-sequence alignment also shows that gaps, introduced to optimize the alignment of the three β-subunit sequences, are not distributed randomly, but are clustered mostly in two regions: 208–230 and 273–305.

To assess quantitatively the structural variation along the β-subunits, standardized similarity scores were calculated for segments along the aligned sequences. The results (Fig. 4) indicate that the degree of similarity varies between different parts of the β-subunits. The *N*-terminal third is the most structurally conserved region.

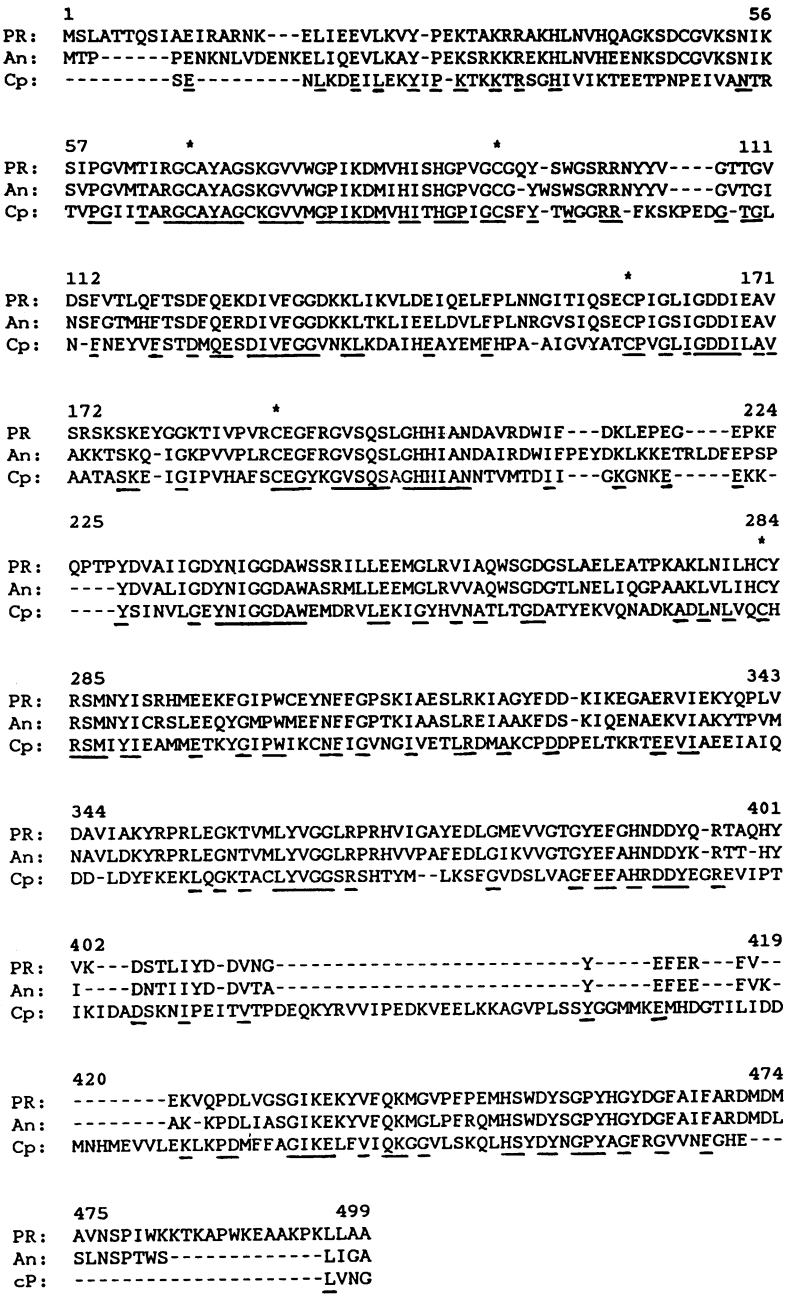

Fig. 5. Comparison of α-subunit sequences from different organisms

The sequences were aligned as described in the Experimental section and in the legend of Fig. 3. PR, *Rhizobium* sp. *Parasponia*; An, *Anabaena* 7120; Cp, *C. pasteurianum*. Underlines, identical residues in the three sequences; asterisks, conserved cysteine residues.

Two minima, around residues 220 and 320, represent regions where insertions/deletions are clustered and conserved residues are relatively scarce.

Comparison of α-subunit sequences

The same method was used to align and assess the similarity of α subunit sequences from *Rhizobium* sp. *Parasponia* (Weinman *et al.*, 1984), *Anabaena* 7120 heterocysts (Golden *et al.*, 1985) and *C. pasteurianum* (Hase *et al.*, 1984). Similarly to β subunits, this group of sequences was derived from phylogenetically separated organisms (Hennecke *et al.*, 1985). Two other sequences available at the time of the comparison, from *Rhizobium* sp. cowpea (Yun & Szalay, 1984) and *B. japonicum* (Kaluza & Hennecke, 1984), were nearly identical with the *Rhizobium* sp. *Parasponia* sequence analysed. The sequence alignment (Fig. 5) shows approx. 30% amino acid residue conservation, including the five cysteine residues observed in previous comparisons (Lammers & Haselkorn, 1983; Weinman *et al.*, 1984). As in the β-subunits, insertions/deletions are clustered in several regions: the *N*-terminus and around residues 240 and 420. In the last-mentioned region, the sequence from *C. pasteurianum* contains an extensive insertion with respect to the other two sequences.

The plot of the standardized scores for the aligned

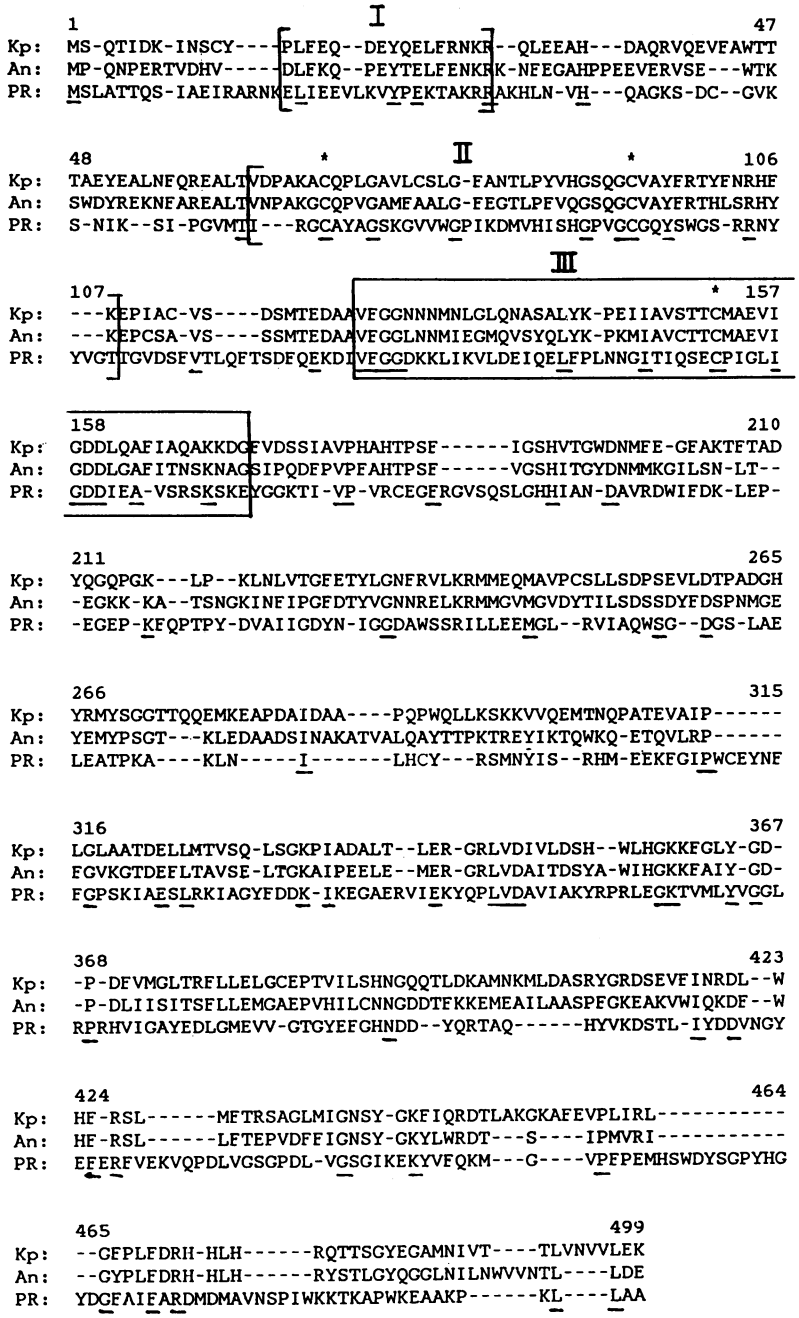

Fig. 6. Comparison between α-subunit and β-subunit sequences

The sequences were aligned as described in the Experimental section and the legend to Fig. 3. Kp, β-subunit of *K. pneumoniae*; An, β-subunit of *Anabaena* 7120; PR, α-subunit of *Rhizobium* sp. *Parasponia*. Underlines, identical residues in the three sequences; asterisks, conserved cysteine residues; bracketed and boxed sequences with Roman numerals, previously proposed regions of homology (Thony *et al.*, 1985); box, sequences showing statistically significant similarity according to the analysis shown in Fig. 4.

sequences (Fig. 4) shows a pattern of similarities resembling that of the β-subunits; the *N*-terminal third is the most highly similar, a region of relatively low similarity (partly due to insertions/deletions) extends from residue 220, and another minimum is evident around residue 340.

Comparison of α- and β-subunit sequences

The relationship between the two subunits of the MoFe protein was studied by comparing one α-subunit sequence with two β-subunit sequences. This type of matching is expected to emphasize conserved, rather than accidental, similarities between the α- and β-subunits.

One alignment included β-subunits from *K. pneumoniae* and *Anabaena* 7120 and the α-subunit from *Rhizobium* sp. *Parasponia* (Fig. 6). The same β-subunits were also aligned with the α-subunit from *Anabaena* 7120 (alignment not shown). It is evident from these alignments that, although the α-subunits share, in general, very little

sequence homology with the β-subunits, some conserved features are apparent. These include two tetrapeptide sequences (also present in all other sequenced α- and β-subunits), and three cysteine residues, with a glycine residue always preceding Cys-92. These represent the three conserved cysteine residues originally found in β-subunits (Thony *et al.*, 1985) and three out of the five conserved cysteine residues in α-subunits (Lammers & Haselkorn, 1983). The plot of the standardized similarity scores (Fig. 4) shows statistically significant similarity between the regions extending approximately between residues 120 and 180 in both α- and β-subunits. This region includes the two identical tetrapeptides and one of the conserved cysteine residues in α- and β-subunits (Fig. 6). Lower, but still greater than 3σ, standardized similarity scores (see the Experimental section for definition) are seen for several other regions. In the absence of additional data, it is difficult for us to assess the significance of similarities just above the 3σ level (1% probability of occurring by chance) in view of the fact that these are the regions that are not highly conserved between different α-subunits or β-subunits (Fig. 4). It is also evident that the alignment of β1 and β2 is somewhat different in the β1β2α1 (Fig. 6) and β1β2β3 (Fig. 3) comparisons, especially in regions that are not highly conserved between the β-subunits. However, the alignments are identical in the regions of the highest similarity between α- and β-subunits.

One concern about the statistical significance analysis of the similarity between α- and β-subunits was that it could be strongly biased by the inclusion of two β-subunit sequences. To eliminate this possibility, the two β-subunits were compared, not just to computer-generated random sequences, but to several natural, presumably non-related, proteins. As an illustration, we show the analysis of the mouse transplantation antigen H-2D^b (Reyes *et al.*, 1982). The standardized scores for this alignment, as well as other proteins tested, do not exceed 3σ, and hence we conclude that the inclusion of two β-subunits does not have a dominating effect on the statistical analysis, and that the similarity patterns are specific for the proteins tested. Thus this approach provides an objective criterion for determining if a test sequence is similar to the consensus of two other related proteins.

DISCUSSION

In this study we determined the sequence of part of the *nifHDKY* operon from *K. pneumoniae* that includes the 3′ end of the *nifD* gene, the *nifD–nifK* intercistronic region, the *nifK* coding region and the sequence downstream to the *nifK* gene, possibly including the start of the *nifY* gene.

Sequences of different *nifD* genes exhibit considerable length and sequence variation at their 3′ end; for example, the sequence of the *nifD* gene from *K. pneumoniae* shows similarity to sequences from *Rhizobium* sp. cowpea and *Rhizobium* sp. *Parasponia*, but differs from that of *C. pasteurianum* and *Anabaena* 7120 vegetative cells. When *Anabaena* cells differentiate into heterocysts, recombination within the *nifD* gene sequence (Golden *et al.*, 1985) generates a 3′ end that matches more closely the sequences from the *Rhizobium* strains and, as shown here, also the sequence from *K. pneumoniae*.

The 55 bp intercistronic distance between the *nifD* and *nifK* genes is longer than the region separating the coding sequences of the *nifH* and *nifD* genes (Scott *et al.*, 1981). It is possible that the *nifD–nifK* intercistronic region may possess some regulatory function. For example, the sequence may provide for highly efficient ribosome binding that, by partly counteracting natural polarity, may allow for the synthesis of comparable amounts of MoFe protein α- and β-subunits. In this respect, it is of interest to note the sequence homology between the regions immediately upstream to the coding sequence of the *nifH* gene, a highly expressed gene (Cannon *et al.*, 1985), and the *nifK* gene.

On the basis of qualitative assessments, the MoFe protein subunits from different organisms were shown to be closely related. However, with the growth in the number of known sequences, we considered that a systematic quantitative analysis was timely. The advantages of simultaneous comparison over the conventional pairwise comparison in aligning three sequences, as demonstrated by Murata *et al.* (1985), are: a consistent alignment between the three sequences is obtained without manual adjustment, and it can reveal homologies that might be missed by the pairwise comparisons.

Comparison of the three different β-subunits indicates that they may share a fourth conserved cysteine residue in addition to the previously identified three cysteine residues (Thony *et al.*, 1985). Although it initially appeared that insertions/deletions were common in the vicinity of this cysteine residue, examination of two additional β-subunit sequences showed that continuous alignment was in fact possible for four of the five sequences compared (Fig. 3*b*). In the four fully aligned sequences, Cys-112 (residue numbers in the sequence from *K. pneumoniae*) was flanked by conserved, or mostly conservatively replaced, amino acid residues. The exceptional sequence is that from *Anabaena* 7120, in which the order of the amino acid residues between the conserved Pro-109 and Val-113 might have been inverted. This deviation may not exclude a function for Cys-112 in iron–sulphur clusters or FeMo-cofactor liganding, as has been proposed for the other three conserved cysteine residues. Thus a total of 18 conserved cysteine residues may be present in the $\alpha_2\beta_2$ MoFe protein.

The alignment of the three β-subunits or three α-subunits required the introduction of gaps, presumably representing insertions/deletions introduced in the course of evolution. The most extensive gaps were introduced to accommodate the alignment of the *C*-terminus of the α-subunit from *C. pasteurianum* with the two more closely related sequences from *Rhizobium* sp. *Parasponia* and *Anabaena* 7120. Clusters of shorter insertions/deletions are evident in other regions of the aligned α-subunits, as well as in β-subunits. It is particularly intriguing to note the presence of such clusters in comparable positions (approximately between residues 210 and 230) in both α- and β-subunits. Another cluster is evident in β-subunits, approximately between residues 290 and 315.

Surprisingly, there is a general resemblance between the plots of standardized similarity scores obtained for α-subunits and β-subunits. For both groups of sequences, the highest scores are obtained for regions included between the *N*-terminus and the region of insertions/deletions around residue 220. In β-subunits this region

contains all four conserved cysteine residues. In α-subunits, it contains four out of the five conserved cysteine residues. Although the relatively high conservation of the *N*-terminal parts of the two MoFe protein subunits was previously noted (Weinman *et al.*, 1984), the graphical presentation indicates that the remaining parts of the proteins also exhibit a parallel pattern of conservation. Although the full significance of this resemblance remains to be clarified, it might suggest that α- and β-subunits are similarly subdivided along their length. It might further suggest that the two subunits are related in their three-dimensional domain structure.

Several lines of evidence suggested that α- and β-subunits may share a common evolutionary origin and some structural similarity (Lundell & Howard, 1981; Yamane *et al.*, 1982). In a recent comparison (Thony *et al.*, 1985), sequence homologies between α- and β-subunits were noted in three separate regions (I, II and III; Fig. 6) located in the *N*-terminal part of the proteins. In agreement with these conclusions, the computer-assisted alignment of α- and β-subunits matched three of the conserved cysteine residues in each of the subunits, as well as the two identical tetrapeptides, and several other identical residues. However, examination of the standardized similarity scores showed that only one of the previously proposed regions of homology, III, exhibited a highly significant similarity. Some similarity could also exist in the *C*-terminal part of region II. However, the similarity in region I and in most of region II fails to reach statistical significance.

The regions of similarity between the two subunits, roughly bounded by the two homologous tetrapeptides, are also highly conserved in each subunit. This region in α-subunits was also identified, in a search of a library of 2372 protein sequences (not shown), as the only sequence similar to the β-subunit from *K. pneumoniae*. (This search also showed that there was no circular permutation in β-subunit sequences.)

It is likely that the sequence relationships revealed in these analyses reflect some features in the tertiary folding of the proteins. According to the simplest interpretation, the more highly conserved regions in α-subunits or β-subunits are responsible for the major folding properties of the polypeptides. The relatively divergent sequences may be present in looped-out regions or perhaps link separate domains in the proteins. It is of interest to note here a possible analogy with the influenza-virus neuraminidase, where amino acid residues that change during antigenic drift are observed to cluster preferentially on the surface loops (Colman *et al.*, 1983). Similarly, in immunoglobulins, the walls of the antigen-binding site consist exclusively of hypervariable regions (Segal *et al.*, 1974; Wu & Kabat, 1970). In both these cases, the variable amino acid residues play a role in the recognition, or in the function of the proteins, but not in their basic architecture.

A possible role of the relatively variable regions could be to optimize the interaction between homologous α- and β-subunits in tetramer formation, or in the interaction of MoFe proteins with the corresponding Fe proteins during nitrogenase action.

We thank C. Felder and D. Littauer for help in translating the sequence alignment program from the VAX to the IBM computer. This work was supported in part by grants from the U.S.–Israel Binational Science Foundation, the H. Gutwirth Fund and the Leo and Julia Forccheimer Center for Molecular Genetics, The Weizmann Institute of Science.

REFERENCES

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

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

Cannon, M., Hill, S., Kavanaugh, E. & Cannon, F. (1985) Mol. Gen. Genet. **198**, 198–206

Colman, P. M., Varghese, J. N. & Laver, W. G. (1983) Nature (London) **303**, 41–44

Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1978) in Atlas of Protein Sequence and Structure (Dayhoff, M. O., ed.), vol. 55, pp. 345–352, National Biomedical Research Foundation, Washington

Golden, J. W., Robinson, S. J. & Haselkorn, R. (1985) Nature (London) **314**, 419–423

Hase, T., Wakabayashi, S., Nakano, T., Zumft, W. G. & Matsubara, H. (1984) FEBS Lett. **166**, 39–43

Hennecke, H., Kaluza, K., Thony, B., Fuhrmann, M., Ludwig, W. & Stackebrandt, E. (1985) Arch. Microbiol. **142**, 342–348

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

Lundell, D. J. & Howard, B. J. (1981) J. Biol. Chem. **256**, 6385–6391

MacNeil, T., MacNeil, D., Roberts, G. P., Supiano, M. A. & Brill, W. S. (1978) J. Bacteriol. **136**, 253–266

Mazur, B. J. & Chui, C. F. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 6782–6786

Merrick, M., Filser, M., Dixon, R., Elmerich, C., Sibold, L. & Houmard, S. (1980) J. Gen. Microbiol. **117**, 509–520

Mortenson, L. E. & Thorneley, R. N. F. (1979) Annu. Rev. Biochem. **48**, 387–418

Murata, M., Richardson, J. S. & Sussman, J. L. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 3073–3077

Needleman, S. B. & Wunsch, L. D. (1970) J. Mol. Biol. **48**, 443–453

Puhler, A. & Klipp, W. (1981) in Biology of Inorganic Nitrogen and Sulfur (Bothe, H. & Trebst, A., eds.), pp. 276–286, Springer-Verlag, Berlin

Reyes, A. A., Schold, M. & Wallace, R. B. (1982) Immunogenetics **16**, 1–9

Roberts, G. P., MacNeil, T., MacNeil, D. & Brill, W. J. (1978) J. Bacteriol. **136**, 267–279

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

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

Segal, D. M., Padlan, E. A., Cohen, G. H., Rudikoff, S., Potter, M. & Davies, D. R. (1974) Proc. Natl. Acad. Sci. U.S.A. **71**, 4298–4302

Smith, D. R. & Calvo, J. M. (1980) Nucleic Acids Res. **8**, 2255–2274

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

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

Wu, T. T. & Kabat, E. A. (1970) J. Exp. Med. **132**, 211–250

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

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

Received 9 March 1987; accepted 25 June 1987