Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 4723–4727, August 1983
Biochemistry

Sequence of the *nifD* gene coding for the α subunit of dinitrogenase from the cyanobacterium *Anabaena*

(nitrogen fixation/blue-green algae/protein sequence conservation)

PETER J. LAMMERS AND ROBERT HASELKORN

Department of Biophysics and Theoretical Biology, The University of Chicago, Chicago, Illinois 60637

Communicated by R. H. Burris, May 18, 1983

ABSTRACT The nucleotide sequence of *nifD*, the structural gene for the α subunit of dinitrogenase from *Anabaena* 7120, has been determined. The coding sequence contains 1,440 nucleotides, which predict an amino acid sequence of 480 residues and M_r of 54,283. The predicted sequence contains eight cysteines, of which five are conserved with respect to adjoining sequences and position relative to the α subunits of dinitrogenase from *Azotobacter*, *Clostridium*, and *Klebsiella*. Because there are also five conserved cysteines in the β subunit of *Anabaena* dinitrogenase [Mazur, B. J. & Chiu, C.-F. (1982) *Proc. Natl. Acad. Sci. USA* 79, 6782–6786], the number of cysteine residues participating as ligands to FeS clusters is likely to be 20 per $\alpha_2\beta_2$ tetramer. This number is sufficient to accommodate the known four Fe_4S_4 clusters, leaving at least four cysteines to be shared among the two FeMo cofactors and the more poorly characterized two-iron center. Although the α- and β-subunit gene sequences are not recognizably homologous, their secondary structures, predicted from the sequences, indicate similar domains around three of the conserved cysteine residues.

Biological nitrogen fixation is catalyzed by the enzyme nitrogenase. The larger of its two component proteins variously termed dinitrogenase, component I, or MoFe protein consists of two pairs of different polypeptides, termed α (M_r 56,000) and β (M_r 59,000) (1). In addition to the polypeptides, dinitrogenase (component I or MoFe protein) contains three distinct metal-ion environments revealed by Mössbauer and EPR studies (2). Of 30–32 total iron atoms, 12–16 are located in two identical MoFe cofactors ($MoFe_{6-8}S_8$ or $MoFe_{6-8}S_9$), which are EPR active ($S = 3/2$) spin-coupled centers easily dissociated from the protein (2–4). Sixteen iron atoms are protein-associated in four Fe_4S_4 clusters, designated P clusters. These differ from classical Fe_4S_4 clusters in that each is composed of two different diamagnetic environments designated D and Fe^{2+}, which are present in a ratio of 3:1. The remaining irons, which comprise $\approx$6% of the Mössbauer signal (2 Fe atoms), are assigned to a poorly characterized "S" center that may be an unusual Fe_2S_2 cluster (2). During a catalytic cycle electrons are transferred from a single Fe_4S_4 center in another protein (dinitrogenase reductase, Fe protein, or component II) to the P clusters of dinitrogenase in a one-electron reduction coupled to ATP hydrolysis. Electrons are then passed to one of the MoFe cofactors where N_2 (or an alternate substrate such as acetylene) is reduced (1).

All nitrogen-fixing organisms studied to date contain similarly constituted dinitrogenase and dinitrogenase reductase (1). Both *in vitro* complementation and spectroscopic methods reveal the conservation of structure among these components, but nothing is known about the detailed three-dimensional structure of the enzyme complex. The *Clostridium* dinitrogenase has been crystallized, but it has not yet been possible to define the liganding of the Fe-S or Mo-Fe centers to the protein (5).

The conserved protein structures of nitrogenase components reflect the conservation of nucleotide sequences coding for these components among the nitrogen-fixing bacteria (6, 7). In *Klebsiella pneumoniae*, the genes coding for dinitrogenase reductase and dinitrogenase α and β subunits are called *nifH*, *nifD*, and *nifK*, respectively (8). Recombinant DNA probes containing *Klebsiella nif* genes have been used to clone DNA fragments with related sequences from *Anabaena* (7, 9), *Azospirillum* (10), *Rhizobium* (11, 12), and *Rhodopseudomonas* (unpublished data).

Determination of the nucleotide sequences of cloned genes has made available the amino acid sequences of several nitrogenase polypeptides. The sequences of dinitrogenase reductase (*nifH*) from *Anabaena* (13), *Klebsiella* (14, 15), and *Rhizobium* (16) were determined this way and compared with the complete amino acid sequence of the corresponding protein from *Clostridium* (17) and the cysteine-containing peptides of the corresponding protein from *Azotobacter* (18). These comparisons defined regions of very strict homology, which include cysteine residues likely to be involved in liganding to the single Fe_4S_4 cluster of dinitrogenase reductase. However, there is no known homology between dinitrogenase reductase and any other iron-sulfur protein. For the larger dinitrogenase subunits, no complete amino acid sequences have been determined directly. Partial sequences for the *Clostridium* α subunit have been published (19), as have the sequences around all the cysteine residues in the α and β subunits of *Azotobacter* dinitrogenase (20). The NH_2-terminal 208 amino acids of the *Klebsiella* α subunit have been determined from the partial sequence of the *nifD* gene (14), whereas the complete amino acid sequence of the *Anabaena* β subunit was determined by sequence analysis of the *Anabaena nifK* gene (21).

We report here the nucleotide sequence of the *Anabaena nifD* gene and the amino acid sequence, obtained by translation of the gene sequence, of the dinitrogenase α subunit. *Anabaena* is now the only organism for which the primary structure of all three nitrogenase polypeptides is known.

MATERIALS AND METHODS

The *Anabaena nifD* gene is cut once by both *Eco*RI and *Hin*dIII (9). The NH_2-terminal part of the gene was cloned in a 1.8-kilobase-pair (kbp) *Hin*dIII fragment in plasmid pAn154.3 (13). The remaining part was cloned in a 2.4-kbp *Hin*dIII fragment in plasmid pAn256 (9). Both plasmids, derivatives of pBR322, were maintained in *Escherichia coli* HB101.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: kbp, kilobase pair.

Plasmid DNA was isolated as described (9). DNA sequence analysis was by the methods of Maxam and Gilbert (22), except for the piperidine cleavage step, which was modified as described by Smith and Calvo (23). All restriction sites were overlapped, except the *Hin*dIII site at the junction of the fragments in pAn154.3 and pAn256. The sequence there was determined, on both plasmids, up to the 5'-labeled nucleotides at the *Hin*dIII site by using 20% acrylamide gels (22). More than 96% of the coding sequence was determined on both strands of DNA.

RESULTS AND DISCUSSION

The *nifH* and *nifD* genes of *Anabaena* 7120 are located on adjacent *Hin*dIII fragments. Only 114 nucleotides separate the *nifH* termination codon from the *nifD* ATG start codon (Fig. 1), whereas the *Anabaena nifK* gene is located approximately 11 kbp downstream from the 3' end of *nifD* (9). The sequence of the *nifH* gene and the first 63 nucleotides of *nifD* has been published (13). The majority of the *nifD* gene is found on the 2.4-kbp *Hin*dIII fragment of pAn256. A detailed restriction map of pAn256 and the strategy used to determine the sequence of *nifD* is also presented in Fig. 1.

The complete sequence of the *Anabaena nifD* gene is shown as the noncoding strand in Fig. 2, along with the predicted amino acid sequence of the dinitrogenase α subunit. The *Anabaena nifD* coding sequence is 1,440 nucleotides long and terminates with a single ochre codon. This sequence defines a 480-residue polypeptide of M_r 54,283. This is in good agreement with values (M_r = 54,000–56,000) determined for the *Anabaena* MoFe protein subunits (24). Unlike the acidic iron-sulfur proteins such as rubredoxin and the ferredoxins (25), the α-subunit polypeptide is rich in basic amino acids, some of which are clustered, and is predicted to have a charge of +11 at neutral pH. Approximately 520 bases of 3'-flanking sequence are also presented. This region is A+T-rich relative to the *nifD* coding sequence and contains numerous nonsense codons. There are no open reading frames, starting with methionine, longer than 25 residues on either strand of this 520-base-pair region.

The designated start codon was established by comparison with the available NH_2-terminal amino acid sequences of the proteins from *Azotobacter* and *Clostridium* (refs. 19 and 20; Fig. 3). There are potential start codons, in phase, at positions −96 and 1. Both are preceded by reasonable ribosome-binding sequences, and no interceding termination codons are present. Because neither the NH_2-terminal amino acid sequence of the *Anabaena* dinitrogenase α subunit nor the 3' sequence of *Anabaena* 16S rRNA is known, the actual *in vivo* translation start site cannot be assigned with certainty. The possibility exists that the α subunit is translated with an NH_2-terminal leader peptide that is later processed to form the mature protein. Although the products of several *nif* genes in *Klebsiella* are required for dinitrogenase "maturation" (8), there is no evidence to show that maturation includes processing a leader peptide.

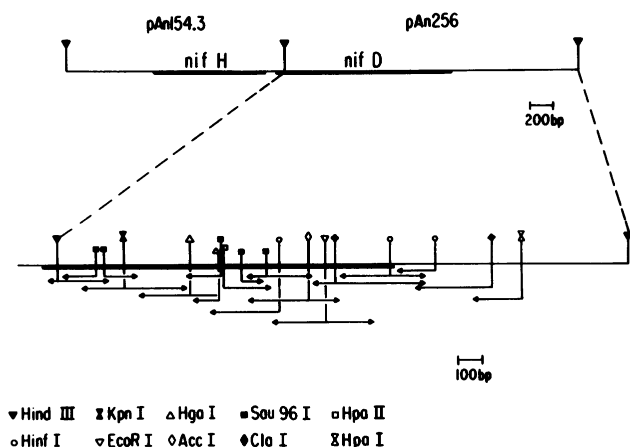

FIG. 1. Restriction map and sequence analysis strategy for the *nifD* gene of *Anabaena* 7120. The map and sequence of the *Hin*dIII fragment of pAn154.3 are given in ref. 13. The *Hin*dIII site at the junction of the two cloned fragments is at residues 65–70 of the *nifD* coding sequence. Fragments shown below were subjected to sequence analysis (see text). bp, Base pairs.

A comparison of the predicted amino acid sequence of the α subunit from *Anabaena* dinitrogenase with the partial sequences available from *Klebsiella* (14), *Clostridium* (19), and *Azotobacter* (20) is shown in Fig. 3. The overall homology of the α-subunit proteins is quite good. The *Anabaena* sequence is 66% homologous with the *Klebsiella* protein (136 of the available 205 residues are identical) and 47% homologous with the *Clostridium* protein (84 of 180 residues are identical). Five of eight cysteine residues in the *Anabaena* sequence show substantial homology with the other proteins. These cysteines occur at positions 64, 90, 156, 185, and 282. It is unlikely that the other three cysteines are conserved, at least between *Azotobacter* and *Anabaena*, because the comparison includes the sequences of all cysteine-containing tryptic peptides from *Azotobacter*. Of the five homologous cysteine peptides only the one containing C-156 exhibits homology with one of the known iron-sulfur proteins. The sequence Glu-Cys-Pro-Val/Ile-Gly-X-Ile is also found at residues 17–23 in the (anaerobic) bacterial ferredoxins (25). As noted by Lundell and Howard (20), this sequence may contain important secondary structure unique to Fe_4S_4 proteins. The fact that the other four homologous cysteine peptides are unique is not surprising, given the unusual spectral properties of the P clusters in dinitrogenase. The possibility of replacement of thiolate ligands by other nucleophiles or even 5-coordinate ligation has been suggested to explain the spectral data (2).

Because complete amino acid sequences are now available for the *Anabaena* α and β dinitrogenase subunits, it is possible to determine the total number of conserved cysteine residues in the $\alpha_2\beta_2$ enzyme by comparison with the cysteine-containing peptide sequences from the *Azotobacter* subunits (20) and partial α-subunit sequences from *Klebsiella* and *Clostridium* (14, 19). These residues are very likely to be involved as ligands to the iron-sulfur clusters of the MoFe protein. The *nifK* protein contains four conserved cysteines at positions 70, 95, 153, and 371. An additional peptide in *Anabaena nifK* differs only by a displacement of cysteine 111 by two residues from its position in *Azotobacter* (21). The *nifD* protein contains five conserved cysteines (Fig. 3). Thus, there are a total of 20 conserved cysteines per dinitrogenase tetramer, based on the available sequence data. If all four Fe_4S_4 centers are ligated totally by cysteine, they would require a total of 16, leaving 4 cysteines for bonding to the pair of MoFe centers or to the poorly characterized S center, which probably contains two Fe atoms, or to both. Neither the Fe_4S_4 centers nor the S center corresponds to known, typical centers *in situ* (26). Thus, it is possible that there are some non-thiolate bonds to the Fe-S centers, in which case a larger number of cysteines would be available for liganding to the MoFe cofactors.

Computer-assisted comparisons (27) of the nucleotide and predicted amino acid sequences of *Anabaena nifD* and *nifK* do not reveal any significant homologies, suggesting that these genes evolved separately or diverged very rapidly from a common ancestral gene. We have also compared the two subunits with respect to predicted secondary structures using the procedure of

```
-117                            -96
TAG GAGACACGGA GACAGGAGAT GAGGAGCAAT TCCTCTTCCC ACTCTCCCTT CCCGACTCCT CACTCTCCCA AATATACTTC
-34                                                                                         36
TATTCCCCCA TTCGTAAGAG TCACTGAGGC AGAT ATG ACA CCT CCT GAA AAC AAG AAT CTT GTA GAT GAA
                                      Met Thr Pro Pro Glu Asn Lys Asn Leu Val Asp Glu
                                                                                    99
AAT AAG GAA CTT ATT CAA GAA GTT CTG AAA GCT TAT CCC GAA AAA TCT CGC AAA AAG CGC GAA
Asn Lys Glu Leu Ile Gln Glu Val Leu Lys Ala Tyr Pro Glu Lys Ser Arg Lys Lys Arg Glu
                                                                                   162
AAA CAC CTC AAC GTC CAC GAA GAA AAC AAG TCT GAT TGC GGC GTT AAG TCT AAC ATC AAA TCC
Lys His Leu Asn Val His Glu Glu Asn Lys Ser Asp Cys Gly Val Lys Ser Asn Ile Lys Ser
                                                                                   225
GTT CCT GGT GTA ATG ACC GCC CGT GGT TGT GCT TAT GCA GGT TCT AAG GGT GTG GTT TGG GGT
Val Pro Gly Val Met Thr Ala Arg Gly Cys Ala Tyr Ala Gly Ser Lys Gly Val Val Trp Gly
                                                                                   288
CCT ATT AAG GAC ATG ATC CAC ATC AGC CAC GGG CCT GTA GGT TGC GGT TAC TGG TCT TGG TCT
Pro Ile Lys Asp Met Ile His Ile Ser His Gly Pro Val Gly Cys Gly Tyr Trp Ser Trp Ser
                                                                                   351
GGT CGT CGT AAC TAC TAC GTT GGT GTA ACT GGT ATC AAC TCT TTC GGT ACC ATG CAC TTC ACC
Gly Arg Arg Asn Tyr Tyr Val Gly Val Thr Gly Ile Asn Ser Phe Gly Thr Met His Phe Thr
                                                                                   414
TCA GAC TTC CAA GAA CGT GAC ATC GTG TTC GGT GGT GAC AAA AAA CTC ACT AAA CTC ATT GAA
Ser Asp Phe Gln Glu Arg Asp Ile Val Phe Gly Gly Asp Lys Lys Leu Thr Lys Leu Ile Glu
                                                                                   477
GAA CTC GAT GTT CTT TTC CCT CTC AAC CGT GGT GTT TCC ATT CAA TCT GAA TGT CCC ATT GGA
Glu Leu Asp Val Leu Phe Pro Leu Asn Arg Gly Val Ser Ile Gln Ser Glu Cys Pro Ile Gly
                                                                                   540
TCT ATT GGG GAT GAC ATC GAA GCT GTT GCT AAG AAA ACT TCT AAG CAA ATT GGT AAG CCT GTT
Ser Ile Gly Asp Asp Ile Glu Ala Val Ala Lys Lys Thr Ser Lys Gln Ile Gly Lys Pro Val
                                                                                   603
GTA CCC TTA CGT TGC GAA GGT TTC CGT GGT GTG TCT CAG TCC TTA GGA CAC CAC ATC GCT AAC
Val Pro Leu Arg Cys Glu Gly Phe Arg Gly Val Ser Gln Ser Leu Gly His His Ile Ala Asn
                                                                                   666
GAC GCT ATC CGT GAC TGG ATT TTC CCA GAA TAC GAC AAG CTC AAG AAA GAA ACC AGA CTT GAC
Asp Ala Ile Arg Asp Trp Ile Phe Pro Glu Tyr Asp Lys Leu Lys Lys Glu Thr Arg Leu Asp
                                                                                   729
TTC GAG CCA AGC CCC TAT GAT GTA GCT CTA ATC GGT GAC TAC AAC ATC GGT GGT GAC GCT TGG
Phe Glu Pro Ser Pro Tyr Asp Val Ala Leu Ile Gly Asp Tyr Asn Ile Gly Gly Asp Ala Trp
                                                                                   792
GCC AGC CGG ATG CTG TTG GAA GAA ATG GGC TTA CGT GTA GTA GCT CAG TGG TCT GGT GAT GGT
Ala Ser Arg Met Leu Leu Glu Glu Met Gly Leu Arg Val Val Ala Gln Trp Ser Gly Asp Gly
                                                                                   855
ACA CTC AAC GAG TTG ATC CAA GGC CCT GCT GCT AAG TTA GTC CTC ATC CAC TGC TAC CGT TCT
Thr Leu Asn Glu Leu Ile Gln Gly Pro Ala Ala Lys Leu Val Leu Ile His Cys Tyr Arg Ser
                                                                                   918
ATG AAC TAC ATC TGC CGT AGT TTG GAA GAA CAA TAC GGT ATG CCT TGG ATG GAG TTC AAC TTC
Met Asn Tyr Ile Cys Arg Ser Leu Glu Glu Gln Tyr Gly Met Pro Trp Met Glu Phe Asn Phe
                                                                                   981
TTC GGC CCC ACC AAG ATT GCT GCT TCT TTA CGT GAA ATC GCA GCT AAG TTT GAT TCT AAG ATT
Phe Gly Pro Thr Lys Ile Ala Ala Ser Leu Arg Glu Ile Ala Ala Lys Phe Asp Ser Lys Ile
                                                                                  1044
CAA GAA AAC GCT GAG AAG GTA ATT GCT AAG TAC ACA CCA GTA ATG AAT GCT GTA CTA GAT AAA
Gln Glu Asn Ala Glu Lys Val Ile Ala Lys Tyr Thr Pro Val Met Asn Ala Val Leu Asp Lys
                                                                                  1107
TAC CGC CCT CGC TTG GAA GGT AAC ACC GTA ATG TTG TAC GTA GGT GGT CTA CGT CCT CGC CAC
Tyr Arg Pro Arg Leu Glu Gly Asn Thr Val Met Leu Tyr Val Gly Gly Leu Arg Pro Arg His
                                                                                  1170
GTT GTT CCC GCT TTT GAA GAC CTG GGT ATC AAA GTA GTT GGT ACT GGC TAT GAA TTC GCT CAC
Val Val Pro Ala Phe Glu Asp Leu Gly Ile Lys Val Val Gly Thr Gly Tyr Glu Phe Ala His
                                                                                  1233
AAT GAC GAT TAC AAA CGT ACC ACC CAC TAC ATC GAT AAC GCC ACC ATC ATT TAC GAT GAC GTT
Asn Asp Asp Tyr Lys Arg Thr Thr His Tyr Ile Asp Asn Ala Thr Ile Ile Tyr Asp Asp Val
                                                                                  1296
ACC GCC TAC GAA TTT GAA GAG TTC GTA AAA GCT AAG AAG CCT GAT TTA ATT GCT TCT GGT ATT
Thr Ala Tyr Glu Phe Glu Glu Phe Val Lys Ala Lys Lys Pro Asp Leu Ile Ala Ser Gly Ile
                                                                                  1359
AAA GAG AAG TAC GTC TTC CAA AAG ATG GGT CTT CCC TTC CGT CAA ATG CAC TCT TGG GAT TAC
Lys Glu Lys Tyr Val Phe Gln Lys Met Gly Leu Pro Phe Arg Gln Met His Ser Trp Asp Tyr
                                                                                  1422
TCC GAA CTT GGC GAC GGG GTG CAG ATG TCA GAT GAG GTA AGG TTT TTT TGT GAG GGG AGA AAA
Ser Glu Leu Gly Asp Gly Val Gln Met Ser Asp Glu Val Arg Phe Phe Cys Glu Gly Arg Lys
                    1440                                                          1493
AAG AGT CTA TTT TTA GCC TAA   ATACAGGTTG TAGGGTTATC TGGGAACAAT ACATTTTTGC GGCTTAAGAC
Lys Ser Leu Phe Leu Ala End
                                                                                  1573
AATATAACTG TTATTGGGTC TGGTTTTTGG GTTTTGGTTA GATTTGGGGT ATATAAATAA AAGACTATTT AATTGTTATT
                                                                                  1653
ATTTGATTAA TGTGGGTCGG TTAGGCGATC GCTTTGAATC ACTAAGACTA GGAAGTATGC GATGAGGTAT GCGACAGAAA
                                                                                  1733
ATCTACATTC CAGCTACTAT ATAGAAAATA ATTAAAGTTA TTCAAGGATA GAAAAGAAAT GTAAAATTAG AGGATGCTAA
                                                                                  1813
AAACATCCTC TACAATAGAG TTGTGAGTCT CTTCAACACA ACTATTGATA TAAATTGCGA ATAATAAACT TAAGATACTG
                                                                                  1893
AGCTTTGTTA GTCAATATCA AAATCCATGA ATATCGATTT TTAGTTCACC TTTATGATAC TTGTATGACA TGAAGGACAA
                                                                        1963
AGAGTCATCA TGTTTTGAAT ATCAGCAGAA CCTCCTTCAC TATAGGGAAC TAAGTGATGA CCATTAGCT
```

FIG. 2. Nucleotide sequence of the *nifD* gene of *Anabaena* 7120. The sequence begins with the termination codon of the *nifH* gene (13) at position −117 with respect to the probable start of the protein sequence. There is an alternative site for translation at −96 (the underlined ATG), which, if used, would produce a protein 32 amino acids longer. This NH_2-terminal extension is not seen in mature dinitrogenase α subunit from *Clostridium* (19) or *Azotobacter* (20).

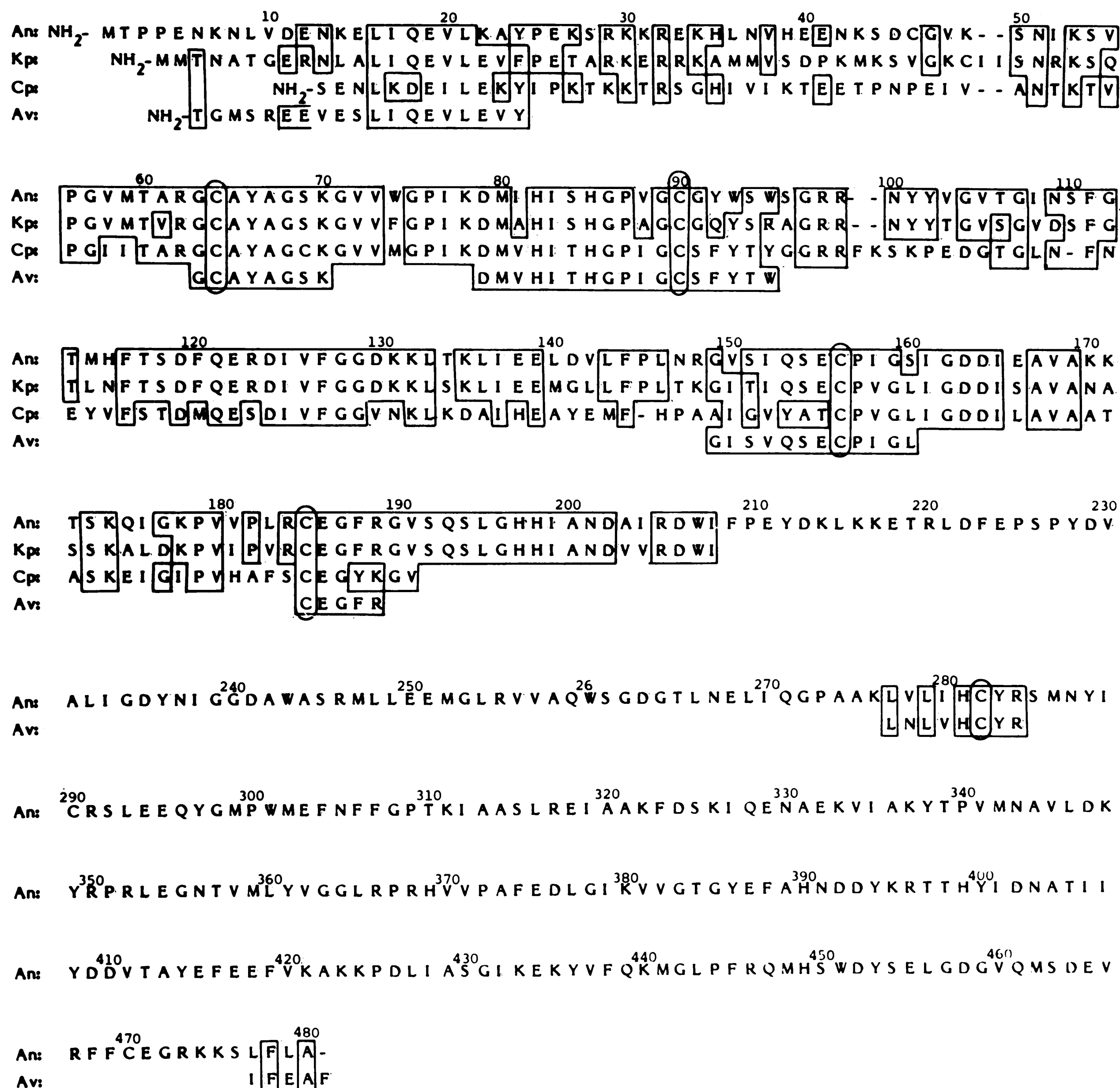

FIG. 3. Comparison of the predicted amino acid sequence of the α subunit of dinitrogenase from *Anabaena* 7120 (An) with partial sequences of the corresponding proteins from *Klebsiella* (Kp), *Clostridium* (Cp), and *Azotobacter* (Av). Data for the other proteins are from refs. 14, 19, and 20, respectively. Conserved cysteine residues are circled; other conserved regions are boxed.

Chou and Fasman (28). The α subunit is predicted to contain 30% α helix, 31% β sheet with 25 β turns, whereas the β subunit should contain 25% α helix, 28% β sheet, and 28 β turns. A plot of the regions of α helix, β sheet, β turns, and coil in each subunit is presented in Fig. 4, along with the positions of conserved cysteine residues. The only similarity appears to be the positioning of three conserved cysteines (α-Cys-64, -90, -156 and β-Cys-70, -95, -153) in the NH_2-terminal third of each subunit. In addition, α-Cys-64 and β-Cys-70 occur in random coil and α-Cys-90 and β-Cys-95 both occur in β regions. Both α-Cys-156 and β-Cys-153 are preceded by a coil-α-β configuration. However, α-Cys-156 occurs in a region predicted to contain random coil, whereas β-Cys-153 occurs in a short stretch of α helix. The significance of these relationships will have to await more detailed tertiary structure information, but they suggest that dinitrogenase subunits fall into the second category of related proteins described by Richardson *et al.* (29), in which similar folding patterns exist within proteins having no clear sequence homology. Low-resolution structural homology between the α and β subunits of dinitrogenase from *Clostridium* has already been suggested on the basis of secondary peaks in the rotation function calculated from x-ray crystallographic data (5).

We are indebted to Drs. Stephanie Curtis, Steven Robinson, and Nilgun Tumer for their assistance and critical reading of this manuscript, to Victoria DeRose for technical assistance, and to Crystal Sherman for preparing Figs. 2 and 3. We are also grateful to Prof. F. J. Kezdy for the use of his program for predicting protein secondary structure. This work was supported by Grant GM 21823 from the National Institutes of Health and Grant 82-CRCR-1-1056 from the U.S. Department of Agriculture, Science and Education Administration, through the Competitive Research Grants Office. P.J.L. is a postdoctoral trainee of the National Institute of Child Health and Human Development (HD 07136).

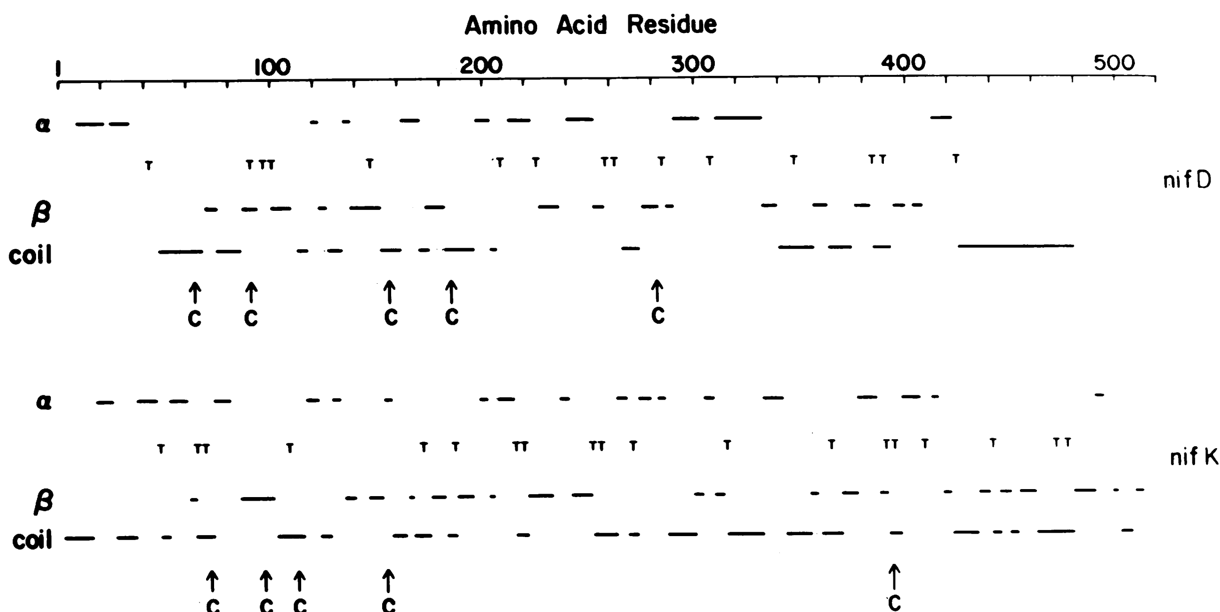

FIG. 4. Comparison of predicted protein conformations (28) of the α and β subunits of *Anabaena* 7120 dinitrogenase, based on the amino acid sequences of Fig. 3 and ref. 21. "T" refers to predicted β turns and "C" refers to the positions of conserved cysteine residues.

1. Mortenson, L. E. & Thorneley, R. N. F. (1979) *Annu. Rev. Biochem.* **48**, 387–418.
2. Huynh, B. H., Henzl, M. T., Christner, J. A., Zimmerman, R., Orme-Johnson, W. H. & Münck, E. (1980) *Biochim. Biophys. Acta* **623**, 124–138.
3. Nelson, M. J., Levy, M. A. & Orme-Johnson, W. H. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 147–150.
4. Yang, S.-S., Pan, W.-H., Friesen, G. D., Burgess, B. K., Corbin, J. L., Stiefel, E. I. & Newton, W. E. (1982) *J. Biol. Chem.* **257**, 8042–8048.
5. Yamane, T., Weininger, M. S., Mortenson, L. E. & Rossmann, M. G. (1982) *J. Biol. Chem.* **257**, 1221–1223.
6. Ruvkun, G. & Ausubel, F. M. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 191–195.
7. Mazur, B. J., Rice, D. & Haselkorn, R. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 186–190.
8. Roberts, G. P. & Brill, W. J. (1981) *Annu. Rev. Microbiol.* **35**, 207–235.
9. Rice, D., Mazur, B. J. & Haselkorn, R. (1982) *J. Biol. Chem.* **257**, 13157–13163.
10. Quiviger, B., Franche, C., Lutfalla, G., Rice, D., Haselkorn, R. & Elmerich, C. (1982) *Biochimie* **64**, 495–502.
11. Ruvkun, G. B., Sundaresan, V. & Ausubel, F. M. (1982) *Cell* **29**, 551–559.
12. Hennecke, H. (1981) *Nature (London)* **291**, 354–355.
13. Mevarech, M., Rice, D. & Haselkorn, R. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6476–6480.
14. Scott, K. F., Rolfe, B. G. & Shine, J. (1981) *J. Mol. Appl. Genet.* **1**, 71–81.
15. Sundaresan, V. & Ausubel, F. M. (1981) *J. Biol. Chem.* **256**, 2808–2812.
16. Torok, I. & Kondorosi, A. (1981) *Nucleic Acids Res.* **9**, 5711–5723.
17. Tanaka, M., Hainu, M., Yasunobu, K. T. & Mortenson, L. E. (1977) *J. Biol. Chem.* **252**, 7093–7100.
18. Hausinger, R. D. & Howard, J. B. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3826–3830.
19. Hase, T., Nakano, T., Matsubara, H. & Zumft, W. G. (1981) *J. Biochem. (Tokyo)* **90**, 295–298.
20. Lundell, D. J. & Howard, J. B. (1981) *J. Biol. Chem.* **256**, 6385–6391.
21. Mazur, B. J. & Chiu, C.-F. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6782–6786.
22. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
23. Smith, D. R. & Calvo, J. M. (1980) *Nucleic Acids Res.* **8**, 2255–2275.
24. Hallenbeck, P. C., Kostel, P. J. & Benemann, J. R. (1979) *Eur. J. Biochem.* **98**, 275–284.
25. Yasunobu, K. T. & Tanaka, M. (1973) in *Iron-Sulfur Proteins*, ed. Lovenberg, W. (Academic, New York), Vol. 2, pp. 29–123.
26. Stephens, P. J., McKenna, C. E., Smith, B. E., Nguyen, H. T., McKenna, M.-C., Thomson, A. J., Devlin, F. & Jones, J. B. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 2585–2589.
27. Queen, C. L. & Korn, L. J. (1980) *Methods Enzymol.* **65**, 595–609.
28. Chou, P. Y. & Fasman, G. D. (1978) *Annu. Rev. Biochem.* **47**, 251–276.
29. Richardson, J. S., Richardson, D. C. & Thomas, K. A. (1976) *J. Mol. Biol.* **102**, 221–235.