JOURNAL OF BACTERIOLOGY, Sept. 1991, p. 5457–5469
0021-9193/91/175457-13$02.00/0

Vol. 173, No. 17

Nucleotide Sequence and Genetic Analysis of the *Azotobacter chroococcum nifUSVWZM* Gene Cluster, Including a New Gene (*nifP*) Which Encodes a Serine Acetyltransferase

DAVID J. EVANS,[1]† ROBERT JONES,[2] PAUL R. WOODLEY,[1] J. RYMON WILBORN,[3] AND ROBERT L. ROBSON[3]*

AFRC-IPSR Unit of Nitrogen Fixation, University of Sussex, Brighton BN1 9RQ, United Kingdom[1]; Thinking Machines Corporation, Cambridge, Massachusetts 02142-1214[2]; and Department of Biochemistry, University of Georgia, Athens, Georgia 30602[3]

Received 4 February 1991/Accepted 25 June 1991

Nucleotide sequence was obtained for a region of 7,099 bp spanning the *nifU*, *nifS*, *nifV*, *nifW*, *nifZ*, and *nifM* genes from *Azotobacter chroococcum*. Chromosomal mutations constructed at several sites within the locus confirmed a requirement for this region for expression of the molybdenum nitrogenase in this organism. The genes are tightly clustered and ordered as in *Klebsiella pneumoniae* except for two additional open reading frames (ORFs) between *nifV* and *nifW*. The arrangement of genes in *A. chroococcum* closely matches that described for *Azotobacter vinelandii*. The polypeptide encoded by ORF4 immediately downstream from *nifV* is 41% identical over 186 amino acids to the product of the *cysE* gene from *Escherichia coli*, which encodes serine acetyltransferase (SAT), a key enzyme in cysteine biosynthesis. Plasmids which potentially express ORF4 complemented *E. coli* JM39, a cysteine auxotroph which lacks SAT. SAT activity was detected in crude extracts of one such complemented strain. A strain of *A. chroococcum* carrying a chromosomal disruption of ORF4 grew normally with ammonium as the N source but more slowly than the parental strain when N_2 was the sole N source. These data suggest that ORF4 encodes a *nif*-specific SAT required for optimizing expression of nitrogenase activity. ORF4 was assigned the name *nifP*. *nifP* may be required to boost rates of synthesis or intracellular concentrations of cysteine or methionine. Sequence identity between *nifV* and *leuA* gene products suggests that *nifV* may catalyze a condensation reaction analogous to that carried out by isopropylmalate synthase (LEUA) but in which acetyl coenzyme and α-ketoglutarate are substrates for the formation of homocitrate, the proposed product of NIFV activity.

Members of the genus *Azotobacter* are free-living, obligately aerobic, heterotrophic diazotrophs found in soil and freshwater environments. They are especially interesting from the point of view of N_2 fixation because, depending on the species, they possess two or three genetically distinct nitrogenase systems whose expression is determined by metal availability (8). The genetics of N_2 fixation in these organisms is complicated not only by their possession of alternative nitrogenases but also by their tolerance of O_2 when fixing N_2 (61). Our studies have concentrated on *Azotobacter chroococcum* MCD1, a derivative of ATCC 4412. This organism makes a typical molybdenum-dependent nitrogenase when molybdenum is available (77) which is replaced in the absence of molybdenum and presence of vanadium by a vanadium-dependent nitrogenase (60).

Many of the genes required for N_2 fixation were first defined in the facultative anaerobe *Klebsiella pneumoniae* M5A1, which possesses a molybdenum nitrogenase system only and fixes N_2 in anaerobic or, at best, microaerobic environments. In this organism, a maximum of 20 specific *nif* genes are required for N_2 fixation. They are organized as a contiguous cluster spanning approximately 24 kb containing seven or eight transcriptional units (2, 16, 57). The *nifH*, *nifD*, and *nifK* genes respectively encode the Fe-protein (component 2) and the α and β subunits of the MoFe-protein (component 1) that together form the nitrogenase enzyme complex. Several other genes, including *nifV*, *nifB*, *nifE*, *nifN*, and *nifH*, are required for the formation or insertion of an Fe- and Mo-containing cofactor (FeMo-co) into the MoFe-protein. *nifM* is required for both stability and activity of the Fe-protein. The products of *nifF* and *nifJ* form a specific electron transfer pathway to nitrogenase. *nifQ* seems to be involved in molybdenum assimilation. *nifA* and *nifL* encode a positive transcriptional activator and antiactivator, respectively, of *nif* promoters which are transcribed by the form of RNA polymerase containing the sigma factor produced by the *ntrA* gene. The role of other genes, such as *nifU*, *nifS*, and two other recently discovered genes, *nifW* and *nifZ*, are uncertain.

In *A. chroococcum*, at least five regions of the genome may be involved in N_2 fixation. The organization of *nif* genes in region I is similar to that of the central portion of the *K. pneumoniae nif* cluster and includes *nifHDK*, *nifE/N*-like DNA, *nifUSV*, and *nifM* (19, 32). Region II contains *nifB* and a *nifA*-like gene. Region III contains the structural genes for the vanadium nitrogenase (*vnfHDGK*) (62, 63) and is closely linked to region IV, which contains *vnfEN* and *vnfA*, which are also specifically required for the vanadium nitrogenase system (36). Region V contains sequences homologous to the *fixABC* genes required for N_2 fixation in *Rhizobium meliloti* (18), although their function in *A. chroococcum* is unknown.

The two nitrogenase systems in *A. chroococcum* are encoded by separate sets of structural genes. Some genes involved in FeMo-co or FeVa-co synthesis (e.g., *nifEN* and *vnfEN*) and regulation (*nifA* and *vnfA*) are duplicated and

* Corresponding author.
† Present address: Department of Microbiology, University of Reading, Reading RG1 5AQ, United Kingdom.

specific for each system. However, *nifV* and *nifB*, important for maturation of MoFe-protein, and *nifM*, important for Fe-protein stability and activity, do not appear reiterated and therefore may serve in both systems. Mutations in *nifB* (29) and *nifM* in *A. vinelandii* (35) affect all three nitrogenase systems. The roles and regulation of these common genes as compared with the sets of specific genes is particularly interesting.

The objective of this work was to characterize the cluster of genes, including *nifV* and *nifM*, present in region I of the genome of *A. chroococcum* in order to understand their precise role in the molybdenum and vanadium nitrogenase systems in this organism. The nucleotide sequence of the region containing the *nifU*, *nifS*, *nifV*, *nifW*, *nifZ*, and *nifM* genes of *A. chroococcum* was determined. The region includes two open reading frames (ORF4 and ORF5) with no counterparts in the *K. pneumoniae nif* gene cluster. One of these open reading frames (ORF4, *nifP*) encodes a serine acetyltransferase (SAT) which plays a role in N_2 fixation in *A. chroococcum*.

MATERIALS AND METHODS

Bacterial strains, plasmids, and bacteriophage. *A. chroococcum* MCD105 (78) and strains constructed in this work were routinely grown in air at 30°C on RM or AM medium (59) with antibiotics added, where required, at the following concentrations (micrograms per milliliter): streptomycin, 20; nalidixic acid, 20; and kanamycin, 0.25. For growth of Nif^- strains, the medium was supplemented with ammonium acetate at 10 mM. *Escherichia coli* 71/18 [Δ(*lac-pro*) F′(*lacI lacZ* M15 proAB)] (49), which served as a host for all plasmids, was grown on Luria-Bertani (LB) medium (43) and maintained on minimal medium supplemented with thiamine at 0.1 μg/ml. Carbenicillin was added, where needed, at 50 μg/ml. *E. coli* JM39 (*cysE51 tfr-8*) (33) was obtained from B. Bachmann and maintained on minimal medium with cystine added to 100 μg/ml. *K. pneumoniae* UNF2050 (*nif-2104* [47]) and UNF828 (*nif-2568*::Tn*5* [46]) are strains bearing defined mutations in *nifM*. *K. pneumoniae* strains were routinely grown on LB medium or on a nitrogen-free defined medium (NFDM) with histidine added, where necessary, to 20 μg/ml. UNF50231, a His^+ revertant of the Nif^+ strain UNF5023 (15), was used as a nitrogen-fixing control strain in complementation studies. *K. pneumoniae* strains harboring pBR325-derived plasmids were selected on NFDM medium (15) containing tetracycline at 10 μg/ml and kanamycin, where required, at 25 μg/ml. Histidine was added at 20 μg/ml to the medium for UNF2050. Plasmids pDE15a and pDE15b contain the 5.1-kb *Kpn*I fragment, previously used to complement *K. pneumoniae nifV* and *nifM* mutants (19), cloned in either orientation in the vector pEMBL18+ (13). pSAR4, pSAR5, and pSAR6 contain *Sau*3A-*Kpn*I fragments of pDE15b cloned in the *Eco*RI site of pBR325 and are described in the text. The suffix a or b added to pSAR-series plasmids refers to the orientation in which the insert is cloned in pBR325; an a indicates that transcription from the pBR325 chloramphenicol acetyltransferase (CAT) promoter is in the same direction as the presumed transcription of the cloned *A. chroococcum nif* DNA. Bacteriophages IR1 (13) and M13KO7 (Pharmacia) were used as helper phages during the preparation of sequencing template from pEMBL-derived plasmids.

Genetic techniques. *E. coli* and *K. pneumoniae* strains were transformed by using the methods of Lederberg and Cohen (39) and Merrick et al. (48), respectively. Complementation of *K. pneumoniae nifM* mutants was defined as anaerobic growth on NFDM agar in 2 to 4 days at 30°C. Gene replacement mutagenesis of *A. chroococcum* was carried out with plasmid pCU101 used as the suicide vector to introduce mutations constructed in vitro into *A. chroococcum* as described previously (58).

Nucleic acid manipulations. Plasmid DNA was extracted essentially according to the method of Birnboim and Doly (7). *A. chroococcum* genomic DNA was prepared according to Robson et al. (59), and hybridizations to check mutant strains were performed as previously described (19). Restriction enzymes were purchased from Bethesda Research Laboratories, Boehringer Mannheim, or New England BioLabs and were used in TAS buffer (33 mM Tris-acetate [pH 7.9], 65 mM potassium acetate, 10 mM magnesium acetate, 0.5 mM dithiothreitol, 4 mM spermidine hydrochloride, pH 7.0) with the exception of *Sau*3A1 and *Kpn*I, which were used according to the supplier's instructions. DNA was electrophoresed through agarose gels in TAE buffer as described by Maniatis et al. (43) and extracted from low-melting-point agarose (Bethesda Research Laboratories ultrapure) by a freeze-squeeze technique modified from Thuring et al. (71).

Nucleotide sequence determination and analysis. Unidirectional ordered deletions of plasmids pDE15a and pDE15b formed the basis for DNA sequencing studies and were generated by partial *Sau*3A digestion as described by Robson et al. (62). Additional defined subclones or specific deletions of pDE15a were created to sequence regions deficient in *Sau*3A sites and to resolve regions containing ambiguities. Gaps were completed by using synthetic oligonucleotide primers. Sequence was determined at least twice from individual templates and on both strands. Single-stranded template was produced from pEMBL-derived plasmids essentially according to Dente et al. (13) and sequenced by the dideoxynucleotide method (64) using [α-^{35}S]dATP (6). Reactions were performed in 96-well microtiter plates (Falcon 3911) and were denatured at 90°C for 10 min prior to electrophoresis. Deoxy- and dideoxynucleotides were obtained from Pharmacia; Klenow fragment and [α-^{35}S]dATP were from Amersham International plc. Gel readings were assembled into a single "contig" by using the computer program DBUTIL (69), and computer analyses were done with the University of Wisconsin Genetics Computer Group programs (14). Initial sequence data base searches were performed with the programs FASTN and FASTP (41). Further protein sequence comparisons used the algorithm of Smith and Waterman (68) as implemented on the Connection Machine CM-2 data parallel computer (31). This software allows multiple sequence alignments to be used in data base searches, which can markedly increase the sensitivity of the method (21, 30).

Enzyme and other assays. SAT was measured spectrophotometrically essentially as described by Kredich and Tomkins (38). A decrease in A_{230} due to cleavage of the thioester bond of acetyl coenzyme A (acetyl-CoA) was monitored by using a Cary 2200 dual-beam spectrophotometer. The assay was carried out at 37°C in 1-cm-light-path cuvettes in a reaction volume of 1 ml which, when complete, contained the following: Tris-HCl (pH 7.6), 50 μmol; L-serine, 1 μmol; acetyl-CoA, 0.4 μmol; and cell extract, 50 to 100 μg. The reaction was monitored for 5 min without added serine to allow measurement of the background level of cleavage of the acetyl-CoA thioester bond. The assay proper was started with the addition of serine, and the rate of serine-dependent acetyl-CoA cleavage was calculated after deducting the background rate. Specific activities are ex-

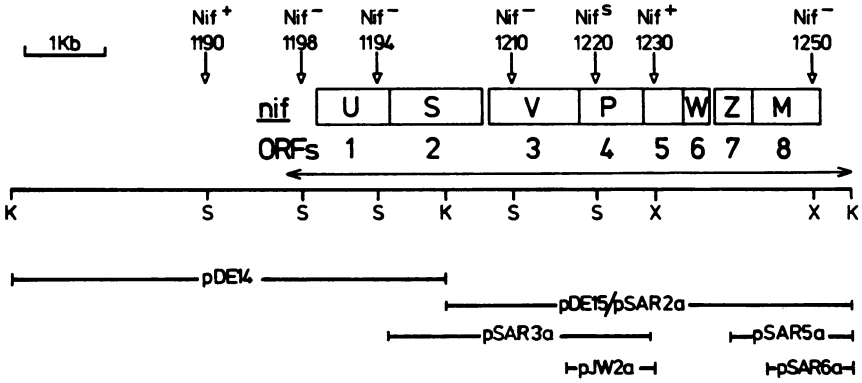

FIG. 1. Physical and genetic map of the *nifU*-to-*nifM* gene cluster in *A. chroococcum*. Shown is a restriction map for a portion of the chromosome cloned in two primary plasmids pDE14 and pDE15 which were described previously (19). Restriction sites: K, *Kpn*I; S, *Sal*I; X, *Xho*I. Nucleotide sequence was determined for a region of 7,099 bp indicated by the horizontal arrows. Subfragments of this region used to construct plasmids for complementation studies are indicated by the bars at the bottom. The positions of eight ORFs are indicated by the boxes, and their genetic assignment is shown within. Sites of insertion of the Km^r gene from pUC4K into the chromosome of *A. chroococcum* are indicated at the top, together with the strain number and the phenotype with respect to N_2-fixing capacity in medium containing molybdenum. Nif^s refers to slow growth with N_2 as the N source.

pressed as nanomoles per hour per milligram of protein. Background levels of serine-independent acetyl-CoA cleavage were approximately 3 and 5.5 nmol/mg/min for *E. coli* and *A. chroococcum*, respectively. Cell extracts treated at 90°C for 10 min were devoid of activity with and without serine addition. Nitrogenase activities in whole cells were measured by the acetylene reduction assay in which acetylene and ethylene were determined by using a flame ionization detector. Cultures were capped with Suba-Seals, and acetylene was added to 10% of the gas headspace. Acetylene reduction activities of *K. pneumoniae* strains were not determined as we have previously shown that strains complemented for growth on N_2 display near-wild-type acetylene reduction levels, whereas strains not complemented for anaerobic growth show levels of acetylene reduction similar to the activity of the mutant strain alone (19). Protein in whole cells or crude extracts was measured by using the Folin-Ciocalteau reagent (42) with bovine serum albumin (fraction V; Sigma Chemical Co.) as a standard. Growth of bacterial cultures was monitored with a Klett-Summerson photoelectric colorimeter with a no. 64 filter.

RESULTS

Location of *nifM*. *nifU*, *nifS*, *nifV*, and *nifM* are clustered in a region of the *A. chroococcum* genome spanning about 7 kb (19). In our previous study, the approximate position of each gene was shown by (i) hybridization studies using various *K. pneumoniae nif* gene probes and (ii) the ability of plasmids pLC11, pSAR2a, and pSAR3a containing overlapping subfragments of this region (Fig. 1) to complement *nifU*, *nifS*, *nifV*, and *nifM* mutants of *K. pneumoniae* (19). The precise position of *nifM* could not be located by hybridization studies using the *K. pneumoniae nifM* gene probe. Therefore, prior to determining the nucleotide sequence of this region, we located *nifM* in further complementation studies. A set of nested deletions was first constructed within the insert in plasmid pDE15, which carries a 5.1-kb *Kpn*I fragment in the plasmid vector pEMBL18+. Once the deletions had been constructed, inserts were recloned into the *Eco*RI site in pBR325 potentially under the control of the CAT gene promoter. Plasmid pSAR5a contained the smallest insert (Fig. 1), at 1.5 kb, which complemented both *K. pneumoniae nifM* mutant strains (UNF 828 and UNF 2050) for growth with N_2 as the sole N source. Complementation was lost in plasmid pSAR6a, which contained an insert of only 1.05 kb. The failure of pSAR6a to complement the *nifM* mutants was not attributable to an inhibitory effect because it did not affect expression of N_2 fixation in the Nif^+ control, UNF50321. These experiments placed *nifM* within 1.5 kb from the right-hand *Kpn*I site of the insert in pDE15 as shown in Fig. 1.

Nucleotide sequencing. The nucleotide sequence of 7,099 bp spanning the 5.1-kb *Kpn*I insert in pDE15 and 2 kb of the adjacent region borne on plasmid pDE14 is shown in Fig. 2. Eight complete ORFs were identified, all reading in the same direction (Fig. 1; Table 1) and possessing the biased codon usage (21) typical of *A. chroococcum* genes. Assignment of genes was based on deduced amino acid sequence identity to *K. pneumoniae nif* genes (Table 1; Fig. 3, 4, and 7) (2, 5, 53). The assignment of ORF1, -2, and -3 as *nifU*, *nifS*, and *nifV*, respectively, is consistent with our complementation results. pSAR 3a (Fig. 1) complemented *nifS* and *nifV* but not *nifU* mutants and contains an insert of 3,485 bp from an *Eco*RI site at position 1225 in the C terminus of *nifU* to the *Pst*I site at 4710 in ORF5. pSAR2a complemented *nifV* but not *nifS* mutants and contains a fragment defined by the *Kpn*I sites at 1961 in *nifS* and at 7099. ORF8 is *nifM*, as defined by sequence identity to the *K. pneumoniae nifM* gene and also by the complementation results presented above. pSAR5a, which complements the *nifM* mutant, contains a fragment from the *Sau*3A1 site at position 5664 in the C-terminal coding region of ORF7 to the *Kpn*I site at 7099. pSAR6a, which fails to complement this mutation, contains a fragment starting at a *Sau*3A1 site at 6082 within ORF8 and terminates at the same *Kpn*I site. ORF6 and -7 display sequence identity to *K. pneumoniae nifW* and *nifZ*, respectively. Two ORFs located between *nifV* and *nifW* in *A. chroococcum* apparently have no counterparts in the *K. pneumoniae nif* gene cluster (2).

The likely initiation codon for all eight genes or ORFs is ATG, which is preceded in each case by a polypurine tract characteristic of potential ribosome binding sites, including the motif 5′-GGAG-3′ situated 5 to 11 nucleotides upstream

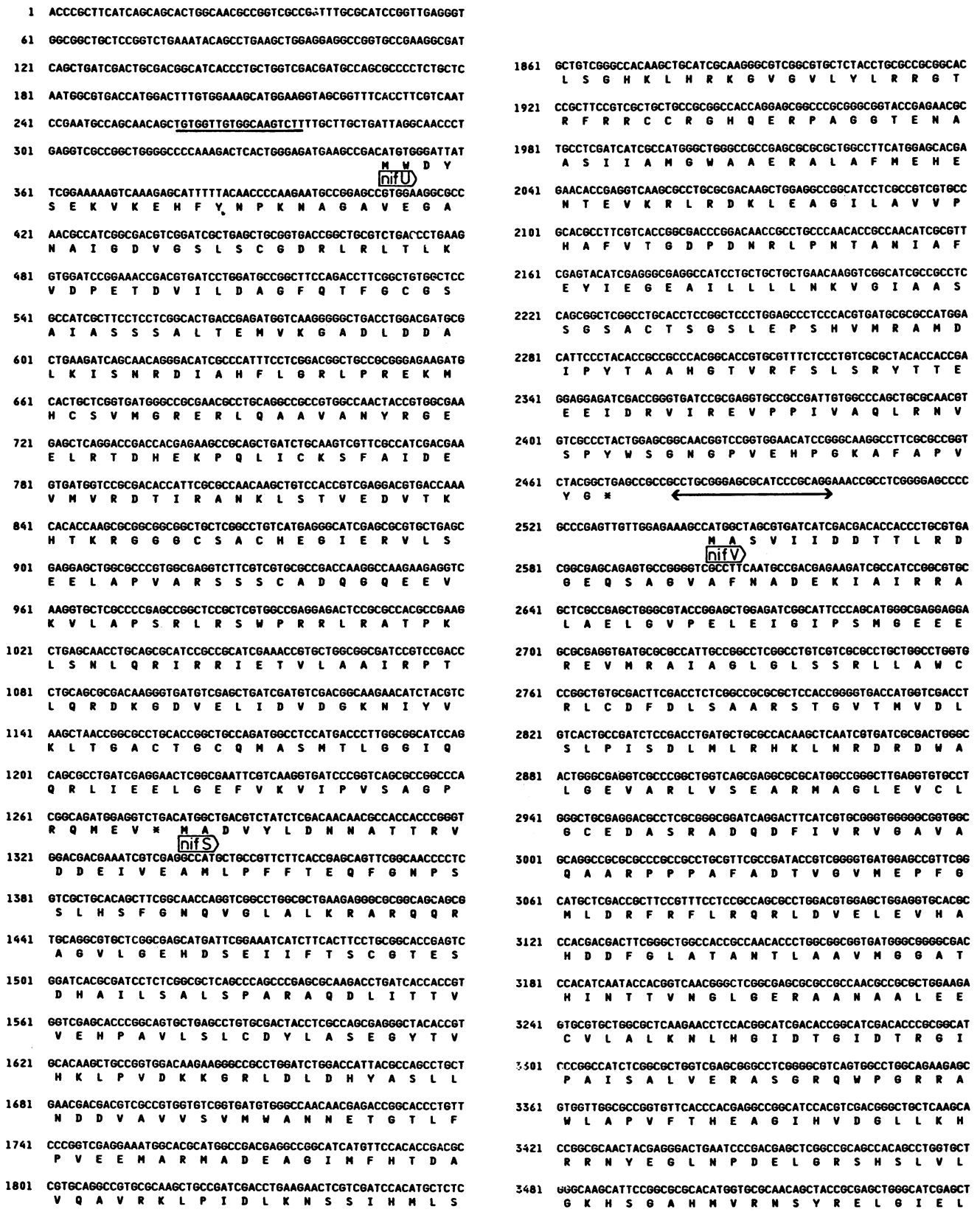

FIG. 2. Nucleotide sequence of the *nifU*-to-*nifM* gene cluster from *A. chroococcum*. Predicted amino acid sequences for the several ORFs identified from the nucleotide sequence are given below in single-letter code. Gene assignments and directions of transcription are indicated by the boxed arrows in the space below. Potential *nif* promoter sequences are underlined, and an inverted repeat sequence is marked beneath by the double-arrow line.

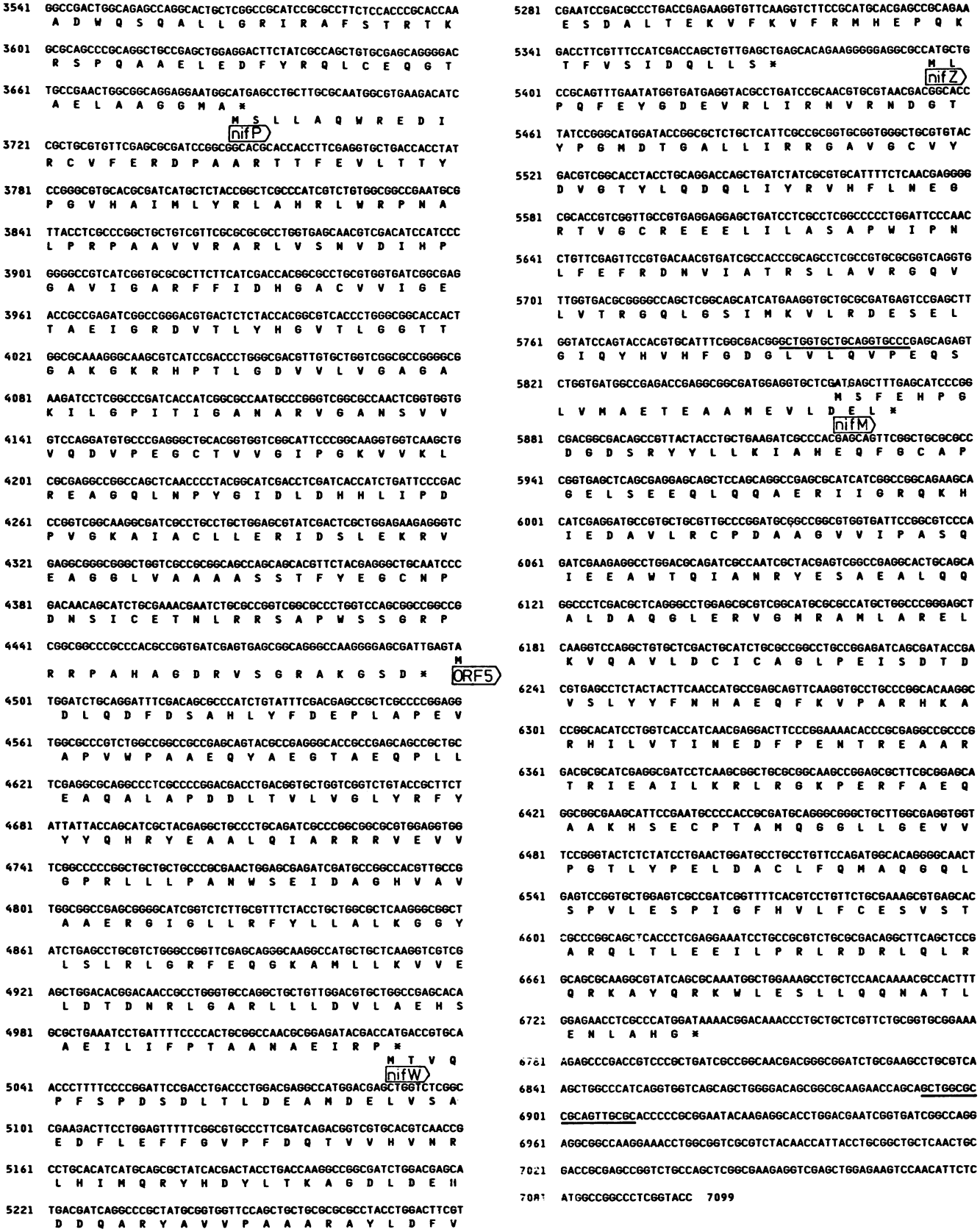
nifP
ORF5
nifW
nifZ
nifM

TABLE 1. Properties of *A. chroococcum nif* gene products

Sequence[a]	ORF	Gene[b]	No. of amino acids[c]	Mol wt[d]	pI	% Identity[e]
349–1278	1	*nifU*	309	33,932	7.56	51.1
1280–2470	2	*nifS*	396	43,295	6.2	58.1
2543–3691	3	*nifV*	382	41,489	5.68	42.9
3688–4497	4	*nifP*	269	28,577	9.04	41.0
4500–5033	5		177	19,757	4.73	
5030–5374	6	*nifW*	114	13,126	4.30	28.6
5395–5871	7	*nifZ*	158	17,768	4.59	43.9
5861–6742	8	*nifM*	293	32,978	5.89	29.1

[a] Numbers refer to positions in sequences shown in Fig. 2.
[b] Gene assignment made by complementation and protein sequence identity.
[c] Including the N-terminal methionine.
[d] Calculation includes the N-terminal methionine.
[e] Identity to corresponding *K. pneumoniae nif* gene products except for *nifP*, for which identity is to the *E. coli cysE* gene product.

(20, 65). All ORFs terminate with TGA except for *nifM* (TAA). Most of the genes, or potential genes, overlap except for *nifW* and *nifZ*, for which the intergenic region is 20 bp, and *nifS* and *nifV*, which are separated by 72 bp containing an inverted repeat sequence (indicated by arrows in Fig. 2) which could form a stem-loop structure at the RNA level with a calculated ΔG of -21.1 kcal (ca. -88.3 kJ). In the cases of *nifV* and ORF4 and of ORF5 and *nifW*, the translational stop signal for the first gene and the translational initiation signal for the second gene overlap. The overlap between *nifZ* and *nifM* extends across the coding region to the extent of 11 bp.

Three regions within the 7,099-bp sequence underlined in Fig. 2 display close similarity to the *nif* consensus promoter defined by Beynon et al. (4). The sequence 5′-TGTGGCAA GTCTTTTGCTT-3′ occurs 65 bp 5′ to *nifU*. The sequence 5′-GCTGGTGCTGCAGGTGCCC-3′ is situated within the *nifZ* gene, 68 bp upstream from the translational initiation codon of *nifM*. A third *nif* promoter-like sequence 5′-GCTGGCGCCGCAGTTGCGC-C-3′ is located 150 bp 3′ to *nifM*, suggesting that this region may encode further gene products involved in diazotrophy of *A. chroococcum*. However, none of these presumptive *nif* promoters is preceded at the appropriate distance by the sequence TGT-n_{10}-ACA, which corresponds to a *nifA*-specific upstream activator (9).

Analysis of predicted *nif* gene products. The predicted properties of the *A. chroococcum nifU*, *nifS*, *nifV*, *nifW*, *nifZ*, and *nifM* gene products are listed in Table 1.

NIFU and NIFS. NIFU may be a metalloprotein. Two pairs of cysteine residues located toward the C terminus of the predicted polypeptide are arranged in a -Cys-X-X-Cys- configuration typical of the spacing of cysteines in many nonheme iron proteins. These cysteine residues are also conserved in NIFU polypeptides from *Azotobacter vinelandii*, *K. pneumoniae*, and *Anabaena* strain 7120 (Fig. 3). The sequence -Lys-Ala-Gly-Gly-Gly-Cys-Ser-Ala-Cys-His- starting at residue 169 in *A. chroococcum* NIFU, which contains the first example of the -Cys-X-X-Cys- motif, also resembles part of the heme-binding domains in *c*-type cytochromes, especially those from *Rhodopseudomonas* species (1). The *nifS* gene products are relatively highly conserved between *Azotobacter* species, *K. pneumoniae*, and *Anabaena* strain 7120 (Fig. 3). Especially striking is an extensive block of identity toward the C terminus of the predicted NIFS polypeptides.

NIFV. Amino acid sequences for *nifV* gene products from *Azotobacter* species and *K. pneumoniae* are aligned in Fig. 4. NIFV is involved in the formation of homocitrate, a component of FeMo-co which is incorporated into the MoFe-protein (23, 24). NIFV is postulated to be homocitrate synthase (23). In yeasts and fungi, homocitrate synthase carries out the condensation between acetyl-CoA and α-ketoglutarate (70, 72). It is particularly interesting that NIFV exhibits similarity to the *leuA* gene product (isopropylmalate synthetase; E.C. 4.1.3.12) from both *Salmonella typhimurium* (56) *Saccharomyces cerevisiae* (3) (Fig. 4). Isopropylmalate synthase carries out the condensation between acetyl-CoA and α-ketoisovalerate. This suggests that NIFV catalyzes a reaction analogous to that catalyzed by LEUA, consistent with the proposal that acetyl-CoA and α-ketoglutarate are substrates for NIFV (Fig. 5). Umbarger (74) also draws a parallel between the reaction carried out by isopropylmalate synthase and homocitrate synthase. Therefore, the two major blocks of identity between NIFV and LEUA, which are placed at approximately equivalent positions in the two enzymes, could be binding sites for acetyl-CoA or for the structurally common portion of α-ketoglutarate and α-ketoisovalerate (Fig. 5).

ORF4 (*nifP*) encodes SAT. The ORF4 polypeptide shows 41% identity to the *cysE* gene product (CYSE) from *E. coli* (Fig. 6), which encodes SAT, a key enzyme in cysteine biosynthesis in *E. coli* (12). The possibility that ORF4 codes for a similar enzyme was tested in complementation studies with *E. coli* JM39, a cysteine auxotroph defective in *cysE* and lacking SAT activity. JM39 was transformed with plasmids pSAR2a and pSAR3a constructed for earlier *nif* complementation studies and containing ORF4 intact (Fig. 1). Both JM39(pSAR2a) and JM39(pSAR3a) grew well in minimal medium lacking cystine, consistent with the idea that ORF4 may encode SAT. Both *nifV* and ORF4 are common to pSAR2a and pSAR3a; therefore, a plasmid (pJW2a) that carries only ORF4 was made. pJW2a carries a fragment of 1,158 bp containing ORF4 cloned under the control of the *lacZ* promoter in pTZ19R plus 140 bp of the 3′ coding region of *nifV* and 125 bp at the 5′ coding region of ORF5. pJW2a also complemented the *cysE* mutation in JM39. SAT activity was compared in cell extracts of *E. coli* 5K (a cysteine prototroph), JM39(pJW2a), and JM39 grown in LB. The first two strains had specific activities of 34 and 43 nmol/mg/min, whereas no activity was detected in JM39. SAT in *E. coli* and *S. typhimurium* is constitutive but subject to feedback inhibition by low levels of cysteine. SAT activity in *E. coli* JM39(pJW2a) crude extracts was also inhibited by cysteine, although the concentration required to cause 50% inhibition (50 μM) was 50-fold higher than that reported to inhibit the *E. coli* enzyme (38).

Although ORF4 and CYSE differ in length by only four amino acid residues, the region of similarity is displaced in the two proteins (Fig. 6). CYSE has a 74-residue extension at the N terminus, whereas ORF4 is accordingly longer at the C terminus. The essential catalytic domain therefore appears confined to the 186-residue portion common to each protein. ORF4 also shows similarity to the *E. coli lacA* gene product, thiogalactoside acetyltransferase (22), and the *Rhizobium leguminosarum nod*L gene product (Fig. 6), suspected to be an acetyltransferase (17). The region common to this class of known or suspected acetyltransferases therefore may be important in binding acetyl-CoA.

ORF5, NIFW, NIFZ, and NIFM. The finding of a SAT gene adjacent to *nifV* suggested that other genes in this cluster might be involved in sulfur assimilation. However,

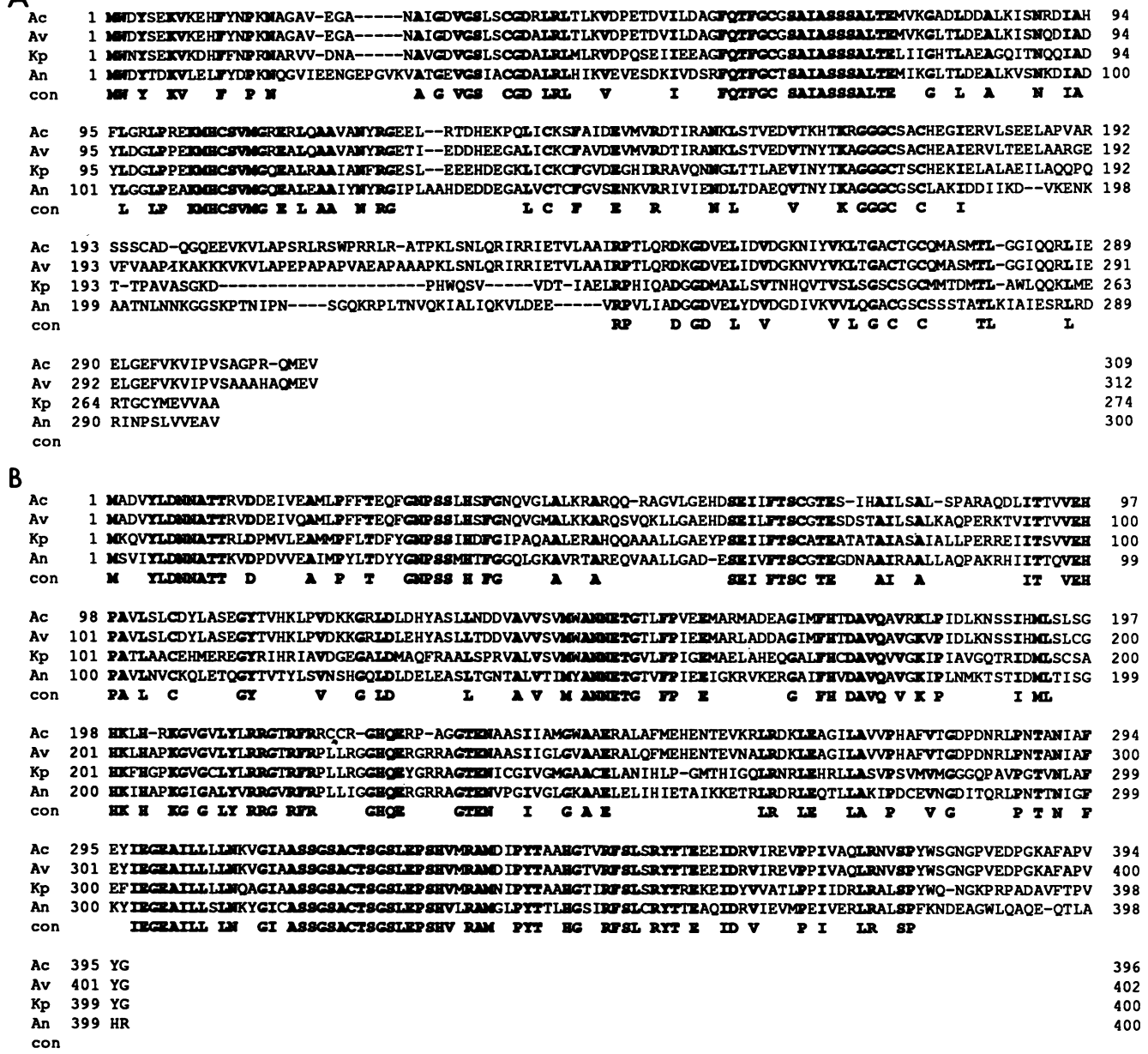

FIG. 3. Alignments of predicted amino acid sequences for the *nifU* (A) and *nifS* (B) gene products from *A. chroococcum* (Ac), *A. vinelandii* (Av) (5), *K. pneumoniae* (Kp) (2, 5); and *Anabaena* strain 7120 (An) (50). In the multiple sequence alignments, fully conserved positions are noted in the consensus line (con) and printed in bold font.

no other significant similarities were detected when comparisons were made between each of the *nif* genes considered here and several other *cys* genes from *E. coli*. These included *cysP*, encoding a thiosulfate-binding protein; *cysT*, *cysW*, and *cysA*, encoding the sulfate permease (26, 67); and *cysK* (10, 40) and *cysM* (67), encoding *O*-acetylserine (thiol) lyase isozymes A and B, respectively.

Alignments of *nifW*, *nifZ*, and *nifM* gene products from several organisms are shown in Fig. 7. NIFW polypeptides appear poorly conserved between *Azotobacter* species and *K. pneumoniae*. *fixO* from *Azorhizobium caulinodans* (34) shows 32% identity to *Azotobacter nifW* (Fig. 7), suggesting that these gene products are functionally analogous. *nifZ* gene products are quite highly conserved between *Azotobacter* species and *K. pneumoniae*. Also, there is 23% sequence identity between the translated product of a partial ORF (ORF2) from *Rhodobacter capsulatus* (44) and the N terminus of *nifZ* (Fig. 7), suggesting that ORF2 in *R. capsulatus* is analogous to *nifZ*.

The sequences of the *nifM* products from *A. chroococcum* and *K. pneumoniae* contain 26% identity dispersed throughout their lengths (Fig. 7). The dispersed nature of this identity may account for our previous lack of success in identifying a *nifM*-like gene in *A. chroococcum* by heterologous hybridization (19). Pagani et al. (52) demonstrated that reactivation of an inactive 2Fe-containing form of *K. pneu-*

```
Ac NifV   9 TTLRDGEQSAGVAFNADEKIAIRRALAELGVPELEIGIPSMGEEEREVMRAIAGLGLSS-RLLAWCRLCDFDLSAARST--GVTMVDL--SLPISDLMLR 103
Av NifV   9 TTLRDGEQSAGVAFNADEKIAIARALAELGVPELEIGIPSMGEEEREVMHAIAGLGLSS-RLLAWCRLCDVDLAAARST--GVTMVDL--SLPVSDLMLH 103
Kp NifV   9 TTLRDGEQSPGVAFRTSEKVAIAEALYAAGITAMEVGTPAMGDEEIARIQLVRRQLPDA-TLMTWCRMNALEIRQSADL--GIDWVDI--SIPASDKLRQ 103
St LeuA  10 TTLRDGEQALQASLSAKEKLQIALALERMGVDVMEVGFPVSSPGDFESVQTIARTIKNS-RVCALARCVEKDIDVAAQALKVADAFRIHTFIATSPMHIA 108
Sc LeuA  66 TDLRDGNQSLPDPMSVEQKKEYFHKLVNIGFKEIEVSFPSASQTDFDFTRYAVENAPDDVSIQCLVQSREHLIKRTVEALTGAKKATIHTYLATSDMFRE 165
con          T LRDG Q          K      L   G    E   P                                                    S

Ac NifV 104 HKLNRDRDWALGEVARLVSEARMAGLE----------VCLGCEDASRADQDFIVRVGAVAQAA------RPPPA-FADTVGVMEPFGMLDRFRFLRQRL- 185
Av NifV 104 HKLNRDRDWALREVARLVGEARMAGLE----------VCLGCEDASRADLEFVVQVGEVAQAA------GARRLRFADTVGVMEPFGMLDRFRFLSRRL- 186
Kp NifV 104 YKLREPLAVLLERLAMFIHLAHTLGLK----------VCIGCEDASRASGQTLRAIAEVAQQC------AAARLRYADTVGLLDPFTTAAQISALRDVW- 186
St LeuA 109 TKLRRTLDEVIERAVYMVKRARNYTDD----------VEFSCEDAGRTPVDDLARVVEAAINA------GARTINIPDTVGYTMPFEFAGIISGLYERVP 192
Sc LeuA 166 IVFNMSREEAISKAVEATKLVRKLTKDDPSQQATRWSYEFSPECFSDTPGEFAVEICEAVKKAWEPTEENPIIFNLPATVEVASPNVYADQIEYFATHIT 265
con                                                E                                   TV    P

Ac NifV 186 DVE---LEVHAHDDFGLATANTLAAVMGGATHINTTVNGLGERAANAALEECVLAL---KNLHGIDTGIDTRGIPAISALVERASGRQWPGRRAWL-APV 278
Av NifV 187 DME---LEVHAHDDFGLATANTLAAVMGGATHINTTVNGLGERAGNAALEECVLAL---KNLHGIDTGIDTRGIPAISALVERASGRQVAWQKSVVGAGV 280
Kp NifV 187 SGE---IEMHAHNDLGMATANTLAAVSAGATSVNTTVLGLGERAGNAALETVALGL---ERCLGVETGVHFSALPASCQRVAEAAQRAIDPQQPLVGELV 280
St LeuA 193 NIDKAIISVHTHDDLGIAVGNSLAAVHAGARQVEGAMNGIGERAGNCALEEVIMAIKVRKDIMNVHTNINHHETGAPARPSVQICNIADPSQQSDCRQRR 292
Sc LeuA 266 EREKVCISTHCHNDRGCGVAATELGMLAGADRVEGCLFGNGERTGNVDL--VTVAMNMYTQ--GVSPNLDFSDLTSVLDVVERCNKIPVSQRAPYGGDLV 361
con                 H H D G           GA        G GER  N  L

Ac NifV 279 FTHEAGIHVDGLLK                                                                                    292
Av NifV 281 FTHEAGIHVDGLLK                                                                                    294
Kp NifV 281 FTHESGVHVAALLR                                                                                    294
St LeuA 293 FRHSSGIHQDGVLK                                                                                    306
Sc LeuA 362 VCAFSGSHQDAIKK                                                                                    375
con              G H
```

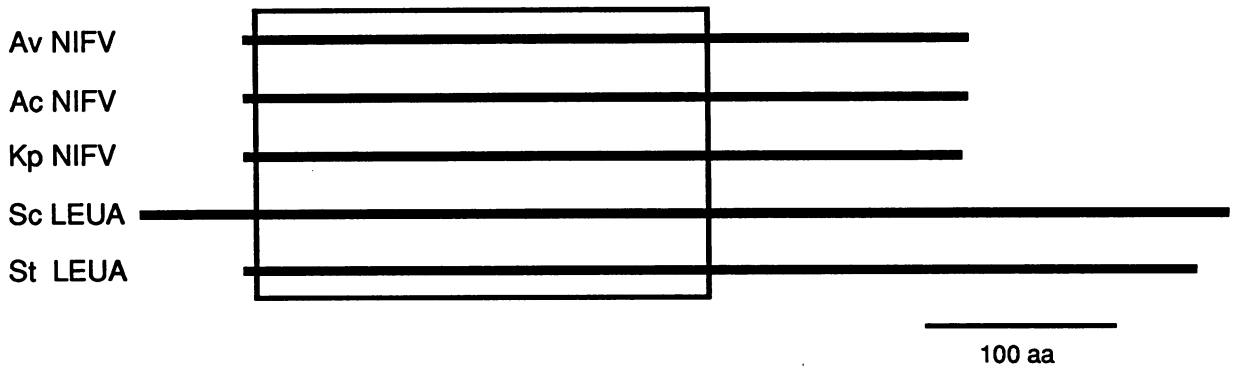

FIG. 4. Alignments of predicted amino acid sequences for the *nifV* and *leuA* gene products from *A. chroococcum* (Ac), *A. vinelandii* (Av) (5); and *K. pneumoniae* (Kp) (2, 5). *leuA* gene products are from *S. typhimurium* (St) (56) and *Saccharomyces cerevisiae* (Sc) (3). In the multiple sequence alignments, fully conserved positions are noted in the consensus line (con) and printed in bold font. aa, Amino acids.

moniae Fe-protein could be catalyzed by the enzyme bovine liver rhodanese, and they speculated that NIFM may act as a *nif*-specific sulfur transferase. However, alignments of NIFM and rhodanese (55) amino acid sequences display no apparent similarities (data not shown). Furthermore, the predicted tertiary structure of *nifM* showed no obvious similarity with the structure of rhodanese determined at 0.25-nm resolution (55).

Mutagenesis of the *nifU* to *nifM* cluster in *A. chroococcum*. Functionality of the region described above was established by mutagenesis. The kanamycin resistance gene cassette (Km^r) from pUC4K (75) was introduced at several locations in the *nifU* to *nifM* region of the *A. chroococcum* chromosome by gene replacement. Sites of insertions are illustrated in Fig. 1. Mutations were constructed in MCD105, which normally fixes N_2 in the presence of molybdenum. Introduction of the Km^r gene into the *Sal*I site 1.1 kb upstream of the sequenced region did not adversely affect N_2 fixation or growth. However, inserts in the *Sal*I sites at positions 154, 1117 (in *nifU*), and 2813 (in *nifV*) and in the *Xho*I site at position 6617 (in *nifM*) resulted in Nif^- phenotypes.

The *nifV* mutant (MCD1210) did not show any significant acetylene reduction activity, which is surprising in view of the fact that *nifV* mutants of *K. pneumoniae* fail to reduce N_2 yet still exhibit significant acetylene reduction activity (45, 46). The activity in *K. pneumoniae* may be explained by the substitution of homocitrate with some other carboxylic acid, possibly citrate, which leads to a partially active enzyme. Indeed, citrate can be incorporated into $NIFV^-$ MoFe-protein from *A. vinelandii* in vitro and in reconstitution experiments with $NIFV^-$ MoFe-protein (23), and acetylene-reducing activity is restored. In *A. chroococcum*, it is possible that no such substitution is possible in vivo because citrate levels are low.

A mutant (MCD1220b) with the Km^r gene in the *Sal*I site within the SAT-encoding gene at position 3886 grew nor-

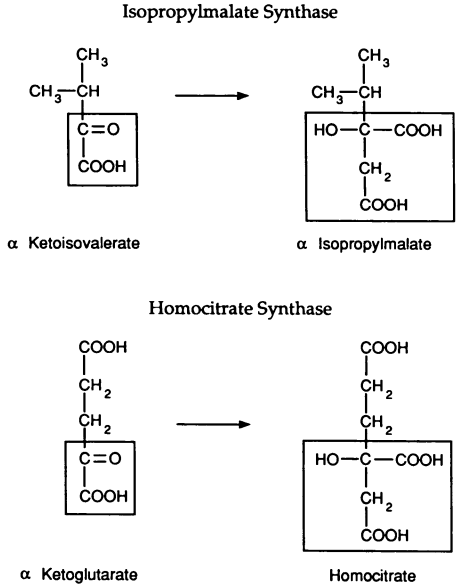

FIG. 5. Similarity between the reactions catalyzed by isopropylmalate synthase and homocitrate synthase, indicated by the boxed moieties which are similar in each reaction.

mally when ammonium was supplied but only slowly with a doubling time of 5 h, compared with 3 h for MCD105, in N_2-fixing conditions. Also, the appearance of acetylene reduction activity was delayed by 2 to 3 h in MCD1220b as compared with MCD105 when organisms were transferred from ammonium-rich to ammonium-free medium. When grown under N_2-fixing conditions, SAT specific activity in MCD1220 was significantly lower than in MCD105: specific activities were 6.7 and 12.4 nmol/min/mg, respectively. The effect on N_2 fixation seen in MCD1220 could be explained by polarity of the mutation on downstream genes. However, strains (MCD1230a and -b) with the Km^r cassette introduced in both orientations into ORF5 at the *Xho*I site at position 4620 were indistinguishable from the parent strain. The phenotype of MCD1220 therefore appears to arise solely from inactivation of the SAT gene.

DISCUSSION

The *nifU*, *nifS*, *nifV*, *nifW*, *nifZ*, and *nifM* genes in *A. chroococcum* are arranged as in *K. pneumoniae* except that two additional ORFs are located between *nifV* and *nifW*. The gene cluster in *A. chroococcum* closely matches that described for *A. vinelandii* (28) with respect to gene order, spacing between genes, and the presence of the additional ORFs between *nifV* and *nifW*. The overall nucleotide sequence identity between the two *Azotobacter* species over this region is 90.03%.

One of the additional ORFs (ORF4) found in *A. chroococcum*, which corresponds to ORF7 in the major *nif* cluster of *A. vinelandii* (27, 28), probably encodes SAT, given the sequence identity between ORF4 and CYSE, complementation of the *cysE* mutation in JM39, and restoration of serine-dependent, acetyl-CoA cleavage activity in JM39(pJW2a). There are no reports describing the cysteine biosynthetic pathway in this genus, though *O*-acetylserine is a likely intermediate. Disruption of ORF4 does not cause cysteine auxotrophy or affect growth on ammonia adversely, and therefore normal biosynthesis of cysteine in *A. chroococcum* appears independent of this gene. However, ORF4 appears necessary for optimal N_2 fixation. Therefore, *A. chroococcum* contains at least two SAT genes. One encodes an isozyme, detectable in extracts of MCD1220b, which probably serves a housekeeping function, whereas ORF4 encodes a second isozyme required additionally under N_2-fixing conditions. We propose to give ORF4 the designation *nifP*. Our results are at variance with those described for *A. vinelandii*, in which mutation of the corresponding ORF (ORF7) results in no detectable phenotype, though mutation of ORF8 (corresponding to ORF5 in this study) does lead to slow diazotrophy (28).

The existence of a *nif*-specific SAT gene in *A. chroococcum* implies that higher intracellular levels, or rates of synthesis, of *O*-acetylserine are required when organisms are fixing N_2. Therefore, *O*-acetylserine may have some specific role in N_2 fixation per se. Increased levels of *O*-acetylserine in the cell may stimulate expression of genes for sulfate uptake and assimilation as occurs in enteric bacteria (37). If, however, *O*-acetylserine is involved in cysteine biosynthesis in this genus, then it is possible that maximal rates of synthesis of active nitrogenase components or other *nif* gene products require higher rates of cysteine and/or methionine synthesis. Nitrogenase is abundant in N_2-fixing cells, and the Fe-protein from *A. chroococcum* contains 2.09% cysteine and 4.6% methionine while MoFe-protein contains 2.22% cysteine and 3.13% methionine (77). There are no data concerning the average amino acid content of proteins in *A. chroococcum*; however, proportions of cysteine and methionine averaged for many families of proteins are only marginally lower than found in nitrogenase, at 1.7 and 3.1%, respectively (11). Therefore, it seems unlikely that maximal rates of synthesis of nitrogenase polypeptides place an especial burden on cysteine biosynthesis specifically. However, this explanation cannot be ruled out, and it is also possible that other *nif* gene products whose functions are unknown at present may serve an analogous role for the production of other amino acids that are relatively abundant in nitrogenase.

Alternatively, NIFP may have a role in optimizing manufacture of metal sulfur clusters for either Fe-protein or MoFe-protein because there is evidence that cysteine is the sulfur donor to nonheme metal sulfur centers in *E. coli* (76). Synthesis of metal centers in both molybdenum nitrogenase components may require high intracellular concentrations of cysteine. This possibility is supported by the finding that FeMo-co synthesis is stimulated by exogenously supplied cysteine or cystine in $NifQ^-$ and Mol^- mutants of *K. pneumoniae* (73). In enteric bacteria at least, SAT is the key enzyme controlling levels of cysteine and is highly sensitive to feedback inhibition by cysteine. The observation that relatively high concentrations of cysteine are required to inhibit NIFP activity, at least in crude extracts in *E. coli*, is consistent with it allowing the accumulation of cysteine in *A. chroococcum*. The overlapping arrangement of *nifV*, *nifP*, ORF5, and *nifW* may simply reflect genetic economy, or it may be functionally significant. Such an arrangement could cause translational coupling (51), possibly ensuring that stoichiometric amounts of each gene product are formed. This may be necessary if two or more of these gene products form a complex. It is interesting that both NIFV and NIFP probably both catalyze condensation reactions involving acetyl group transfer from acetyl-CoA to α-ketoisovalerate and serine, respectively. Also, both reactions are normally associated with amino acid biosynthesis. In *E. coli* and *S.*

A

```
Ac NifP  10 DIRCVFERDPAARTTFEVLTTYPGVHAIMLYRLAHRLWRPNALPRPAAVVR--ARLVSNVDIHPGAVIGARFFIDHGACVVIGETAEIGRDVTLYHGVTL 107
Av NifP  10 DIRCVFERDPAARTTFEVLTTYPGVHAIIGYRLANRLWR-AAWRYPARLLSFVARMLSNVDIHPGATIGERFFIDHGACVVIGETAEIGNDVTLYHGVTL 108
Ec CysE  84 DIQAVRTRDPAVDKYSTPLLYLKGFHALQAYRIGHWLW-NQGRRALAIFLQNQVSVTFQVDIHPAAKIGRGIMLDHATGIVVGETAVIENDVSILQSVTL 182
con         DI  V  RDPA       L    G HA   YR    LW        A            VDIHP A IG     DH    V GETA I  DV     VTL

Ac NifP 108 GGTTGAKGKRHPTLGDVVLVGAGAKILGPITIGANARVGANSVVVQDVPEGCTVVGIPGKVV                                      169
Av NifP 109 GGTSWNKGKRHPTLRDGVLVGAGAKILGPITVGAGARVGANSVVVQDVPDGCTVVGIPGKVV                                      170
Ec CysE 183 GGTGKSGGDRHPKIREGVMIGAGAKILGNIEVGRGAKIGAGSVVLQPVPPHTTAAGVPARIV                                      244
con         GGT    G RHP     V  GAGAKILG I  G  A  GA SVV Q VP   T  G P   V
```

```
Ac NifP  120 TLGDVVLVGAGAKILGPITIGANARVGANSVVVQDVPEGCTVVGIPGKV  168
Av NifP  121 TLRDGVLVGAGAKILGPITVGAGARVGANSVVVQDVPDGCTVVGIPGKV  169
Ec CysE  195 KIREGVMIGAGAKILGNIEVGRGAKIGAGSVVLQPVPPHTTAAGVPARI  243
Rl NodL  133 SIGRHAWIGGGAIILPGVTIGDHAVIGAGSVVTRDVPAGSTAMGNPARV  181
Ec LacA  133 TIGNNVWIGSHVVINPGVTIGDNSVIGAGSIVTKDIPPNVVAAGVPCRV  181
con                  G    I     G    GA S V   P     G P
```

B

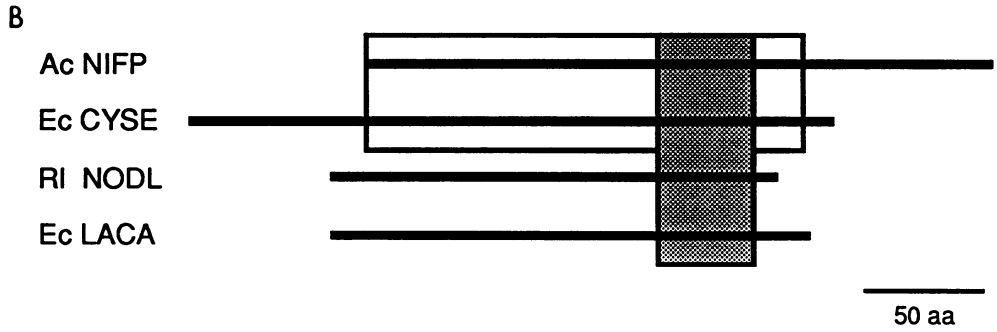

FIG. 6. (A) Alignment for predicted amino acid sequences for ORF4 (*nifP* gene product) and SAT (*cysE* gene product). The upper alignments show similarity between the predicted amino acid sequences for ORF4 (*nifP*) from *A. chroococcum* (Ac), ORF7 from the major *nif* gene cluster from *A. vinelandii* (Av) (27, 28), and the SAT-encoding gene, *cysE*, from *E. coli* (Ec) (12). The lower alignment shows the most conserved region observed between these sequences and predicted amino acid sequences of gene products of *nodL* from *R. leguminosarum* (Rl) (17) and *lacA* from *E. coli* (Ec) (22). In the multiple sequence alignments, fully conserved positions are noted in the consensus line (con) and printed in bold font. (B) Relative position of this conserved region (indicated by the stippled box). The extent of significant sequence identity between the *nifP* and *cysE* gene products is also indicated (open box). aa, Amino acids.

typhimurium, SAT is normally complexed with *O*-acetylserine (thiol) lyase, which catalyzes the final reaction in cysteine biosynthesis (37). Thus, the amino-terminal portion of CYSE, with no counterpart in NIFP, may be important for formation of this complex. Likewise, the C-terminal extension of NIFP, with no counterpart in CYSE, may have an analogous role in a complex with a different protein, e.g., NIFV.

Though the requirement for *nifM* for the maturation and stability of NIFH has long been documented (25, 57), the precise role of NIFM remains elusive. Recent searches of the protein data bases also do not provide any obvious clues. The possibility that NIFM acts as a specific sulfur transferase involved in metal cluster insertion or assembly in the Fe-protein has been discussed elsewhere (52).

The *nifU*-to-*nifM* cluster in *A. chroococcum* may be transcribed as a single operon because *nifM* was expressed from the CAT gene promoter placed 3.8 kb upstream in pSAR2a (19), whereas *nifP* can be expressed from the same promoter placed 750 bp further upstream in pSAR3a. This finding suggests that there are no strong transcription termination signals in this region that are recognized in *K. pneumoniae* or *E. coli*. Therefore, the inverted repeat located between *nifS* and *nifV* appears not to prevent transcription of downstream genes at least in these backgrounds. Previous data also show that significant levels of expression of these gene products in enteric bacteria cannot proceed from internal promoters recognized by RNA polymerase containing the normal housekeeping sigma factor (σ^{70}) or the alternative sigma factor produced by *ntrA* (σ^{54}). The nucleotide sequence data shows that although there are potential *ntrA*-dependent promoters immediately upstream of *nifM* and *nifU*, neither is preceded at an appropriate distance by a $TGTN_{10}ACA$ upstream activator (9).

These potential promoter sequences appear conserved at similar locations in the corresponding *A. vinelandii* gene cluster, although in this organism an additional potential promoter together with an upstream activator is located upstream of ORF8 (28). In *A. vinelandii*, it is proposed that transcription of *nifU* through ORF6 (*nifP*) is initiated from a *nif* promoter situated upstream of a small ORF (ORF6) preceding *nifU*. ORF8 through *nifM* might be cotranscribed

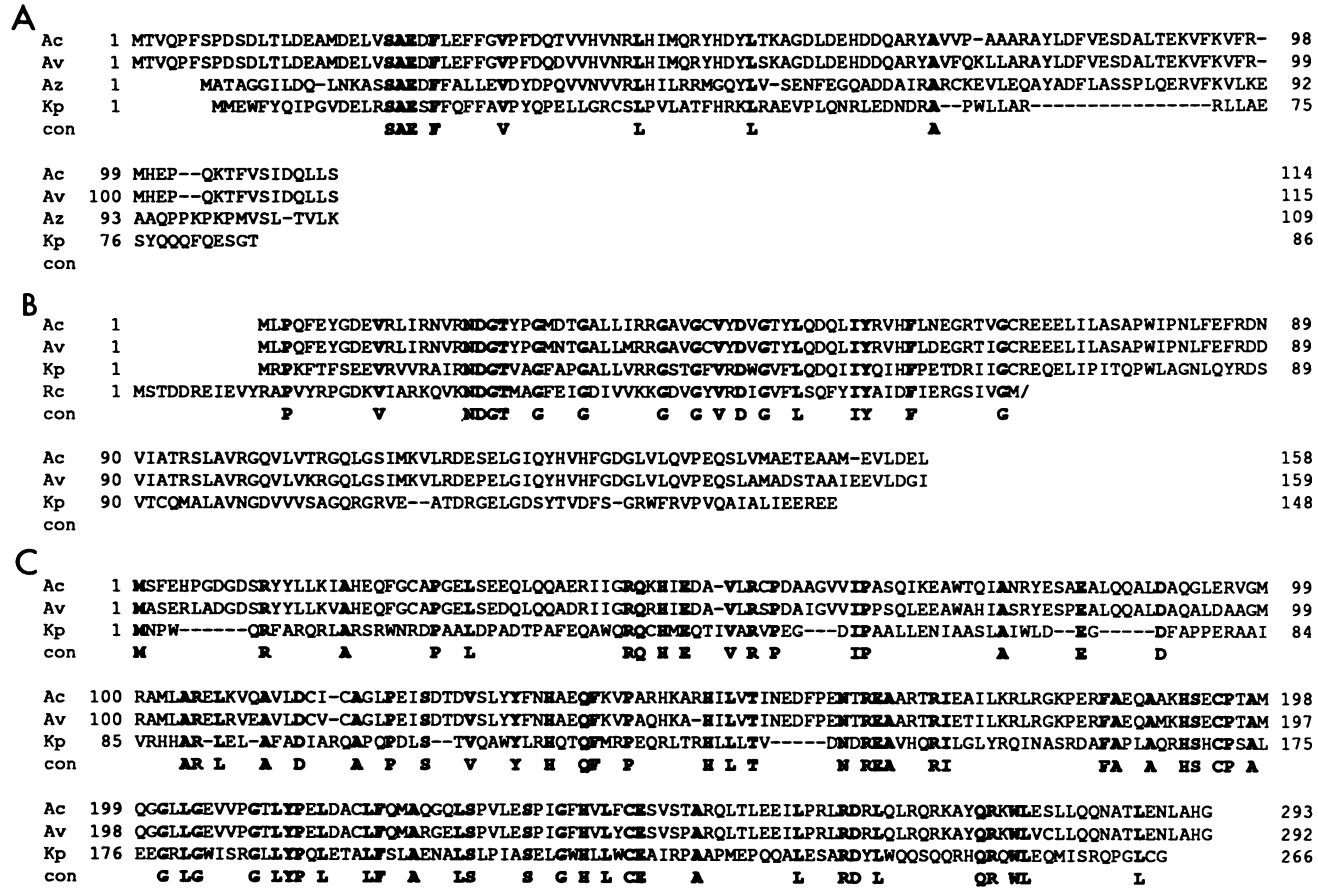

FIG. 7. Alignments of predicted amino acid sequences for *nifW* (A), *nifZ* (B), and *nifM* (C) gene products from *A. chroococcum* (Ac), *A. vinelandii* (Ac) (27), and *K. pneumoniae* (Kp) (2, 53). Also shown is an alignment between the *fixO* gene product from *Azorhizobium caulinodans* (Az) (34) and the *nifW* gene products. A partial ORF from *Rhodobacter capsulatus* (44) with a high degree of similarity to *nifZ* is shown in the *nifZ* alignments. In the multiple sequence alignments, fully conserved positions are noted in the consensus line (con) and printed in bold font.

with these genes or from the promoter immediately preceding ORF8. This arrangement may also allow boosting of expression of the distal genes under some conditions.

The apparently complex and potentially overlapping transcriptional organization of this gene cluster in the azotobacters mirrors the situation in *K. pneumoniae* (4, 54, 66). The reason for the apparently complex transcriptional arrangement in these two genera is unclear. In the case of the azotobacters, it may be linked to their possession of alternative nitrogenase systems. Transcription of the structural genes for the alternative vanadium nitrogenase in *A. chroococcum* requires an alternative activator (*vnfA*), and mutation of this gene does not block expression of the Mo-nitrogenase system (36). The possibility exists that if, as discussed earlier, the *nifUM* gene cluster is required for expression of both nitrogenase systems, then its transcription may be independent of both *nifA* and *vnfA* and depend instead upon a third, as yet unidentified, activator.

ACKNOWLEDGMENTS

We thank R. Dixon, C. Kennedy, M. Merrick, and R. Pau for useful discussions and A. L. Menon for critically reading the manuscript. We are particularly grateful to Wyatt Paul and Mike Merrick for making results available prior to publication.

REFERENCES

1. **Ambler, R. P., R. G. Bartsch, M. Daniel, M. D. Karmen, L. McLellan, T. E. Meyer, and J. van Beeumen.** 1981. Amino acid sequences of bacterial cytochromes c′ and c-556. Proc. Natl. Acad. Sci. USA **78**:6845–6857.
2. **Arnold, W., A. Rump, W. Klipp, U. B. Prifer, and A. Puhler.** 1988. Nucleotide sequence of a 24,206-base pair DNA fragment carrying the entire nitrogen fixation gene cluster of *Klebsiella pneumoniae*. J. Mol. Biol. **203**:715–738.
3. **Beltzer, J. P., L.-F. L. Chang, A. E. Hinkkanen, and G. B. Kohlhaw.** 1986. Structure of yeast Leu 4. The 5′ flanking region contains features that predict two modes of control and two productive translation starts. J. Biol. Chem. **261**:5160–5167.
4. **Beynon, J., M. Cannon, V. Buchanan-Wollaston, and F. Cannon.** 1983. The *nif* promoters of *Klebsiella pneumoniae* have a characteristic primary structure. Cell **34**:665–671.
5. **Beynon, J., A. Ally, M. Cannon, F. Cannon, M. Jacobson, V. Cash, and D. Dean.** 1987. Comparative organization of nitrogen fixation-specific genes from *Azotobacter vinelandii* and *Klebsiella pneumoniae*: DNA sequence of the *nifUSV* genes. J. Bacteriol. **169**:4024–4029.
6. **Biggin, M. D., J. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **77**:7342–7346.
7. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1525.

8. **Bishop, P. E., and R. D. Joerger.** 1990. Genetics and molecular biology of alternative nitrogenase systems. Annu. Rev. Plant Physiol. Plant Mol. Biol. **41**:109–125.
9. **Buck, M., S. Millar, M. Drummond, and R. Dixon.** 1986. Upstream activator sequences are present in the promoters of nitrogen fixation genes. Nature (London) **320**:374–378.
10. **Byrne, C. R., R. S. Monroe, K. A. Ward, and N. M. Kredich.** 1988. DNA sequences of the *cysK* regions of *Salmonella typhimurium* and *Escherichia coli* and linkage of the *cysK* regions to *ptsH*. J. Bacteriol. **170**:3150–3157.
11. **Dayhoff, M. O., R. E. Dayhoff, and L. T. Hunt.** 1976. Atlas of protein sequence and structure, vol. 5, suppl. 2. National Biomedical Research Foundation, Washington, D.C.
12. **Denk, D., and A. Böck.** 1987. L-Cysteine biosynthesis in *Escherichia coli*: nucleotide sequence and expression of the serine acetyl transferase (*cysE*) gene from the wild-type and a cysteine-excreting mutant. J. Gen. Microbiol. **133**:515–525.
13. **Dente, L., G. Cesareni, and R. Cortese.** 1983. pEMBL: a new family of single stranded plasmids. Nucleic Acids Res. **11**:1645–1655.
14. **Deveraux, J., P. Haeberli, and O. Smithies.** 1985. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
15. **Dixon, R., C. Kennedy, A. Kondorosi, V. Krishnapillai, and M. Merrick.** 1977. Complementation analysis of *Klebsiella pneumoniae* mutants defective in nitrogen fixation. Mol. Gen. Genet. **157**:189–198.
16. **Dixon, R. A.** 1984. The genetic complexity of nitrogen fixation. J. Gen. Microbiol. **130**:2745–2755.
17. **Downie, J. A.** 1989. The *nodL* gene from *Rhizobium leguminosarum* is homologous to the acetyl transferases encoded by *lacA* and *cysE*. Mol. Microbiol. **3**:1649–1651.
18. **Earl, C. D., C. W. Ronson, and F. M. Ausubel.** 1987. Genetic and structural analysis of the *Rhizobium meliloti fixA*, *fixB*, *fixC*, and *fixX* genes. J. Bacteriol. **169**:1127–1136.
19. **Evans, D., R. Jones, P. Woodley, and R. Robson.** 1988. Further analysis of nitrogen fixation genes in *Azotobacter chroococcum*: identification and expression in *Klebsiella pneumoniae* of *nifS*, *nifV*, *nifM*, and *nifB* genes and localization of *nifE*/*N*-, *nifU*-, *nifA*- and *fixABC*-like genes. J. Gen. Microbiol. **134**:931–942.
20. **Gold, L., D. Pribnow, T. Schneider, S. Shinedling, B. S. Singer, and G. Stormo.** 1981. Translation initiation in prokaryotes. Annu. Rev. Microbiol. **35**:365–403.
21. **Gribskov, M.** 1984. The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression. Nucleic Acids Res. **12**:539–549.
22. **Hediger, M. A., D. F. Johnson, D. P. Nierlich, and F. Zabin.** 1985. DNA sequence of the lactose operon: the *lacA* gene and the transcription termination region. Proc. Natl. Acad. Sci. USA **82**:6414–6418.
23. **Hoover, T. R., J. Imperial, P. W. Ludden, and V. K. Shah.** 1989. Homocitrate is a component of the iron-molybdenum cofactor of nitrogenase. Biochemistry **28**:2768–2771.
24. **Hoover, T. R., A. D. Robertson, R. L. Cerny, R. N. Hayes, J. Imperial, V. K. Shah, and P. W. Ludden.** 1987. Identification of the V factor needed for synthesis of the iron-molybdenum cofactor as homocitrate. Nature (London) **329**:855–857.
25. **Howard, K. S., P. A. McLean, F. B. Hansen, P. V. Lemley, K. S. Koblan, and W. H. Orme-Johnson.** 1986. *Klebsiella pneumoniae nifM* gene product is required for stabilization and activation of nitrogenase iron protein in *Escherichia coli*. J. Biol. Chem. **261**:772–778.
26. **Hryniewicz, M., A. Sirko, A. Palucha, A. Böck, and D. Hulanicka.** 1990. Sulfate and thiosulfate transport in *Escherichia coli* K-12: identification of a gene encoding a novel protein involved in thiosulfate binding. J. Bacteriol. **172**:3358–3366.
27. **Jacobson, M. R., K. E. Brigle, L. T. Bennett, R. A. Setterquist, M. S. Wilson, V. L. Cash, J. Beynon, W. E. Newton, and D. R. Dean.** 1989. Physical and genetic map of the major *nif* gene cluster from *Azotobacter vinelandii*. J. Bacteriol. **171**:1017–1027.
28. **Jacobson, M. R., V. L. Cash, M. C. Weiss, N. F. Laird, W. E. Newton, and D. R. Dean.** 1989. Biochemical and genetic analysis of the *nifUSVWZM* cluster from *Azotobacter vinelandii*. Mol. Gen. Genet. **219**:49–57.
29. **Joerger, R. D., and P. E. Bishop.** 1988. Nucleotide sequence and genetic analysis of the *nifB* to *nifQ* region from *Azotobacter vinelandii*. J. Bacteriol. **170**:1475–1487.
30. **Jones, R., and E. Lander.** Unpublished data.
31. **Jones, R., W. Taylor, X. Zhang, J. P. Mesirov, and E. Lander.** 1989. Protein sequence comparison on the Connection Machine CM-2, p. 99–107. *In* G. Bell and T. Marr (ed.), Computers and DNA. SFI studies in the science of complexity, vol. VII. Addison-Wesley Publishing Co., Redwood City, Calif.
32. **Jones, R., P. Woodley, and R. Robson.** 1984. Cloning and organization of some genes for nitrogen fixation from *Azotobacter chroococcum* and their expression in *Klebsiella pneumoniae*. Mol. Gen. Genet. **197**:318–327.
33. **Jones-Mortimer, M. C.** 1968. Positive control of sulphate reduction in *Escherichia coli*. Isolation, characterization and mapping of cysteineless mutants of *E. coli* K12. Biochem. J. **110**:589–595.
34. **Kaminsky, P. A., F. Norel, N. Desnoues, A. Kush, G. Salzano, and C. Elmerich.** 1988. Characterization of the *fixABC* region of *Azorhizobium caulinodans* ORS571 and identification of a new nitrogen fixation gene. Mol. Gen. Genet. **214**:496–502.
35. **Kennedy, C., R. Gamal, R. Humphrey, J. Ramos, K. Brigle, and D. Dean.** 1986. The *nifH*, *nifM* and *nifN* genes of *Azotobacter vinelandii*: characterization by Tn*5* mutagenesis and isolation from pLAFR1 gene banks. Mol. Gen. Genet. **205**:318–325.
36. **Kerber, L. L., and R. L. Robson.** Unpublished data.
37. **Kredich, N. M.** 1987. Biosynthesis of cysteine, p. 419–428. *In* F. C. Neidhardt, J. L. Ingraham, B. Magasanik, K. B. Low, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
38. **Kredich, N. M., and G. M. Tomkins.** 1966. The enzymic synthesis of L-cysteine in *Escherichia coli* and *Salmonella typhimurium*. J. Biol. Chem. **241**:4955–4965.
39. **Lederberg, E. M., and S. N. Cohen.** 1974. Transformation of *Salmonella typhimurium* by plasmid deoxyribonucleic acid. J. Bacteriol. **119**:1072–1074.
40. **Levy, S., and A. Danchin.** 1988. Phylogeny of metabolic pathways: O-acetylserine sulphydrylase A is homologous to the tryptophan synthase beta subunit. Mol. Microbiol. **2**:777–783.
41. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
42. **Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall.** 1951. Protein measurement with the Folin phenol reagent. J. Biol. Chem. **193**:265–275.
43. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
44. **Masepohl, B., W. Klipp, and A. Puhler.** 1988. Genetic characterization and sequence analysis of the duplicated *nifA*/*nifB* gene region of *Rhodobacter capsulatus*. Mol. Gen. Genet. **212**:27–37.
45. **McLean, P. A., and R. A. Dixon.** 1981. Requirement of *nifV* for production of wild-type nitrogenase enzyme in *Klebsiella pneumoniae*. Nature (London) **292**:655–656.
46. **Merrick, M., M. Filser, R. Dixon, C. Elmerich, L. Sibold, and J. Houmard.** 1980. The use of translocatable genetic elements to construct a fine-structure map of the *Klebsiella pneumoniae* nitrogen fixation (*nif*) gene cluster. J. Gen. Microbiol. **117**:509–520.
47. **Merrick, M., M. Filser, C. Kennedy, and R. A. Dixon.** 1978. Polarity of mutations induced by insertion of transposons Tn*5*, Tn*7* and Tn*10* into the *nif* gene cluster of *Klebsiella pneumoniae*. Mol. Gen. Genet. **165**:103–111.
48. **Merrick, M. J., J. R. Gibbins, and J. R. Postgate.** 1987. A rapid and efficient method for plasmid transformation of *Klebsiella pneumoniae* and *Escherichia coli*. J. Gen. Microbiol. **133**:2053–2057.
49. **Messing, J., B. Groneborn, B. Muller-Hill, and P. H. Hofscheider.** 1977. Filamentous coliphage M13 as a cloning vehicle: insertion of a *Hin*dIII fragment of the regulatory region in M13 replicative form in vitro. Proc. Natl. Acad. Sci. USA **74**:3642–3646.

50. **Mulligan, M. E., and R. Haselkorn.** 1989. Nitrogen-fixation (nif) genes of the cyanobacterium *Anabaena* sp. strain PCC7120: the *nif fdxN-nifS-nifU* operon. J. Biol. Chem. **264**:19200–19207.
51. **Oppenheim, D. S., and C. Yanofsky.** 1980. Translational coupling during expression of the tryptophan operon of *Escherichia coli*. Genetics **95**:785–795.
52. **Pagani, S., M. Eldridge, and R. R. Eady.** 1987. Nitrogenase of *Klebsiella pneumoniae*: rhodanese catalysed restoration of activity of the inactive 2Fe species of the Fe-protein. Biochem. J. **244**:485–488.
53. **Paul, W., and M. Merrick.** 1987. The nucleotide sequence of the *nifM* gene of *Klebsiella pneumoniae* and identification of a new gene: *nifZ*. Eur. J. Biochem. **170**:259–265.
54. **Paul, W., and M. Merrick.** 1989. The roles of the *nifW*, *nifZ*, and *nifM* genes of *Klebsiella pneumoniae* in nitrogenase biosynthesis. Eur. J. Biochem. **178**:675–682.
55. **Ploegman, J. H., G. Drent, K. H. Kalk, W. G. J. Hol, R. L. Heinrickson, P. Keim, L. Weng, and J. Russell.** 1978. The covalent and tertiary structure of bovine liver rhodanese. Nature (London) **273**:124–129.
56. **Ricca, E., and J. M. Calvo.** 1990. The nucleotide sequence of *leuA* from *Salmonella typhimurium*. Nucleic Acids Res. **18**: 1290.
57. **Roberts, G. P., T. MacNeil, D. MacNeil, and W. J. Brill.** 1978. Regulation and characterization of protein products coded by the *nif* (nitrogen fixation) genes of *Klebsiella pneumoniae*. J. Bacteriol. **166**:180–186.
58. **Robson, R. L.** 1986. Nitrogen fixation in strains of *Azotobacter chroococcum* bearing deletions of a cluster of genes coding for nitrogenase. Arch. Microbiol. **146**:74–79.
59. **Robson, R. L., J. A. Chesshyre, C. Wheeler, R. Jones, P. R. Woodley, and J. R. Postgate.** 1984. Genome size and complexity in *Azotobacter chroococcum*. J. Gen. Microbiol. **130**:1603–1612.
60. **Robson, R. L., R. R. Eady, T. H. Richardson, R. W. Miller, M. Hawkins, and J. R. Postgate.** 1986. The alternative nitrogenase of *Azotobacter chroococcum* is a vanadium enzyme. Nature (London) **322**:388–390.
61. **Robson, R. L., and J. R. Postgate.** 1980. Oxygen and hydrogen in biological nitrogen fixation. Annu. Rev. Microbiol. **34**:183–207.
62. **Robson, R. L., P. Woodley, and R. Jones.** 1986. Second gene (*nifH**) coding for a nitrogenase iron protein is adjacent to a gene coding for a ferredoxin-like protein. EMBO J. **5**:1159–1163.
63. **Robson, R. L., P. R. Woodley, R. N. Pau, and R. R. Eady.** 1989. Structural genes for the vanadium nitrogenase from *Azotobacter chroococcum*. EMBO J. **8**:1217–1224.
64. **Sanger, F., S. Nicklen, and A. R. Coulsen.** 1977. DNA sequencing with chain termination inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
65. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementary to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
66. **Sibold, L.** 1982. The polar effect on *nifM* of mutations in the *nifU*, *-S*, *-V* genes of *Klebsiella pneumoniae* depends on their plasmid or chromosomal location. Mol. Gen. Genet. **186**:569–571.
67. **Sirko, A., M. Hryniewicz, D. Hulanicka, and A. Böck.** 1990. Sulfate and thiosulfate transport in *Escherichia coli* K-12: nucleotide sequence and expression of the *cysTWAM* gene cluster. J. Bacteriol. **172**:3351–3357.
68. **Smith, T. F., and M. S. Waterman.** 1981. Identification of common molecular subsequences. J. Mol. Biol. **147**:195–197.
69. **Staden, R.** 1982. Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. Nucleic Acids Res. **10**:4731–4751.
70. **Strassman, M., and L. N. Ceci.** 1964. Enzymatic formation of homocitric acid, an intermediate in lysine biosynthesis. Biochem. Biophys. Res. Commun. **14**:262–267.
71. **Thuring, R. W. J., J. P. M. Sander, and P. Borst.** 1975. A freeze-squeeze method for recovering DNA from agarose gels. Anal. Biochem. **66**:213–220.
72. **Tucci, A. F., and L. N. Ceci.** 1972. Homocitrate synthase from yeast. Arch. Biochem. Biophys. **153**:742–750.
73. **Ugalde, R. A., J. Imperial, V. K. Shah, and W. J. Brill.** 1985. Biosynthesis of the iron-molybdenum cofactor and the molybdenum cofactor in *Klebsiella pneumoniae*: effect of sulfur source. J. Bacteriol. **164**:1081–1087.
74. **Umbarger, H. E.** 1987. Biosynthesis of the branched-chain amino acids, p. 352–367. *In* F. C. Neidhardt, J. L. Ingraham, B. Magasanik, K. B. Low, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
75. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
76. **White, R. H.** 1983. Origin of the labile sulfide in the iron-sulfur proteins of *Escherichia coli*. Biochem. Biophys. Res. Commun. **112**:66–72.
77. **Yates, M. G., and K. Planqué.** 1975. Nitrogenase from *Azotobacter chroococcum*. Purification and properties of the component proteins. Eur. J. Biochem. **60**:467–476.
78. **Yates, M. G., and R. L. Robson.** 1985. Mutants of *Azotobacter chroococcum* defective in hydrogenase activity. J. Gen. Microbiol. **131**:1459–1466.