Structural homologies between the amino acid sequence of *Clostridium pasteurianum* MoFe protein and the DNA sequences of *nifD* and *K* genes of phylogenetically diverse bacteria

Toshiharu Hase, Sadao Wakabayashi, Tohru Nakano, Walter G. Zumft* and Hiroshi Matsubara

Department of Biology, Faculty of Science, Osaka University, Toyonaka, Osaka 560, Japan and **Lehrstuhl für Mikrobiologie, Universität Karlsruhe, Kaiserstrasse 12, D-7500 Karlsruhe 1, FRG*

Received 3 November 1983

The complete amino acid sequence of the larger (α-) subunit and about 70% of the total sequence of the smaller (β-) subunit of the MoFe protein from *Clostridium pasteurianum* was determined by analyses of peptides derived from BrCN cleavage and by digestions with trypsin, staphylococcal protease and lysylendo-peptidase of the separated subunits. The α-subunit has 529 amino acid residues, giving an M_r value of 58774. This is the first complete sequence for the α-subunit of an isolated MoFe protein. In comparing the sequences of both subunits to those from other sources, 5 out of 9 cysteines in the α-subunit and 3 out of 6 in the β-subunit are invariant, thus suggesting a function as ligands to FeS and MoFeS clusters in the MoFe protein. All of these cysteines are located in the amino terminal halves of both subunits.

Amino acid sequence *Nitrogenase* *MoFe protein* *Sequence homology* Clostridium

1. INTRODUCTION

Biological nitrogen fixation, found in a wide variety of prokaryotes, is catalyzed by an enzyme complex composed of the MoFe protein (dinitrogenase) and the Fe protein (dinitrogenase reductase). The MoFe protein is an $\alpha_2\beta_2$ tetramer of M_r 220000–245000, and contains 30–32 Fe atoms, 2 Mo atoms and inorganic sulfur [1]. Genetic analysis of nitrogenase in *Klebsiella pneumoniae* has revealed that *nifH, nifD,* and *nifK* are the structural genes for the Fe protein and the two subunits of the MoFe protein, respectively [2]. We have been studying the amino acid sequences of both subunits, the larger (α-) and the smaller (β-) one of the MoFe protein from *Clostridium pasteurianum* and have reported that the N-terminal 179 amino acid sequence of the α-subunit is homologous with those of the *nifD* gene products of other organisms [3]. Recently, the amino acid sequence of the *Anabaena* β-subunit was deduced by sequencing the *Anabaena nifK* gene [4]. Those of *Anabaena* [5] and *Rhizobium Parasponia* [6] α-subunits were also derived from *nifD* genes, while no complete amino acid sequence has been determined by direct analysis of the isolated proteins.

We report here the complete sequence of the α-subunit of clostridial MoFe protein together with the partial structure of the β-subunit, and discuss structural characteristics found by the comparison with other MoFe proteins.

Abbreviations: N-, amino-: C-, carboxyl-: Cm-, *S*-carboxymethyl-: HPLC, high performance liquid chromatography; TLC, thin-layer chromatography

Published by Elsevier Science Publishers B.V.
00145793/84/$3.00

2. MATERIALS AND METHODS

The MoFe protein from *C. pasteurianum* W5 was purified as in [7]. The *S*-carboxymethylated (Cm-) MoFe protein was separated into the two subunits on a DE-52 column after maleylation or succinylation [3], or without any modification [8].

The Cm- and maleylated or succinylated α- and β-subunits were cleaved separately with BrCN, trypsin and staphylococcal protease and the resulting peptides were fractionated as in [3]. The Cm-α-subunit was digested with lysylendopeptidase at substrate:enzyme = 100:1 (w/w) in 0.1 M Tris-HCl (pH 8.0) at 40°C overnight and the digest was separated by HPLC under the conditions in [9]. The amino acid compositions of peptides were determined as usual and the sequences were obtained by automatic solid-phase Edman degradation [10] and/or a manual procedure [11]. Phenylthiohydantoin derivatives were identified by HPLC and TLC as in [3,9].

3. RESULTS AND DISCUSSION

A summary of the sequence studies of the α-subunit is given in fig.1. Peptides were named ac-

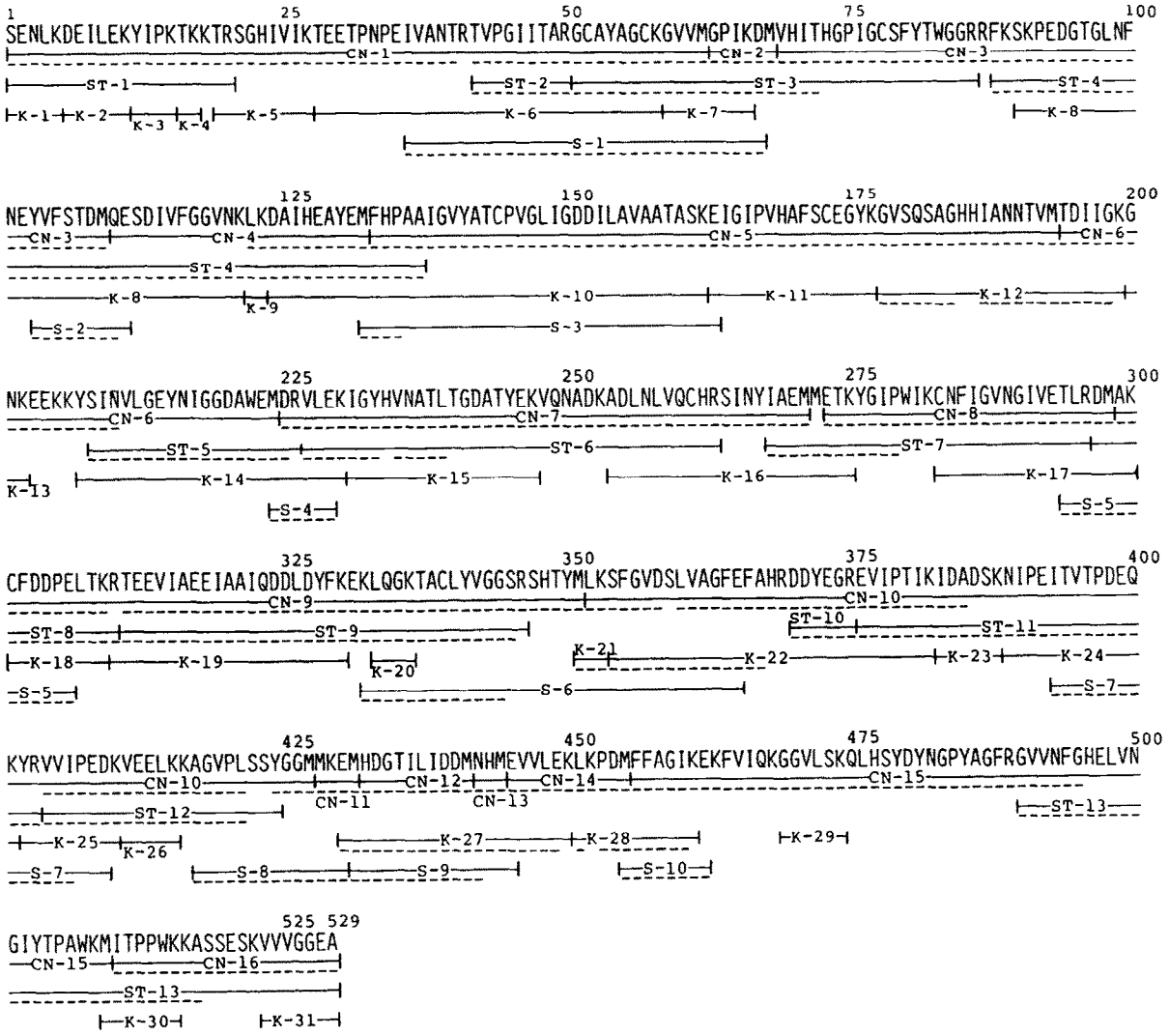

Fig.1. A summary of the sequence studies of the α-subunit. Nomenclature of peptides is explained in the text. Dashed lines below the sequences of peptides show the residues that were identified.

cording to the cleavage method used for their generation: CN- (BrCN peptides of Cm-maleylated protein), ST- and S- (tryptic and staphylococcal protease peptides of Cm-succinylated protein, respectively), and K- (lysylendopeptidase peptides of Cm-protein). Sequence analyses were performed mainly on CN- and ST-peptides. Long peptides, CN-1, CN-3, CN-5, CN-7, CN-9, CN-10 and CN-15, and ST-4, ST-9, ST-11 and ST-13 were further digested with staphylococcal protease or trypsin in order to obtain subfragments which were used for completing the sequences of the original peptides. Peptides CN-1 to CN-5, CN-6 to CN-9 and CN-15 to CN-16 could be easily linked by comparing the sequences of these peptides with ST-peptides. The N-terminal sequence of CN-1 was identical with the previously reported N-terminal sequence of the α-subunit [12]. The carboxyl(C-) terminus of CN-16 was alanine, suggesting from the specificity of the cleavage method that this peptide should be located at the C-terminus of the original protein. This agreed with the fact that one of the C-termini of the two subunits of clostridial MoFe protein had to be alanine [13]. S- and K-peptides provided information on non-overlapping regions of CN- and ST-peptides and also supplemental sequences for unidentified residues of these peptides. K-12 pro-

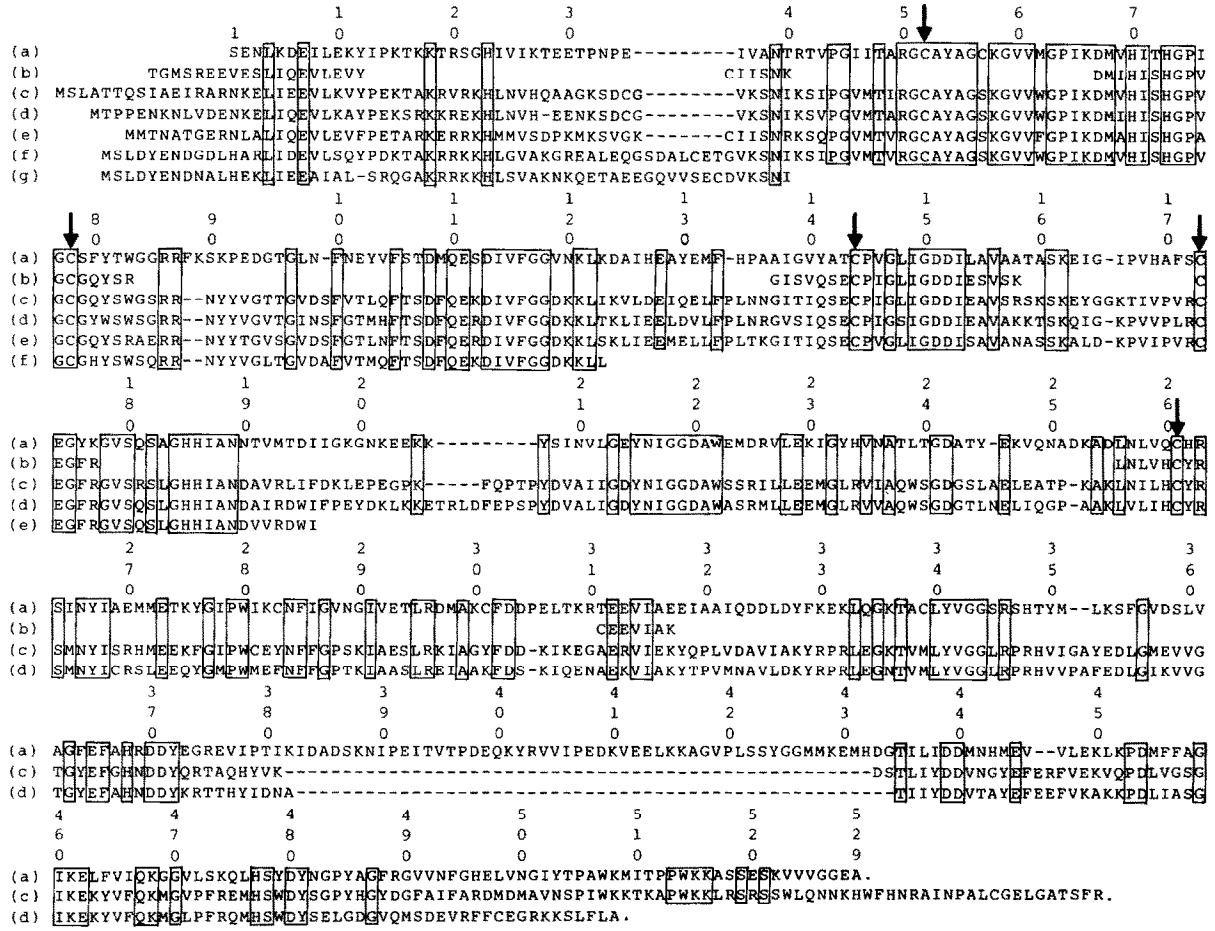

Fig.2. A comparison of sequences of the α-subunits of MoFe proteins. (a) α-Subunit from *C. pasteurianum* (here), (b) tryptic peptides containing cysteine residues of *A. vinelandii* MoFe protein [18], (c-g) the α-subunits predicted from *nifD* genes from *Rhizobium Parasponia* [6], *Anabaena* [5], *Klebsiella pneumoniae* [19], *Rhizobium trifolii* [20] and *R. meliloti* [21], respectively. Numbering refers to the sequence of clostridial sequence. Residues identical in all sequences are enclosed in boxes, and invariant cysteines are marked with arrows.

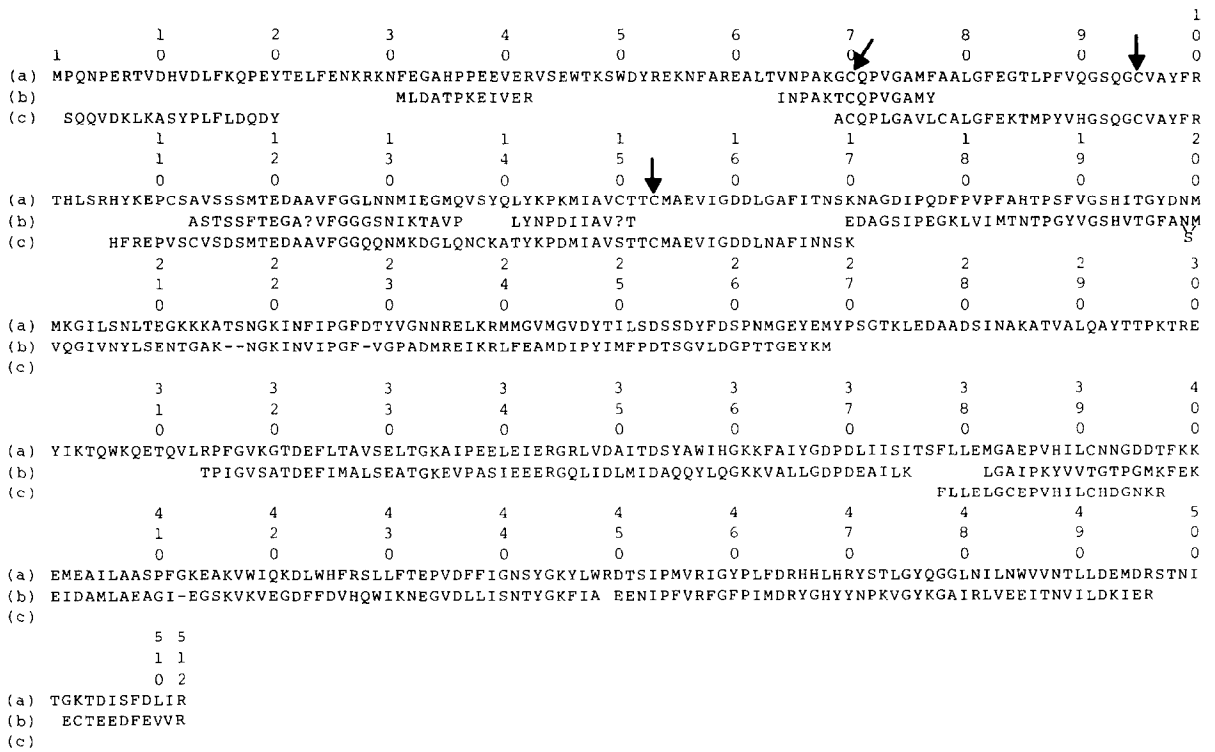

Fig.3. A comparison of sequences of the β-subunits of MoFe proteins. (a) The β-subunit predicted from *nifK* gene of *Anabaena* [4], (b) the β-subunit from *C. pasteurianym* (here) (c) tryptic peptides containing cysteine residues of *A. vinelandii* MoFe protein [18]. Numbering refers to the sequence of *Anabaena* protein. Peptides in (b) and (c) were tentatively aligned to obtain a high homology with *Anabaena* sequence. Invariant cysteines are marked with arrows.

vided the overlap between CN-5 and CN-6, and S-8, S-9, S-10, K-27 and K-28 overlaps from CN-10 to CN-15. Although K-21 provided only 3 overlapping residues to link peptides CN-9 and CN-10, the correctness of this alignment was confirmed by the partial sequence and amino acid composition of S-6. Thus, all CN-peptides could be aligned correctly from CN-1 to CN-16, completing the total sequence of the α-subunit as shown in fig.1. In [3], the residue at position 41 was not identified by direct analysis of its phenylthiohydantoin derivative of amino acid and was tentatively assigned as lysine. The sequence of S-1, however, clearly showed this residue to be arginine.

The calculated amino acid composition of the α-subunit from the complete sequence (numbers in parentheses) agreed well with that obtained from a direct analysis of the original protein: Asx 56.8(54), Thr 29.3(30), Ser 24.7(24), Glx 55.3((51), Pro 21.4(22), Gly 50.9(51), Ala 37.0(36), Cys 8.39(9), Val 42.1(42), Met 16.6(17), Ile 40.8(44), Leu 28.9(26), Tyr 21.8(23), Phe 18.4(18), Lys 41.9(46), His 15.4(16), Arg 14.9(14) and Trp 2.79(5).

A comparison of the sequence of the clostridial α-subunit with those of other organisms is shown in fig.2. All sequences except those of *Clostridium* and *Azotobacter* proteins are deduced from DNA sequences of *nifD* genes. Three long deletions or insertions are introduced to obtain high homology: deletions between residues 35 and 36, according to the numbering of the clostridial sequence and between residues 206 and 207, and an insertion from residues 382 to 432. The overall homology was calculated among the α-subunits whose complete sequences were available. The clostridial sequence is identical by 40 and 41% with that of *Rhizobium Parasponia* and *Anabaena*, respectively, whereas the latter two are identical to each other by 68%. The clostridial α-subunit also shows relatively low homology to the α-subunits from other organisms. This sequence divergence is also found among the

Fe proteins of *Clostridium* and other organisms [14,15]. At present, we have obtained about 70% of the total sequence of the β-subunit of clostridial MoFe protein and a comparison of this sequence with that of *Anabaena* and *Azotobacter* is shown in fig.3. The N-terminus of the β-subunit of the clostridial MoFe protein corresponds to residue 31 in the sequence of *Anabaena* β-subunit. Therefore, the difference in the molecular mass of the β-subunit from these two organisms is mainly attributed to this N-terminal extension in *Anabaena* β-subunit.

The MoFe protein contains four [4Fe-4S] clusters, possibly two MoFe cofactors and a minor, unidentified species of iron [1]. In the α-subunit of the clostridial MoFe protein, 5 out of a total of 9 cysteine residues at positions 52, 78, 144, 183 and 261 are invariant as shown in fig.2, and the sequences surrounding these cysteines show more than 70% conservation. Thus, several or all of these residues are expected to function as thiol ligands to the clusters. In the β-subunit, 3 cysteine residues are invariant as shown in fig.3. Therefore, 16 cysteine residues per $\alpha_2\beta_2$ are conservative among all MoFe proteins so far studied and they are distributed in the N-terminal regions of both subunits. The possibility that nucleophiles other than thiolate group function as ligands has been suggested [16]. In the α-subunit, there are several invariant amino acids with a nucleophilic side chain, such as tyrosine, phenyl-alanine, glutamine or asparagine. It is notable that Tyr-54, Phe-105, Gln-110, Phe-116, Asn-189, Tyr-215 and Tyr-342 are found in highly conserved regions.

In vitro complementation experiments of the nitrogenase components show limited cross-reactivity for the clostridial MoFe protein with almost all Fe proteins from other organisms [17]. We find several characteristic structures in the clostridial α-subunit as shown in fig.2: a long insertion from residues 382 to 432 and unique sequences from residues 88 to 95 and residues 134 to 143. These features may provide the structural basis for explaining the weak cross-reactivity of the clostridial MoFe protein.

ACKNOWLEDGEMENTS

We thank Dr J. Shine for providing the unpublished DNA sequences and Dr G. Schatz for helpful comments and criticism. This work was supported by the Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan and by the Deutsche Forschungs-Gemeinschaft.

REFERENCES

[1] Zumft, W.G. (1981) in: Biology of Inorganic Nitrogen and Sulfur (Bothe, H. and Trebst, A. eds) pp. 116–140, Springer, Berlin.
[2] Roberts, G.P., MacNeil, T., MacNeil, D. and Brill, W.J. (1978) J. Bacteriol. 136, 267–279.
[3] Hase, T., Nakano, T., Matsubara, H. and Zumft, W.G. (1981) J. Biochem. (Tokyo) 90, 295–298.
[4] Mazur, B.J. and Chui, C.F. (1982) Proc. Natl. Acad. Sci. USA 79, 6782–6786.
[5] Lammers, P.J. and Haselkorn, R. (1983) Proc. Natl. Acad. Sci. USA 80, 4723–4727.
[6] Weinman, J., Fellows, F., Gresshoff, P., Shine, J. and Scott, K. (1983) in: Abstracts for the 5th International Symposium on Nitrogen Fixation.
[7] Zumft, W.G. (1978) Eur. J. Biochem. 91, 345–350.
[8] Lundell, D.J. and Howard, J.B. (1979) J. Biol. Chem. 253, 3422–3426.
[9] Takahashi, Y., Hase, T., Wada, K. and Matsubara, H. (1983) Plant Cell Physiol. 24, 189–198.
[10] Laursen, R.A. (1971) Eur. J. Biochem. 20, 89–102.
[11] Blombäck, B., Blombäck, M., Edman, P. and Hessel, B. (1966) Biochim. Biophys. Acta 115, 371–396.
[12] Zumft, W.G., Hase, T. and Matsubara, H. (1980) in: Molybdenum Chemistry of Biological Significance (Newton, W.E. and Otsuka, S. eds) pp. 59–72, Plenum, New York.
[13] Chen, J.S., Multani, J.S. and Mortenson, L.E. (1973) Biochim. Biophys. Acta 310, 51–59.
[14] Hausinger, R.P. and Howard, J.B. (1982) J. Biol. Chem. 257, 2483–2490.
[15] Scott, K.F., Rolfe, B.G. and Shine, J. (1983) Nucleic Acids Res., in press.
[16] Huynh, B.H., Henzl, M.T., Christner, J.A., Zimmerman, R., Orme-Johnson, W.H. and Münck, E. (1980) Biochim. Biophys. Acta 623, 124–138.
[17] Emerich, D.W. and Burris, R.H. (1978) J. Bacteriol. 134, 936–943.
[18] Lundell, D.J. and Howard, J.B. (1981) J. Biol. Chem. 256, 6385–6391.
[19] Scott, K.F., Rolfe, B.G. and Shine, J. (1981) J. Mol. Appl. Genet. 1, 71–81.
[20] Scott, K.F., Rolfe, B.G. and Shine, J. (1983) Nucleic Acids Res., in press.
[21] Török, I. and Kondorosi, A. (1981) Nucleic Acids Res. 9, 5711–5723.