JOURNAL OF BACTERIOLOGY, May 1987, p. 1911–1916
0021-9193/87/051911-06$02.00/0

Vol. 169, No. 5

Cloning and Nucleotide Sequence of the *chlD* Locus

STEPHEN JOHANN[1]† AND STEPHEN M. HINTON[2]‡*

Cold Spring Harbor Laboratories[1] and Exxon Liaison, Inc.,[2] Cold Spring Harbor, New York 11724

Received 7 August 1986/Accepted 11 February 1987

The nucleotide sequence of a *Sau*3A1 restriction nuclease fragment that complemented an *Escherichia coli chlD*::Mu *cts* mutant strain was determined. DNA and deduced amino acid sequence analysis revealed two open reading frames (ORFs) that potentially codes for proteins with amino acid sequence homology with binding protein-dependent transport systems. One of the ORFs showed a sequence that encoded a protein with properties that were characteristic of a hydrophobic inner membrane protein. The other ORF, which was responsible for complementing a *chlD* mutant, encoded a protein with conserved sequences in nucleotide-binding proteins and hydrophilic inner membrane proteins in active transport systems. A proposal that the *chlD* locus is the molybdate transport operon is discussed in terms of the *chlD* phenotype.

Under anaerobic conditions in the presence of nitrate, *Escherichia coli* synthesizes a membrane-bound multienzyme complex that is composed of formate dehydrogenase, cytochrome b_{556}, and nitrate reductase (23). The formate-nitrate reductase system is the preferred anaerobic respiration pathway in which formate is oxidized and nitrate is the terminal electron acceptor. Mutants that are selected for chlorate resistance, designated *chl* or *nar*, lost the activity of the formate-nitrate reductase system. The *chlA*, *-B*, *-D*, *-E*, and *-G* mutants are pleiotropic in that they lack nitrate reductase as well as other molybdoenzyme activities. The *nar* mutants have mutations in the nitrate reductase operon and exhibit altered expression of nitrate reductase (28).

The *chlA*, *-B*, *-D*, *-E*, and *-G* mutants have been implicated in the synthesis of the molybdenum cofactor, a molybdopterin species of undefined structure (18), that is common to all molybdoenzymes, with the exception of nitrogenase. The *chlA* and *-E* gene products are thought to be involved in the biosynthesis of the pterin, while the *chlB*, *-D*, and *-G* gene products appear to be required for the processing of molybdenum. The phenotype of the *chlD* and *-G* mutants can be suppressed by the addition of high concentrations (1 mM) of molybdate to the growth medium (9). The results suggest that the *chlD* and *-G* gene products are involved in molybdate transport, processing, or insertion of molybdenum into the cofactor structure, and the high levels of MoO_4^{2-} presumably circumvent the role of these gene products. These observations have prompted us to investigate the *chlD* locus.

We determined the DNA sequence of a *Sau*3A1 restriction nuclease fragment of *E. coli* DNA that complements an *E. coli chlD*::Mu *cts* mutant (28). Two open reading frames (ORFs) were identified in a 1,609-base-pair (bp) restriction fragment, and preliminary evidence based on DNA and amino acid sequence analysis suggests that the *chlD* locus is an operon that is involved in the active transport of molybdenum.

MATERIALS AND METHODS

Organisms and growth media. All strains used in this study were derivatives of *E. coli* K-12. RK5202 (*chlD*::Mu *cts*) is a chlorate-resistant derivative of RK4353 (28). All bacteria were grown at 30°C and PN medium (28) was used. For solid media, purified agar (Difco Labortories, Detroit, Mich.) was added, with a final agar concentration of 1.6% (wt/vol). Carbenicillin was added separately as a sterile solution at a final concentration of 50 μg/ml. Plates were incubated anaerobically in anaerobic chambers (GasPak; Becton Dickinson and Co., Paramus, N.J.) under an atmosphere of H_2 and CO_2.

DNA manipulations. Restriction enzyme digestion, restriction fragment isolation with DEAE membrane (NA-45; Schleicher & Schuell, Inc., Keene, N.H.), and ligation of DNA with T4 DNA ligase were performed as described by the vendor. The restriction nuclease sites were mapped as described previously (14).

Construction of the *E. coli* genomic library. *E. coli* chromosomal DNA was partially digested with the restriction endonuclease *Sau*3A1. The different-sized restriction nuclease fragments were separated in a 5 to 20% (wt/vol) sucrose density gradient containing 1 M NaCl, 20 mM Tris hydrochloride (pH 7.5), and 1 mM sodium EDTA (14). The fractions from the sucrose gradient containing DNA fragments 2 to 4, 4 to 6, 6 to 9, and 9 to 12 kilobases (kb) were identified by agarose gel electrophoresis. Each of the four size-selected *Sau*3A1 restriction nuclease fragments (3 μg) were separately ligated into the expression vector plasmid pUC9 (1 μg) that had been linearized at the *Bam*HI site and dephosphorylated with bacterial alkaline phosphatase (preventing self-circularization). The four genomic libraries of *E. coli* chromosomal DNA with inserts with different sizes were used to transform the *E. coli* mutant strain RK5202 (*chlD*::Mu *cts* [28]) that was made competent (11).

Genetic complementation test. Clones complementing RK5202, a *chlD* mutant strain, were selected by plating the transformed bacterial colonies on PN agar containing carbenicillin, and the plates were incubated anaerobically at 30°C for 3 days. Several colonies from each plate (representing a genomic library of determined-size fragments) were purified on LB agar containing carbenicillin and then tested for temperature sensitivity and autoinduction to detect the presence of phage Mu *cts*. The plasmid from each clone was isolated by the alkaline lysis method (2) and analyzed by agarose gel electrophoresis. The plasmids containing the smallest and the largest inserts of *E. coli* DNA complementing the *chlD* mutant were retained for study.

* Corresponding author.
† Present address: Lederle Laboratories, Pearl River, NY 10965.
‡ Present address: Exxon Corporate Research Co., Annandale, NJ 08801.

DNA sequence determination. DNA sequencing was performed as described previously (14) with the following sequencing strategy. The insert of *E. coli* DNA in plasmid pSJE301 was isolated and digested into smaller DNA fragments with the restriction endonuclease *Sau*3A1, and then the fragments that were ligated into the *Bam*HI site of M13 mp18 were used for DNA sequencing analysis. Unique cloned *Sau*3A1 fragments were identified by their DNA sequence, which was used to determine the sequence of complementary synthetic oligonucleotides (15mer) DNA sequencing primers. The synthetic oligonucleotide primers were used to extend the sequence of each restriction fragment in both directions by using the entire insert of *E. coli* DNA as a template for dideoxy chain-termination reactions (24). This sequencing strategy resulted in approximately 500 bp of DNA sequence from each *Sau*3A1 clone.

Subcloning procedures. pSJE301 and subclones of pSJE301 were constructed to define the region complementing the *chlD*::Mu *c*ts mutant (RK5202). pSJE301 was obtained by cleaving pSJE300 at the *Eco*RI and *Hin*dIII restriction sites at the ends of the pUC9 polylinker and then isolating the 1.6-kb cloned insert of *E. coli* DNA by using standard procedures with the DEAE membrane NA-45. The isolated 1.6-kb restriction nuclease fragment was ligated into the *Eco*RI and *Hin*dIII restriction sites of plasmid pBR322. pSJE310 was obtained by cleaving the isolated restriction fragment at the *Bal*I site and ligating it into the vector pBR322, which was cleaved at the *Hin*dIII, *Bam*HI, and *Bal*I sites. pSJE311 was constructed by cleaving the isolated 1.6-kb insert of *E. coli* DNA at the *Xmn*I site and then ligating it into pBR322, which was cleaved at the *Hin*dIII, *Bam*HI, and *Pvu*II restriction sites. pSJE312 was obtained by cleaving the 1.6-kb restriction fragment at the *Bal*I site and then by ligating it into the vector pBR322, which was cleaved at the *Eco*RI and *Bal*I sites. pSJE313 was constructed by cleaving the insert of *E. coli* DNA at the *Acc*I site and then by ligating it into the vector pBR322, which was cleaved at the *Eco*RI and *Acc*I sites. All of the subclones described above were transformed into competent *E. coli* DH-1, and then they were selected for ampicillin resistance and subsequently screened for tetracycline sensitivity. Plasmids from the clones with the correct antibiotic phenotype were isolated by the alkaline lysis method (2) and analyzed by agarose gel electrophoresis. DNA-DNA hybridization was performed with all derivatives as described previously (14) to ensure that they contained the proper insert of *E. coli* DNA. The ability of each subclone to complement the *E. coli* *chlD*::Mu *c*ts mutant RK5202 was tested by plating transformed RK5202 colonies on PN agar (28) containing carbenicillin, and the plates were incubated anaerobically at 30°C for 3 days.

Computer analysis. DNA sequences were analyzed with programs that are available at Cold Spring Harbor Laboratories (Cold Spring Harbor, N.Y.) (19). Amino acid sequences were analyzed by using the secondary structure algorithm described by Chou and Fasman (3), and the hydropathy profile was predicted by the method described by Kyte and Doolittle (20).

RESULTS

Determination of the *chlD* DNA sequence. The insert of *E. coli* DNA in the pUC9 expression vector (pSJE300) that complemented the *E. coli* mutant strain RK5202 (*chlD*::Mu *c*ts) was subcloned into the *Eco*RI and *Hin*dIII sites of pBR322 (pSJE301). Both strands of the 1,609-bp insert of *E. coli* DNA in pSJE301 were sequenced independently more than once (Fig. 1).

Localization and DNA sequence of the *chlD* gene. The DNA sequence of the insert in pSJE301 was surveyed for ORFs as possible candidates for the *chlD* gene. The physical map of the insert of *E. coli* DNA in pSJE301 that was derived from the DNA sequence was used to determine convenient restriction sites for subcloning the various ORFs (Fig. 2). The subclones of the complementing fragment were ligated into pBR322 (Fig. 2) and introduced into an *E. coli chlD*::Mu *c*ts mutant by transformation. The strains were then tested for anaerobic growth on nitrate. pSJE312 and pSJE313 were able to complement the *E. coli chlD*::Mu *c*ts mutant, which suggests that the region contains the functional *chlD* gene.

The largest ORF in the region containing the functional *chlD* gene was initiated by a ATG codon at position 607 and was followed by 897 nucleotides until a nonsense codon (TAA) was reached at position 1507. The other ORFs (in the other phases in either direction) are relatively short and coded for relatively small peptides. A possible Shine-Dalgarno ribosome recognition sequence (26) at positions 594 and 598 (CGGGG) is 9 bases upstream from the proposed initiation codon within the putative *chlD* gene. The codon usage in the putative *chlD* gene gave an *f* index (frequency of codon preference [16]) of 0.64, which fell within the range of moderately expressed proteins in *E. coli*.

The ChlD protein. The large ORF initiated by ATG at position 607 encoded for a polypeptide of 300 amino acids, with a calculated molecular weight of 33,671. The hydropathy profile, the charged amino acid distribution, and the predicted secondary structure of the ChlD protein were determined. The average hydropathy of ChlD was determined to be −0.15, which falls in the range of hydropathy index values in which hydrophilic membrane proteins are found (20). The ChlD protein contains 45% nonpolar amino acids, 33% polar amino acids, and 22% charged amino acids (37 of 67 charged amino acids are basic). The charged amino acid residues were evenly distributed, which was not a characteristic of hydrophobic membrane proteins. The secondary structure predictions show that the ChlD protein conformation appears to be symmetrical. Both ends of the protein are predicted to be in β-sheet conformation interrupted by β-turns with an α-helix stretch in the middle of the protein.

The DNA sequence flanking of *chlD* gene. The *chlD* gene was flanked on both sides by ORFs that were separated by short intergenic regions (at positions 605 to 606 and 1510 to 1514; Fig. 1). This type of organization strongly suggests that *chlD* is part of a polycistronic operon (4, 7). It has been proposed that short intergenic regions are important for the coordinate expression of macromolecules with multiple subunits or a complex in which the components must be made in equivalent amounts.

DISCUSSION

The amino acid sequence of ChlD, which was deduced from the nucleotide sequence, was aligned with the amino acid sequences of OppD (from the oligopeptide transport system [13]), HisP (from the histidine transport system [12]), and MalK (from the maltose transport system [8]) proteins to show the extent of sequence homology (Fig. 3). These three proteins that are being compared with ChlD are hydrophilic inner membrane proteins belonging to the binding protein-dependent transport systems. The degree of sequence homology between the proteins was the greatest in the middle

```
        ILE PHE PHE ALA TRP LEU LEU VAL ARG CYS SER PHE PRO GLY LYS ALA LEU LEU ASP SER VAL LEU HIS LEU PRO LEU VAL LEU PRO PRO
5'....G ATC TTT TTT GCC TGG TTA CTG GTG CGT TGC AGC TTT CCG GGC AAA GCT CTG CTC GAC AGC GTA CTG CAT CTA CCG CTG GTG TTA CCG CCC
           10          20          30          40          50          60          70          80          90

VAL VAL VAL GLY TYR LEU LEU LEU VAL SER MET GLY ARG ARG GLY PHE ILE GLY GLU ARG LEU TYR ASP TRP PHE GLY ILE THR PHE ALA PHE SER
GTG GTC GTC GGT TAC TTA TTA TTA GTT TCG ATG GGA CGG CGC GGA TTT ATC GGT GAA CGT CTG TAT GAC TGG TTT GGT ATT ACC TTC GCC TTT AGC
     100         110         120         130         140         150         160         170         180

TRP ARG GLY ALA VAL LEU ALA ALA ALA VAL MET SER PHE PRO LEU MET VAL ARG ALA ILE ARG LEU ALA LEU GLU GLY VAL ASP VAL LYS LEU GLU
TGG CGC GGC GCG GTT CTC GCT GCC GCC GTC ATG TCG TTT CCG CTG ATG GTG CGG GCA ATT CGT CTG GCG CTG GAA GGG GTT GAT GTC AAA CTG GAA
190         200         210         220         230         240         250         260         270         280

GLN ALA ALA ARG THR LEU GLY ALA GLY ARG TRP ARG VAL PHE PHE THR ILE THR LEU PRO LEU THR LEU PRO GLY ILE ILE VAL GLY THR VAL LEU
CAG GCC GCA AGA ACA CTG GGG GCC GGG CGC TGG CGC GTT TTC TTT ACT ATC ACG TTA CCG CTG ACC TTA CCG GGA ATT ATT GTT GGT ACG GTA CTG
    290         300         310         320         330         340         350         360         370

ALA PHE ALA ARG SER LEU GLY GLU PHE GLY ALA THR ILE THR PHE VAL SER ASN ILE PRO GLY GLU THR ARG THR ILE PRO SER ALA MET TYR THR
GCT TTT GCT CGT TCT CTC GGT GAG TTT GGT GCA ACC ATC ACC TTT GTG TCG AAC ATT CCT GGT GAA ACG CGA ACC ATT CCT TCT GCC ATG TAT ACC
        390         400         410         420         430         440         450         460         470

LEU ILE GLN THR PRO GLY GLY GLU SER GLY ALA ALA ARG LEU CYS ILE ILE SER ILE ALA LEU ALA MET ILE SER LEU LEU ILE SER GLU TRP LEU
CTG ATC CAG ACC CCC GGC GGG GAA AGT GGA GCG GCG AGA CTG TGC ATT ATT TCT ATT GCG CTG GCG ATG ATC TCC CTG TTG ATT TCA GAA TGG CTG
    480         490         500         510         520         530         540         550         560         570

ALA ARG ILE SER ARG GLU ARG ALA GLY ARG ••• ···  MET LEU GLU LEU ASN PHE SER GLN THR LEU GLY ASN HIS CYS LEU THR ILE ASN GLU THR
GCC AGA ATC AGC CGT GAA CGG GCG GGG CGC TAA TC  ATG CTG GAA CTG AAT TTT TCC CAG ACG TTG GGC AAC CAT TGC CTG ACT ATT AAT GAA ACG
        580         590         600         610         620         630         640         650         660

LEU PRO ALA ASN GLY ILE THR ALA ILE PHE GLY VAL SER GLY ALA GLY LYS THR SER LEU ILE ASN ALA ILE SER GLY LEU THR ARG PRO GLN LYS
CTG CCC GCC AAT GGC ATC ACT GCT ATC TTT GGC GTC TCC GGT GCC GGA AAA ACT TCG CTG ATT AAC GCC ATC AGT GGA CTG ACG CGC CCG CAA AAA
    670         680         690         700         710         720         730         740         750         760

GLY ARG ILE VAL LEU ASN GLY ARG VAL LEU ASN ASP ALA GLU LYS GLY ILE CYS LEU THR PRO GLY LYS ARG ARG VAL GLY TYR VAL PHE GLN ASP
GGG CGG ATT GTC CTC AAT GGG CGG GTA CTA AAT GAT GCC GAA AAA GGT ATC TGC CTG ACG CCG GAA AAG CGT CGC GTT GGC TAT GTT TTT CAG GAT
        770         780         790         800         810         820         830         840         850

ALA ARG LEU PHE PRO HIS TYR LYS VAL ARG GLY ASN LEU ARG TYR GLY MET SER LYS SER MET VAL ASP GLN PHE ASP LYS LEU VAL ALA LEU LEU
GCG CGG CTG TTC CCG CAT TAC AAA GTG CGT GGC AAT CTG CGC TAC GGC ATG TCG AAA AGT ATG GTC GAT CAG TTC GAT AAG CTG GTG GCG CTT TTA
    870         880         890         900         910         920         930         940         950

GLY ILE GLU PRO LEU LEU ASP ARG LEU PRO GLY SER LEU SER GLY GLY GLU LYS GLN ARG VAL ALA ILE GLY ARG ALA LEU LEU THR ALA PRO GLU
GGC ATT GAA CCG TTG CTT GAC CGT TTA CCA GGC AGC CTG TCC GGA GGC GAA AAA CAG CGC GTG GCG ATT GGT CGG GCT TTG CTG ACA GCA CCG GAA
    960         970         980         990        1000        1010        1020        1030        1040        1050

LEU LEU LEU LEU ASP GLU PRO LEU ALA SER LEU ASP ILE PRO ARG LYS ARG GLU LEU LEU PRO TYR LEU GLN ARG LEU THR ARG GLU ILE ASN ILE
TTG CTG TTG CTG GAT GAA CCG CTG GCG TCA CTG GAT ATT CCG CGT AAA CGC GAA CTG TTG CCT TAT CTG CAA CGG CTG ACA CGG GAA ATC AAC ATT
        1060        1070        1080        1090        1100        1110        1120        1130        1140

PRO MET LEU TYR VAL SER HIS SER LEU ASP GLU ILE LEU HIS LEU ALA ASP ARG VAL MET VAL LEU GLU ASN GLY GLN VAL LYS ALA PHE GLY ALA
CCG ATG TTG TAT GTC AGC CAT TCG CTG GAT GAG ATC CTC CAT CTG GCA GAC AGA GTG ATG GTA CTG GAA AAC GGT CAG GTG AAA GCC TTT GGC GCG
    1150        1160        1170        1180        1190        1200        1210        1220        1230        1240

LEU GLU GLU VAL TRP GLY SER SER VAL MET ASN PRO TRP LEU PRO LYS GLU GLN GLN SER SER ILE LEU LYS VAL THR VAL LEU GLU HIS HIS SER
CTG GAG GAA GTG TGG GGC AGT AGC GTG ATG AAT CCG TGG CTG CCG AAA GAG CAA CAA AGT AGC ATT CTG AAA GTG ACG GTG CTT GAG CAT CAT TCC
        1250        1260        1270        1280        1290        1300        1310        1320        1330

ALA LEU ARG ASP ASP ARG LEU ALA LEU GLY ASP GLN HIS LEU TRP VAL ASN LYS LEU ASP GLU PRO LEU GLN LEU ARG TYR TYR PRO HIS SER GLY
GCA TTA CGC GAT GAC CGG CTG GCG CTG GGC GAT CAG CAT TTG TGG GTC AAT AAG CTG GAC GAA CCG CTG CAA CTG CGC TAC TAT CCG CAT TCA GGC
        1350        1360        1370        1380        1390        1400        1410        1420        1430

PHE ARG CYS PHE SER TRP PHE TYR ASN ARG ARG SER LYS PRO ALA PHE VAL THR ILE ALA GLY LYS SER CYS ••• ••• ···  MET THR THR THR ALA
TTC CGA TGT TTC TCT TGG TTT TAC AAC CGC CGC AGC AAA CCA GCA TTC GTA ACT ATT GCG GGC AAA AGT TGT TAA TAG  TT ATG ACG ACA ACG GCC
    1440        1450        1460        1470        1480        1490        1500        1510        1520

ARG TRP LYS TRP ASN TRP LYS SER ALA VAL LYS ARG CYS GLY ARG VAL SER ALA ARG GLY PRO GLY MET ASN TRP ARG
AGG TGG AAG TGG AAC TGG AAG TCG GCG GTA AAA CGC TGT GGG CGC GTA TCA GCC CGT GGG CCA GGG ATG AAC TGG CGA TC....3'
        1540        1550        1560        1570        1580        1590        1600        1609
```

FIG. 1. Nucleotide sequence of the 1,609-bp *Sau*3A1 restriction nuclease fragment in pSJE301 and the deduced amino acid sequence of the ORFs representing the C-terminal portions of *chlJ* and *chlD*. The initiation codon (ATG) for the *chlD* gene at position 607 starts an ORF that encodes for 300 amino acids up to the termination codon (TAA) at position 1507.

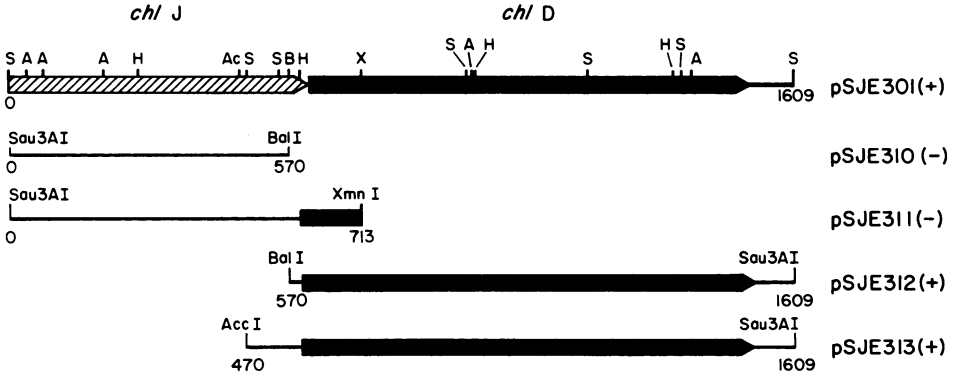

FIG. 2. Correlated physical and genetic map of the *E. coli* DNA complementing a *chlD*::Mu cts mutation. The restriction enzyme cleavage map of the 1,609-bp insert in the plasmid pSJE301 is shown with the location of the ORF, (*chlJ* and -*D*). The solid arrow represents the putative *chlD* gene, and the hatched arrow represents the C-terminal region of the putative *chlJ* gene. The restriction enzyme cut sites are designated as follows: A, *Alu*I; Ac, *Acc*I; B, *Bal*I; H, *Hin*dIII; S, *Sau*3A1; X, *Xmn*I. The large arrows show the potential ORFs with the first (ATG) initiation codon and all termination codons. The subclones of the plasmid pSJE301 are designated pSJE310, pSJE311, pSJE312, and pSJE313, with each segment of cloned DNA indicated by the bracketed line. The *chlD* gene or a portion of it is represented by the solid black arrow. The result of the ability of each subclone to complement the *E. coli* mutant strain RK5202 (*chlD*::Mu cts) is also indicated (+ or −) to the right of each construct.

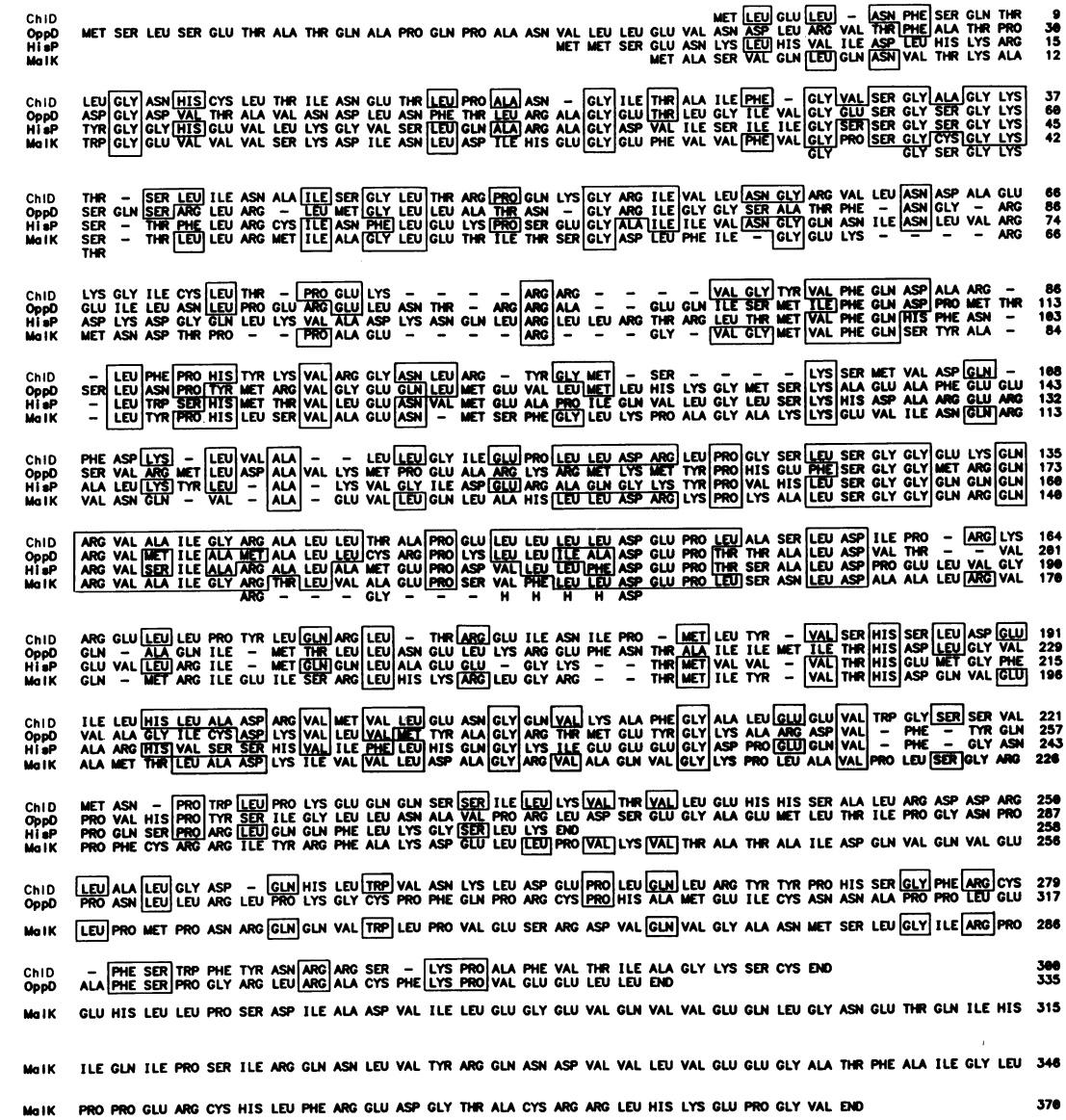

FIG. 3. Aligned amino acid sequences of the ChlD, OppD (12), HisP (15), and MalK (8) proteins. The amino acid sequences were aligned (SEQHP program) for best fit based on the Danhoff protein atlas with a gap penalty of 8 (10). The identical amino acid matches with the ChlD protein are boxed. The alignment of OppD, HisP, and MalK proteins was taken from Higgins et al. (12). The consensus sequence found in nucleotide-binding proteins is shown below the aligned sequence, and H in the consensus sequence represents a conserved hydrophobic residue.

region. It has been proposed (13) that the conserved sequence is a result of a common evolutionary origin for these transport systems, as well as constraints on structure and function. The C-terminal region probably lacks sequence homology because of a specific binding function in each system. The regions of these proteins with extensive sequence homology are consensus amino acid sequences, which have been speculated to be required for the formation of the nucleotide-binding fold (29). Recently, the HisP and MalK proteins have been shown to react with an ATP analog (15), and an OppD-LacZ fusion protein was shown to bind to a Cibacron blue affinity column (which retains ATP-binding proteins) and to react with an analog of ATP (13). The evidence suggests that these proteins have a common nucleotide-binding site.

The amino acid sequence of the ORF upstream from the *chlD* gene is shown in Fig. 1. The cloning of the *Sau*3A1 fragment appears to have interrupted a gene, and the deduced amino acid sequence that is presented (Fig. 1) represents the C-terminal region of the protein. The presence of a conserved sequence (~90 residues from the C terminus) that was found in the ChlJ protein, as well as other hydrophobic inner membrane proteins from several binding protein-dependent transport systems, is shown in Fig. 4. The hydropathy profile and amino acid charge distribution (data not shown) of the C-terminal segment of the suspected Mo transport protein reveals several potential membrane-spanning regions. The presence of this conserved sequence in the ORF and the hydropathic character (hydropathy index, 0.754; the range in which hydrophobic inner mem-

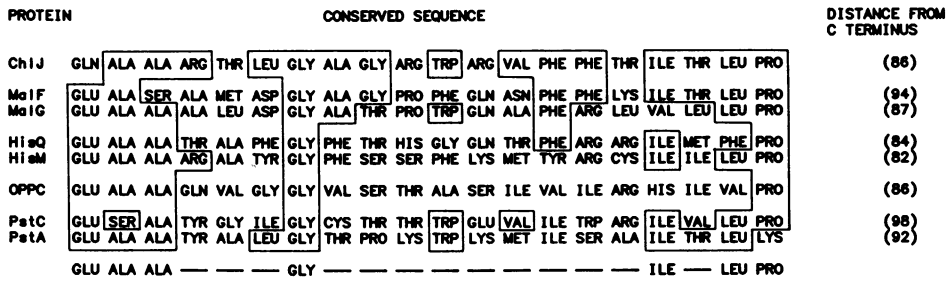

FIG. 4. Amino acid sequence alignment of the consensus sequence ~90 residues from the C terminus of integral inner membrane proteins. The amino acids that are identical to the ChlD protein sequence are boxed, and the conserved residues are indicated below. The alignment of the conserved region in integral inner membrane proteins involved in binding protein-dependent transport systems was taken from Dassa and Hofnung (5).

brane proteins are found [4]) confirms the reading frame predicted by sequence analysis and suggests that it may encode a hydrophobic inner membrane protein that is involved in transport.

The ability to synthesize the molybdenum cofactor which is common to all molybdoenzymes in *E. coli* requires the expression of at least five *chl* gene products. The phenotype of *chlD* and *chlG* is suppressed by the addition of relatively high concentrations of molybdate to the growth medium, which suggests that these gene products may be involved in Mo acquisition. Glaser and DeMoss (9) found, however, that the level of Mo accumulation is the same in *chlD* mutants and wild-type cells. To understand the role that the *chlD* gene plays in Mo metabolism, we determined the complete nucleotide sequence of a 1,609-bp *Sau*3A1 restriction fragment that complemented an *E. coli chlD*::Mu cts mutation. Results of DNA and deduced amino acid sequence analysis of the complementing fragment suggested a role for the *chlD* locus in Mo metabolism. DNA sequence analysis of the region complementing the *chlD* mutant showed only one ORF coding for a protein of the appropriate size. The initiation codon of the putative *chlD* gene was preceded by a potential Shine-Dalgarno sequence (CGGGG) at the appropriate position and was followed by an ORF encoding for 300 amino acids. The deduced amino acid sequence of *chlD* showed regions of homology (i.e., nucleotide-binding fold) with hydrophilic inner membrane proteins in binding protein-dependent transport systems.

In bacteria the known binding protein-dependent transport systems are made up of several proteins (1). The active transport genes are usually contiguous, constituting an operon under the control of a single promoter. The substrate-binding protein located in the periplasmic space is responsible for trapping the substrate in the periplasmic space for its eventual transfer to a group of proteins on the inner membrane. The energy-dependent transport complex that is located on the inner membrane is comprised of two hydrophobic and one hydrophilic membrane proteins. The hydrophilic inner membrane protein (containing the nucleotide-binding site) may be responsible for the energy-coupling mechanism in the active transport process. The DNA sequence upstream from the *chlD* gene contains an ORF, with its translational termination signal being two bases before the initiation site of the putative *chlD* gene. This short intercistronic region is typical for the operon structure of the membrane components in these active transport systems. The hydropathy profile and the charged amino acid distribution of the deduced amino acid sequence of the C-terminal region of the potential gene product reveals several hydrophobic segments (Fig. 2). A region of amino acid sequence homology with other hydrophobic inner membrane transport proteins was also found in the appropriate position from the C terminus. We propose that the ORF upstream from *chlD* is another gene, *chlJ*, in the Mo transport operon coding for a hydrophobic inner membrane protein, which is a component of the Mo transport complex. Five bases downstream from the translational termination signal of *chlD* is another potential ORF that also may be a part of the Mo transport operon. The DNA sequence that is available (~100 bases) is too short to determine the characteristics of the potential gene or its product.

The molecular weight of the *chlD* protein was calculated to be 33,671. From the amino acid composition, it can be predicted that it is a basic protein. Analysis of the deduced amino acid sequence shows that the protein has an even distribution of charged residues and no long hydrophobic segments, which indicates that it is not an integral membrane protein.

The hypothesis that the *chlD* gene product is involved in active MoO_4^{2-} transport suggests that the characteristic phenotype of the *chlD* mutant is a reflection of Mo deprivation. Glaser and DeMoss (9) demonstrated that the *chlD* phenotype can be suppressed by adding molybdate (MoO_4^{2-}) to the growth medium. Furthermore, these investigators showed that Mo accumulation (at relatively low Mo concentrations) is the same in the *chlD* mutant and the wild-type cell. In light of the data presented here, it is possible that the *chlD* mutation causes a defect in Mo transport; the Mo accumulated in the *chlD* mutant is in the periplasmic space bound to the periplasmic binding protein. A number of observations support this hypothesis. *Klebsiella pneumoniae* and *E. coli* appear to have a high-affinity system for Mo transport (17, 27), *K. pneumoniae* and *E. coli* require *chlD* to accumulate Mo (17, 27), the *chlD* gene is required for maximum expression of nitrogenase in *K. pneumoniae* (25) and for maximum expression of molybdoenzymes in *E. coli* (22), the Mo that is accumulated by an *E. coli chlD* mutant is localized in the membrane fragments (6), and *chlD* mutants accumulate Mo at low levels in a highly exchangeable pool; when molybdate is added molybdoenzyme activity is obtained and Mo is in a nonexchangeable pool (25). Shah et al. (25) found that mutants that are deficient in molybdoenzymes (i.e., nitrogenase and nitrate reductase) containing two distinct molybdenum cofactors were *chlD*, which supports the idea that the *chlD* locus is responsible for a common early step (i.e., transport) in the synthesis of both cofactors.

The role that Mo plays in regulating molybdoenzyme expression has not been settled. It might be interesting to reexamine the effect that the *chlD* mutation has on the

expression of molybdoenzymes with regard to Mo regulation. A mutation in the *chlD* locus would result in a Mo-deficient condition in the mutant cell (at low Mo concentrations), and the phenotype of the mutant should be similar to that of wild-type cells that are starved for Mo. The presence of the metal is not obligatory for the expression of the apoenzyme, but a supplement of molybdate (1 mM) to *chlD* mutants and a supplement of molybdate (10 μM) to Mo-starved wild-type cells is required for full expression (27). Pascal and co-workers (21, 22) used *nar-lac* gene fusions to demonstrate that the transcription of the nitrate reductase gene is stimulated by Mo only in a *chlD* mutant background. The other *chl* mutations had no effect on the level of apoenzyme expression. The results suggest that the cell can repress the synthesis of nitrate reductase in response to Mo deprivation; what mediates the response to metal availability is unknown.

ACKNOWLEDGMENTS

We thank V. Stewart for providing strains, and E. Stiefel and G. F. Ames for helpful discussions and for reviewing the manuscript. Special thanks go to B. Eisner for technical and artistic assistance, as well as to C. Keller for assistance in computer analysis.

LITERATURE CITED

1. **Ames, G. F.-L.** 1986. Bacterial periplasmic transport systems: structure, mechanism, and evolution. Annu. Rev. Biochem. **55**:397–425.
2. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
3. **Chou, P., and G. D. Fasman.** 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. **47**:145–147.
4. **Christie, G. E., and T. Platt.** 1980. Gene structure in the tryptophan operon of *Escherichia coli*. Nucleotide sequence of the *trp*C and the flanking intercistronic regions. J. Mol. Biol. **142**:519–530.
5. **Dassa, E., and M. Hofnung.** 1985. Sequence of gene *mal*G in *E. coli* K12: homologies between intergral membrane components from binding protein-dependent transport systems. EMBO J. **4**:2287–2293.
6. **Dubourdieu, M., E. Andrade, and J. Puig.** 1976. Molybdenum and chlorate resistant mutants in *Escherichia coli* K-12. Biochem. Biophys. Res. Commun. **70**:766–773.
7. **Froshauer, S., and J. Beckwith.** 1984. The nucleotide sequence of the gene for *mal*F protein, an inner membrane component of the maltose transport system of *Escherichia coli*. J. Biol. Chem. **259**:10896–10903.
8. **Gilson, E., H. Nikaido, and M. Hofnung.** 1982. Sequence of the *mal*K gene in *E. coli* K12. Nucleic Acids Res. **10**:7449–7458.
9. **Glaser, J. H., and J. A. DeMoss.** 1971. Phenotypic restoration by molybdate of nitrate reductase activity in *chl*D mutants of *Escherichia coli*. J. Bacteriol. **108**:854–860.
10. **Goad, W. B., and M. I. Kanehisa.** 1982. Pattern recognition in nucleic acid sequences. I. A general method for finding local homologies and symmetries. Nucleic Acids Res. **10**:247–263.
11. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557–580.
12. **Higgins, C. F., P. D. Haag, K. Nikaido, F. Ardeshir, G. Garcia, and G. F.-L. Ames.** 1982. Complete nucleotide sequence and identification of membrane components of the histidine transport operon of *S. typhimurium*. Nature (London) **298**:723–727.
13. **Higgins, C. F., I. D. Hiles, K. Whalley, and D. J. Jamieson.** 1985. Nucleotide binding by membrane components of bacterial periplasmic binding protein-dependent transport systems. EMBO J. **4**:1033–1040.
14. **Hinton, S. M., and G. Freyer.** 1986. Cloning, expression and sequencing the molybdenum-pterin binding protein (*mop*) gene of *Clostridium pastuerianum* in *Escherichia coli*. Nucleic Acids Res. **14**:9371–9380.
15. **Hobson, A., R. Weatherwax, and G. F.-L. Ames.** 1984. ATP-binding sites in the membrane components of histidine permease, a periplasmic transport system. Proc. Natl. Acad. Sci. USA **81**:7333–7337.
16. **Ikemura, T.** 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein gene: a proposal for a synonymous codon choice that is optimal for the *E. coli* translation system. J. Mol. Biol. **151**:389–409.
17. **Imperial, J., R. A. Ugalde, V. K. Shah, and W. J. Brill.** 1984. *mol*-mutants of *Klebsiella pneumoniae* requiring high levels of molybdate for nitrogenase activity. J. Bacteriol. **163**:1285–1287.
18. **Johnson, J. L., and K. V. Rajagopalan.** 1982. Structural and metabolic relationship between the molybdenum cofactor and urothione. Proc. Natl. Acad. Sci. USA **79**:6856–6860.
19. **Keller, C., M. Corcoran, and R. J. Roberts.** 1984. Computer programs for handling nucleic acid sequences. Nucleic Acids Res. **12**:379–386.
20. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
21. **Pascal, M.-C., J.-F. Burini, J. Ratouchniak, and M. Chippaux.** 1982. Regulation of the nitrate reductase operon: effects of mutations in *chl*A, B, and E genes. Mol. Gen. Genet. **188**:103–106.
22. **Pascal, M.-C., and M. Chippaux.** 1982. Involvement of a gene of the *chl*E locus in the regulation of the nitrate reductase operon. Mol. Gen. Genet. **185**:334–338.
23. **Ruiz-Herrera, J., and J. A. DeMoss.** 1969. Nitrate reductase complex of *Escherichia coli* K12: participation of specific formate dehydrogenase and cytochrome B_1 components in nitrate reduction. J. Bacteriol. **99**:720–729.
24. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
25. **Shah, V. K., R. A. Uglade, J. Imperial, and W. J. Brill.** 1984. Molybdenum in nitrogenase. Annu. Rev. Biochem. **53**:231–257.
26. **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.
27. **Sperl, G. T., and J. A. DeMoss.** 1975. *chl*D gene function in molybdate activation of nitrate reductase. J. Bacteriol. **122**:1230–1238.
28. **Stewart, V., and C. H. MacGregor.** 1982. Nitrate reductase in *Escherichia coli* K-12: involvement of *chl*C, *chl*E, and *chl*G loci. J. Bacteriol. **151**:788–799.
29. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay.** 1982. Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP requiring enzymes and a common nucleotide binding fold. EMBO J. **1**:945–951.