INFECTION AND IMMUNITY, Dec. 1987, p. 3023–3029
0019-9567/87/123023-07$02.00/0
Copyright © 1987, American Society for Microbiology
Vol. 55, No. 12

Cloning and Sequence Analysis of Cytadhesin P1 Gene from *Mycoplasma pneumoniae*

C. J. SU, V. V. TRYON, AND J. B. BASEMAN*

Department of Microbiology, The University of Texas Health Science Center at San Antonio, San Antonio, Texas 78284

Received 31 July 1987/Accepted 12 September 1987

***Mycoplasma pneumoniae* cytadhesin P1 was purified by monoclonal antibody affinity chromatography followed by preparative sodium dodecyl sulfate-polyacrylamide gel electrophoresis. The N-terminal 18-amino-acid sequence of P1 was determined and used to design two synthetic oligonucleotides, a 14-mer corresponding to amino acids 1 to 5 and an 18-mer corresponding to amino acids 7 to 12. These oligonucleotides served as hybridization probes for the identification of the P1 gene by Southern blot analysis of *M. pneumoniae* DNA. The P1 gene was cloned into plasmid pUC19 and mapped by using appropriate restriction endonucleases. The DNA sequence of the entire P1 gene was determined by subcloning appropriate DNA fragments into bacteriophage M13 and sequencing the DNA by the dideoxy-chain-termination method. The P1 gene contains an open reading frame of 4,881 nucleotides coding for a protein of 1,627 amino acids with a calculated molecular weight of 176,288. Properties of the amino-terminal sequence suggest that protein P1 may be synthesized as a precursor with subsequent processing to a mature protein of a calculated molecular weight of 169,758. Potential antigenic sites were determined by hydrophilicity plots. A computer search revealed that part of the predicted P1 sequence is homologous to cytoskeletal keratin of mammalian species and human fibrinogen alpha chain precursor. These results demonstrate the uniqueness of P1 as a cytadhesin and virulence determinant.**

Mycoplasma-mediated human disease represents a major class of infections which are poorly understood. *Mycoplasma pneumoniae*, the etiological agent of primary atypical pneumonia, possesses a tiplike organelle which permits a highly oriented extracellular parasitism of the respiratory epithelium (5, 11, 26). A surface-localized, trypsin-sensitive protein designated P1 (165 kilodaltons [kDa]) clusters at the mycoplasma terminus and is essential for the cytadherence event (1, 2, 6, 10). Mutants of *M. pneumoniae* which lack P1 or are unable to mobilize and anchor P1 at the tip are avirulent (1, 15, 16). Also, trypsin treatment of wild-type virulent mycoplasmas abrogates the functionality of P1 as an adhesin (1, 2, 10, 11, 23). Thus, it is well established that P1 is a critical virulence determinant but efforts to define structural-functional domains of this hydrophobic, integral membrane protein have failed.

Our previous attempts to clone the P1 gene by introducing recombinant molecules containing *M. pneumoniae* DNA into *E. coli* and screening immunologically for the expression of protein P1 were of limited success (31). This may be due to the premature termination of *M. pneumoniae* proteins in *Escherichia coli* since Yamao et al. (35) recently reported that in *Mycoplasma capricolum* TGA encodes tryptophan rather than a termination signal.

In this report we describe the cloning of the P1 gene using an approach which does not require expression of the P1 protein in *E. coli*. We also present the DNA sequence of the P1 gene and the deduced sequence of the P1 protein.

MATERIALS AND METHODS

Organisms and growth conditions. Virulent hemadsorbing *M. pneumoniae* M129 in the 16th broth passage was grown at 37°C in 32-oz. (ca. 946-ml) glass prescription bottles containing 70 ml of Edward medium as previously described (23). Cells were harvested 72 h after inoculation and stored at −70°C.

E. coli HB101, DH5α, and JM107 were purchased from commercial sources and grown in LB broth (20).

Purification of P1 protein by affinity chromatography. The P1 protein was purified by antibody affinity chromatography according to the method described previously by us (18). Four anti-P1 monoclonal antibodies secreted by independently derived hybridomas (23, 24) were combined and purified by protein A-Sephadex column chromatography. Anti-P1 affinity columns were prepared by coupling 50 mg of purified anti-P1 antibody to 15 ml of cyanogen bromide-activated Sephadex gel (Pharmacia Fine Chemicals, Piscataway, N.J.). Pellets from 100 bottles of *M. pneumoniae* were suspended in 50 ml of 20 mM Tris hydrochloride (pH 8.0), 0.2% sodium deoxycholate (Fisher Scientific Co., Pittsburgh, Pa.), 0.1% sodium dodecyl sulfate (SDS; BDH Chemicals, Poole, England), 10 mM EDTA, and 0.2% Triton X-100 containing 1 mM phenylmethylsulfonyl fluoride. Solubilization of proteins was assisted by passing the cell suspension through successively smaller-gauge needles (22 to 27 gauge). Insoluble material was removed by centrifugation at 100,000 × *g* for 30 min.

Solubilized proteins were applied to the affinity column at 4°C and washed with 5 column volumes of the same buffer minus sodium deoxycholate. Bound protein was eluted with 0.1 M acetic acid (pH 3) containing 0.15 M NaCl and 0.1% SDS. The eluted protein was immediately neutralized with 1.0 M Tris and concentrated in a pressure ultrafiltration concentrator (Amicon Corp., Danvers, Mass.).

The affinity column-derived P1 protein was further purified by preparative polyacrylamide gel electrophoresis (PAGE). After staining, the P1 protein band was cut out of the gel, electroeluted according to Hunkapiller et al. (12), and reprecipitated several times with 80% methanol at −20°C.

During the purification process, the amount of P1 protein was estimated by comparisons with known amounts of

* Corresponding author.

molecular weight standards (Bio-Rad Laboratories, Richmond, Calif.) after Coomassie blue staining.

N-terminal amino acid sequence determination of P1 protein. The purified P1 protein was sequenced from the amino terminus with a gas phase protein sequencer by C. Y. Yang in the Department of Medicine, Baylor College of Medicine, Houston, Tex. Approximately 50 μg of purified P1 was used (300 pmol) for each of three sequence analyses.

Oligonucleotide probes. Two oligonucleotide probes were designed on the basis of the N-terminal amino acid sequence of the P1 protein. These oligonucleotides were synthesized in the Department of Biochemistry, Baylor College of Medicine, and purified by electrophoresis in a 20% polyacrylamide gel containing 8 M urea (3). For use as hybridization probes, the oligonucleotides were labeled at the 5′ end with [γ-^{32}P]ATP by the T4-polynucleotide kinase reaction (20).

Southern blot analysis of *M. pneumoniae* DNA. *M. pneumoniae* DNA was prepared from exponentially growing cells as previously described (31). DNA (12 μg) was digested to completion with appropriate restriction enzymes before electrophoretic separation on 0.7% agarose gels. Gels were stained with ethidium bromide and photographed under UV illumination. DNA was transferred to nitrocellulose filter paper (30) with 20× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), rinsed once with 6× SSC, and then baked at 80°C for 2 h under vacuum. Filters were prehybridized overnight at 37°C in 20 ml of prehybridization solution containing 6× SSC, 60 mM sodium phosphate (pH 7.0), 5× Denhardt solution (bovine serum albumin, polyvinylpyrrolidone, Ficoll [Pharmacia] at 1 mg/ml), and 0.1 mg of denatured herring sperm DNA per ml. Hybridizations were done for 12 h in 10 ml of prehybridization solution plus 10% dextran sulfate and ^{32}P-labeled oligonucleotide probes (3×10^8 cpm) at 25°C (14 base pairs [bp], 14-mer) or 37°C (18 bp, 18-mer). After incubation, filters were rinsed twice with 6× SSC at 4°C (30 min each) and then washed twice in wash solution (3 M tetramethylammonium chloride, 50 mM Tris hydrochloride [pH 8.0], 2 mM EDTA, 0.1% SDS) at the appropriate temperature (14-mer at 37°C and 18-mer at 45°C) for 20 min (34). After being washed, filters were rinsed in 6× SSC at 4°C, dried, and exposed to X-ray film using an intensifying screen.

Cloning of *M. pneumoniae* P1 gene. *M. pneumoniae* DNA (500 μg) was digested to completion with selected restriction enzymes and electrophoresed in 0.7% agarose gels. The DNA of the size range which hybridized to both the 14- and 18-bp oligonucleotide probes was purified by electrophoresis onto DE-81 paper, eluted from the paper with 20 mM Tris hydrochloride, pH 8.0, and 1.5 M NaCl, precipitated with ethanol, and redissolved in TE buffer (20).

Plasmids pUC9 and pUC19 were used as cloning vectors. Vectors were digested with appropriate restriction enzymes, and the 5′-end phosphate was removed by calf intestinal alkaline phosphatase (20). Mycoplasma DNA and vector were mixed at 1:1 molar ratios and ligated at room temperature for 4 h with T4 DNA ligase (20). After incubation the reaction was stopped by adding EDTA to 10 mM and diluted fivefold with distilled H_2O, and the DNA was used to transform competent HB101 or DH5α *E. coli* cells according to the instructions of the manufacturer (Bethesda Research Laboratories, Inc., Gaithersburg, Md.). Transformants were selected on LB agar plates containing 50 μg of ampicillin per ml. Individual colonies were picked and grown overnight in 5 ml of LB broth containing 50 μg of ampicillin per ml. Plasmid DNA was isolated from overnight cultures by the alkaline lysis method (13) and analyzed on agarose gels. To determine which insert-containing plasmids carried the P1 gene, plasmid DNAs were blotted onto nitrocellulose filters. The filters were then hybridized to the ^{32}P-labeled 18-mer oligonucleotide probes, washed, and exposed to film as described above.

DNA sequencing. DNA sequences were determined by the dideoxy-chain-termination method of Sanger et al. (29). M13 sequencing kits were purchased from Bethesda Research Laboratories, and the reactions were performed according to the instructions of the manufacturer, except that deoxy-7-deaza GTP (Boehringer Mannheim Biochemicals, Indianapolis, Ind.) was used in sequencing reactions in place of dGTP (22). Some DNA fragments were sequenced by subcloning appropriate restriction enzyme fragments into an M13 bacteriophage vector (21), and the single-stranded DNA was purified for use as a sequencing template. To sequence the rest of the P1 gene, a large piece of DNA was subcloned into an M13 vector and a series of deletions from the 3′ end were generated by treating the double-stranded DNA with exonuclease III according to the method of Henikoff (8). Subclones with progressive deletions were selected for use as sequencing templates. Both strands of the entire P1 gene were sequenced. Nucleic acid and protein computer analyses were performed by using the Microgenie program (Beckman Instruments, Inc., Palo Alto, Calif.). Comparisons of the P1 DNA and deduced protein sequences were with the most recent releases of the National Institutes of Health GenBank DNA sequence database and the National Biomedical Research Foundation protein sequence database, respectively.

RESULTS

Purification of P1. *M. pneumoniae* cytadhesin protein P1 (165 kDa) was isolated by using monoclonal antibody affinity chromatography as an initial step. Immunoaffinity chromatography selectively enriched for the expected 165-kDa species (Fig. 1A to C). Approximately 400 μg of P1 protein was recovered after the immunoaffinity step from an initial *M. pneumoniae* extract containing 300 mg of total protein. The affinity column-purified P1 was further processed by preparative gel electrophoresis through a 5% polyacrylamide-SDS gel, and the 165-kDa protein was recovered by electroelution. About 60% recovery was achieved after 24 h of elution at room temperature in 50 mM ammonium carbonate containing 0.1% SDS. The eluted protein was then precipitated in 80% methanol to remove SDS. SDS-PAGE analysis of the recovered P1 revealed that the sample contained intact P1 protein (Fig. 1D), and the gel was deliberately overloaded to show the purity of the sample. Finally, the purified protein was shown to be P1 since it reacted with anti-P1 monoclonal antibodies in Western blot analyses (data not shown).

Generation of oligonucleotide probes. The amino-terminal amino acid sequence of P1 protein was determined by gas phase microsequencing. Three separate determinations yielded the sequence shown in Fig. 2. Two oligonucleotide probes complementary to all the possible mRNA combinations encoding different portions of the protein were synthesized, a 14-mer corresponding to amino acids 1 to 5 and an 18-mer corresponding to amino acids 7 to 12 (Fig. 2). Because *M. pneumoniae* may use TGA instead of TGG to encode tryptophan, both C and T were used in the third position of the tryptophan codon of the 18-bp oligonucleotide.

Cloning of the P1 gene. The synthetic oligonucleotides were used to probe the *M. pneumoniae* genome for the P1

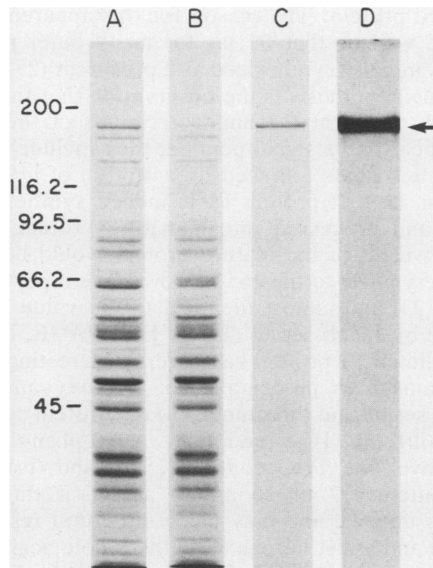


FIG. 1. SDS-PAGE analysis of protein samples during the purification of P1. (A) Total protein extract from *M. pneumoniae*. Arrow at right indicates the position of P1. (B) Same sample as in panel A after a single passage through the anti-P1 affinity column. (C) Protein eluted from the anti-P1 affinity column. (D) P1 after preparative gel electrophoresis and electroelution. Proteins were separated by 7.5% PAGE and stained with Coomassie blue.

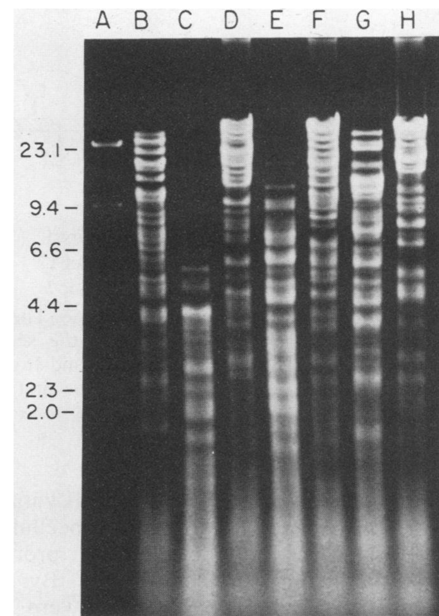


FIG. 3. *M. pneumoniae* DNA (12 μg per lane) digested with different restriction enzymes and separated by 0.7% agarose gel electrophoresis. Lanes: A, standard; B, *Eco*RI; C, *Hae*III; D, *Pst*I; E, *Hin*dIII; F, *Bam*HI; G, *Kpn*I; H, *Sal*I.

gene. *M. pneumoniae* DNA was digested overnight with several restriction endonucleases and separated by 0.7% agarose gel electrophoresis (Fig. 3). These DNAs were transferred to nitrocellulose filters and exposed to the 14- and 18-bp oligonucleotide probes. Both probes hybridized to several DNA bands in each digestion. A 4.3-kilobase (kb) *Hin*dIII fragment hybridized most intensely to both the 14-mer and 18-mer (Fig. 4), strongly implicating this DNA fragment as containing the N-terminal sequence of P1.

To clone this DNA fragment, *M. pneumoniae* DNA was digested with *Hin*dIII, separated by agarose gel electrophoresis, and stained briefly with ethidium bromide. DNA in the 4.3-kb size range was eluted from the gel. The DNA was then ligated into the *Hin*dIII site of pUC9 and transformed into *E. coli* HB101. About 5,000 transformants were obtained, of which 200 transformants were picked, and their plasmid DNA was isolated and analyzed on agarose gels to

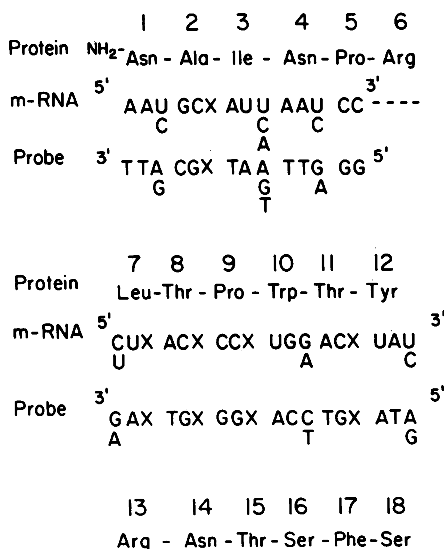


FIG. 2. The N-terminal 18-amino-acid sequence of protein P1 and the 14-mer and 18-mer oligonucleotide probes designed to hybridize to the P1 gene. The 14-mer covers amino acids 1 to 5, and the 18-mer covers amino acids 7 to 12. X = ACGT.

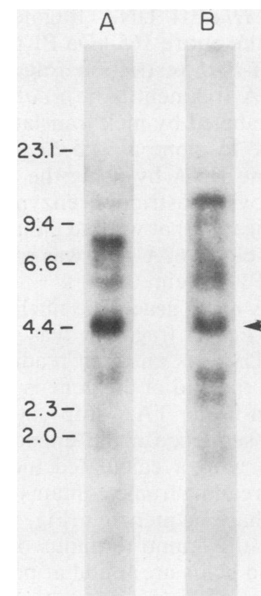
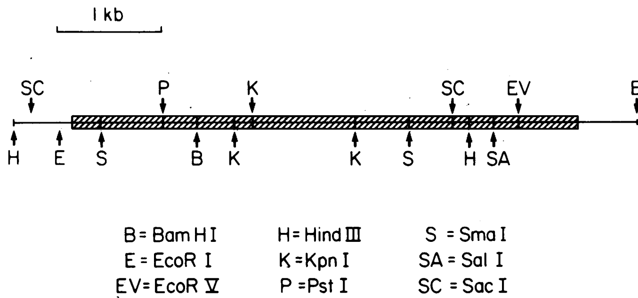


FIG. 4. Southern blot analysis of *M. pneumoniae* genome. *M. pneumoniae* DNA was digested with *Hin*dIII, separated by 0.7% agarose electrophoresis, and transferred to nitrocellulose paper according to the method of Southern (30). The nitrocellulose strip was then hybridized to the 14-mer (A) and 18-mer (B) probes labeled with ^{32}P. A single band (4.3 kb) hybridizes to both probes (arrow).



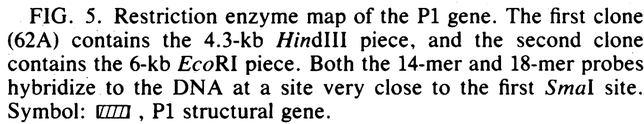


FIG. 5. Restriction enzyme map of the P1 gene. The first clone (62A) contains the 4.3-kb *Hin*dIII piece, and the second clone contains the 6-kb *Eco*RI piece. Both the 14-mer and 18-mer probes hybridize to the DNA at a site very close to the first *Sma*I site. Symbol: ▨, P1 structural gene.

check for insert size. DNAs from 40 plasmids with inserts in the 4- to 5-kb range were blotted onto nitrocellulose paper and hybridized with the 14-mer and 18-mer probes. Three clones hybridized strongly to both probes. By restriction endonuclease analysis, the three clones were shown to contain the same insert, designated 62A (Fig. 5).

The DNA sequence which hybridized to both probes was narrowed to a 350-bp *Hae*III restriction fragment by digesting the 62A plasmid with *Hae*III, separating the DNA on a 5% polyacrylamide gel, and transferring the DNA from the gel onto nitrocellulose paper for hybridization with each individual probe (data not shown). The 350-bp *Hae*III piece was subcloned into M13mp18, and its sequence was determined. It contains both the 14-mer and 18-mer sequences, and most importantly the DNA has an open reading frame which codes for the 18 amino acids found by sequencing the amino terminus of the P1 protein (Fig. 6). Thus, clone 62A contains the gene for protein P1. However, on the basis of the location of the sequenced *Hae*III fragment in the 62A clone, the 4.3-kb *Hin*dIII DNA fragment was not large enough to encode the entire 165-kDa P1 protein. Therefore, we used an *Eco*RI-*Pst*I restriction fragment from 62A to clone a larger DNA fragment. The *Eco*RI-*Pst*I piece from 62A was purified, labeled by nick translation, and used as a hybridization probe to clone a 6-kb *Eco*RI fragment of *M. pneumoniae* genome DNA by using the agarose gel procedure described above. Restriction enzyme analysis of this DNA fragment indicated that we had cloned a piece of DNA which overlapped clone 62A and was sufficiently large to encode the entire P1 protein.

Sequence analysis of P1 gene. The nucleotide sequence of much of the 6-kb *Eco*RI fragment was determined and is shown in Fig. 6. There is an open reading frame of 4,881 nucleotides, and at the end of the gene is a TAG stop codon followed by two in-frame TAA stop codons 21 and 27 bp downstream. This sequence could encode for a protein of 1,627 amino acids with a calculated molecular weight of 176,288. The open reading frame contains the 18 amino acids identified by gas phase sequencing (Fig. 6, box). However, instead of being at the amino terminus of the open reading frame, the 18 amino acids are found at positions 60 to 77 of the deduced protein. The reason for this apparent discrepancy could well be that P1, as for many outer membrane proteins, is initially synthesized as a precursor (25). Consistent with this hypothesis is the observation that the extra 59 amino acids found at the amino terminus of the deduced protein appear like a signal peptide; they include positively charged amino acids followed by a stretch of hydrophobic amino acids (25). If protein P1 is indeed synthesized as a precursor and processed into a mature protein, then the molecular weight of the mature protein would be 169,758, which is very close to the 165 kDa we have reported earlier (1, 17, 18, 23) and almost identical to the value (168 kDa) determined by Jacobs et al. (14) by SDS-PAGE.

The predicted P1 protein has several interesting features: (i) it contains high percentages of hydroxy amino acids (17.7% are serine and threonine); (ii) the protein contains no cysteine; (iii) the TGA codon of tryptophan is slightly preferred over the TGG codon (21:16); and (iv) the high proline content (13 of 26 amino acids) at the carboxy terminus is unusual and may place structural restraints on the protein and assist in regulating the topological organization of the cytadhesin in the membrane (1, 2, 15, 17). Hydrophilicity plots (9) of the predicted protein sequence suggest several potential antigenic sites (Fig. 7) at positions 240 to 260, 280 to 304, 314 to 333, 450 to 479, 680 to 690, 746 to 767, 898 to 913, 1244 to 1260, and 1476 to 1485. A computer search against existing protein sequence files revealed that the predicted P1 sequence is homologous to coat protein A of bacteriophage Ike (protein P1 amino acid numbers 1308 through 1322 compared with bacteriophage amino acid numbers 240 through 254, 73.3% homology; 257 to 290 versus 231 to 264, 41.2% homology), protein 3A of brome mosaic virus (956 to 979 versus 133 to 159, 52% homology), coat proteins vp2 and vp3 of mouse polyomavirus (733 to 746 versus 24 to 38, 66.7% homology), and coat protein A precursor of bacteriophages fd, M-13, and F1 (1296 to 1330 versus 245 to 280, 51.3% homology). The 1290 to 1350 region of P1 also shares extensive homology with cytoskeletal keratin of mammalian species. In addition, two regions of P1 share extensive homology with human fibrinogen alpha chain precursor (337 to 352 versus 338 to 354, 70.6% homology; 822 to 852 versus 544 to 565, 59.1% homology).

Other relevant features of the sequence include elements upstream of the first ATG codon which are similar to those of typical eubacterial promoters (27). Although the base sequences in these regions are almost identical to those found at −10 and −35 in the consensus *E. coli* α^{70} promoter (27), their separation distance is unusually close (14 bp versus 16 to 18 bp in the *E. coli* promoter). No typical ribosome-binding site is observed between the putative promoter and the proposed initiation codon for translation. Also, an imperfect inverted repeat sequence followed by several thymidines is detected 19 bp downstream from the TAG stop codon. This sequence element is a common feature of RNA terminators (28).

DISCUSSION

Our earlier difficulties in characterizing the P1 cytadhesin gene (31) prompted us to approach its isolation by obtaining

FIG. 6. Complete nucleotide sequence and deduced amino acid sequence of the P1 gene. The presumed starting codon of P1 (ATG) is numbered as 1. In the 5′ flanking region, the possible promoter elements (−10 and −35) are underlined. The 18 amino acids which match those determined by protein sequencing of P1 are boxed (nucleotides 178 to 231). In the 3′ flanking region, a sequence with dyad symmetry, which may be a termination signal, is indicated by the arrows, and the asterisk indicates mismatched sequences in this sequence. The complete P1 gene contains 4,881 nucleotides coding for a protein of a calculated 176,288 daltons, which includes an apparent leader peptide (see the text).

AATTCCGTAGTAACGGAATTAACGTACGCGAAGTAGCTTTAAAGTATGGTGGCGGGGTCATATTCAGGCCAGCGGTGC
AGTTCTTAAAAGCAAGCGCGACATAATTCGTGTAGTTCAAGATTGCCAAAAGCAAATTGCTGTATAATTTTAACAACT

30 60
ATG CAC CAA ACC AAA AAA ACT GCC TTG TCC AAG TCC ACT TGG ATT CTC ATC CTC ACC GCC
Met His Gln Thr Lys Lys Thr Ala Leu Ser Lys Ser Thr Trp Ile Leu Ile Leu Thr Ala

90 120
ACC GCC TCC CTC GCG ACG GGA CTC ACC GTA GTG GGA CAC TTC ACA AGT ACC ACC ACG ACG
Thr Ala Ser Leu Ala Thr Gly Leu Thr Val Val Gly His Phe Thr Ser Thr Thr Thr Thr

150 180
CTC AAG CGC CAG CAA TTT AGC TAC ACC CGC CCT GAC GAG GTC GCG CTG CGC CAC ACC AAT
Leu Lys Arg Gln Gln Phe Ser Tyr Thr Arg Pro Asp Glu Val Ala Leu Arg His Thr Asn

210 240
GCC ATC AAC CCG CGC TTA ACC CCG TGA ACG TAT CGT AAC ACG AGC TTT TCC TCC CTC CCC
Ala Ile Asn Pro Arg Leu Thr Pro Trp Thr Tyr Arg Asn Thr Ser Phe Ser Ser Leu Pro

270 300
CTC ACG GGT GAA AAT CCC GGG GCG TGG GCC TTA GTG CGC GAC AAC AGC GCT AAG GGC ATC
Leu Thr Gly Glu Asn Pro Gly Ala Trp Ala Leu Val Arg Asp Asn Ser Ala Lys Gly Ile

330 360
ACT GCC GGC AGT GGC AGT CAA CAA ACC ACG TAT GAT CCC ACC CGA ACC GAA GCG GCT TTG
Thr Ala Gly Ser Gly Ser Gln Gln Thr Thr Tyr Asp Pro Thr Arg Thr Glu Ala Ala Leu

390 420
ACC GCA TCA ACC ACC TTT GCG TTA CGC CGG TAT GAC CTC GCC GGG CGC GCC TTA TAC GAC
Thr Ala Ser Thr Thr Phe Ala Leu Arg Arg Tyr Asp Leu Ala Gly Arg Ala Leu Tyr Asp

450 480
CTC GAT TTT TCG AAG TTA AAC CCG CAA ACG CCC ACG CGC GAC CAA ACC GGG CAG ATC ACC
Leu Asp Phe Ser Lys Leu Asn Pro Gln Thr Pro Thr Arg Asp Gln Thr Gly Gln Ile Thr

510 540
TTT AAC CCC TTT GGC GGC TTT GGT TTG AGT GGG GCT GCA CCC CAA CAG TGA AAC GAG GTC
Phe Asn Pro Phe Gly Gly Phe Gly Leu Ser Gly Ala Ala Pro Gln Gln Trp Asn Glu Val

570 600
AAA AAC AAG GTC CCC GTC GAG GTG GCG CAA GAC CCC TCC AAT CCC TAC CGG TTT GCC GTT
Lys Asn Lys Val Pro Val Glu Val Ala Gln Asp Pro Ser Asn Pro Tyr Arg Phe Ala Val

630 660
TTA CTC GTG CCG CGC AGC GTG GTG TAC TAT GAG CAG TTG CAA AGG GGG TTG GGC TTA CCA
Leu Leu Val Pro Arg Ser Val Val Tyr Tyr Glu Gln Leu Gln Arg Gly Leu Gly Leu Pro

690 720
CAG CAG CGA ACC GAG AGT GGT CAA AAT ACT TCC ACC ACC GGG GCA ATG TTT GGC TTG AAG
Gln Gln Arg Thr Glu Ser Gly Gln Asn Thr Ser Thr Thr Gly Ala Met Phe Gly Leu Lys

750 780
GTG AAG AAC GCC GAG GCG GAC ACC GCG AAG AGC AAT GAA AAA CTC CAG GGC GCT GAG GCC
Val Lys Asn Ala Glu Ala Asp Thr Ala Lys Ser Asn Glu Lys Leu Gln Gly Ala Glu Ala

810 840
ACT GGT TCT TCA ACC ACA TCT GGA TCT GGC CAA TCC ACC CAA CGT GGG GGT TCG TCA GGG
Thr Gly Ser Ser Thr Thr Ser Gly Ser Gly Gln Ser Thr Gln Arg Gly Gly Ser Ser Gly

870 900
GAC ACC AAA GTC AAG GCT TTA AAA ATA GAG GTG AAA AAG AAA TCG GAC TCG GAG GAC AAT
Asp Thr Lys Val Lys Ala Leu Lys Ile Glu Val Lys Lys Lys Ser Asp Ser Glu Asp Asn

930 960
GGT CAG CTG CAG TTA GAA AAA AAT GAT CTC GCC AAC GCT CCC ATT AAG CGG AGC GAG GAG
Gly Gln Leu Gln Leu Glu Lys Asn Asp Leu Ala Asn Ala Pro Ile Lys Arg Ser Glu Glu

990 1020
TCG GGT CAG TCC GTC CAA CTC AAG GCG GAC GAT TTT GGT ACT GCC CTT TCC AGT TCG GGA
Ser Gly Gln Ser Val Gln Leu Lys Ala Asp Asp Phe Gly Thr Ala Leu Ser Ser Ser Gly

1050 1080
TCA GGC GGC AAC TCC AAT CCC GGT TCC CCC ACC CCC TGA AGG CCG TGG CTT GCG ACT GAG
Ser Gly Gly Asn Ser Asn Pro Gly Ser Pro Thr Pro Trp Arg Pro Trp Leu Ala Thr Glu

1110 1140
CAA ATT CAC AAG GAC CTC CCC AAA TGA TCC GCC TCG ATC CTG ATT CTG TAC GAT GCG CCT
Gln Ile His Lys Asp Leu Pro Lys Trp Ser Ala Ser Ile Leu Ile Leu Tyr Asp Ala Pro

1170 1200
TAT GCG CGC AAC CGT ACC GCC ATT GAC CGC GTT GAT CAC TTG GAT CCC AAG GCC ATG ACC
Tyr Ala Arg Asn Arg Thr Ala Ile Asp Arg Val Asp His Leu Asp Pro Lys Ala Met Thr

1230 1260
GCG AAC TAT CCG CCC AGT TGA AGA ACG CCC AAG TGA AAC CAC CAC GGT TTG TGG GAC TGA
Ala Asn Tyr Pro Pro Ser Trp Arg Thr Pro Lys Trp Asn His His Gly Leu Trp Asp Trp

1290 1320
AAG GCG CGC GAT GTT TTG CTC CAA ACC ACC GGG TTC TTC AAC CCG CGC CGC CAC CCC GAG
Lys Ala Arg Asp Val Leu Leu Gln Thr Thr Gly Phe Phe Asn Pro Arg Arg His Pro Glu

1350 1380
TGG TTT GAT GGC GGG CAG ACG GTC GCG GAT AAC GAA AAG ACC GGG TTT GAT GTG GAT AAC
Trp Phe Asp Gly Gly Gln Thr Val Ala Asp Asn Glu Lys Thr Gly Phe Asp Val Asp Asn

1410 1440
TCT GAA AAC ACC AAG CAG GGC TTT CAA AAG GAA GCT GAC TCC GAC AAG TCG GCC CCG ATC
Ser Glu Asn Thr Lys Gln Gly Phe Gln Lys Glu Ala Asp Ser Asp Lys Ser Ala Pro Ile

1470 1500
GCC CTC CCG TTT GAA GCG TAC TTC GCC AAC ATT GGC AAC CTC ACC TGG TTC GGG CAA GCG
Ala Leu Pro Phe Glu Ala Tyr Phe Ala Asn Ile Gly Asn Leu Thr Trp Phe Gly Gln Ala

1530 1560
CTT TTG GTG TTT GGT GGC AAT GGC CAT GTT ACC AAG TCG GCC CAC ACC GCG CCT TTG AGT
Leu Leu Val Phe Gly Gly Asn Gly His Val Thr Lys Ser Ala His Thr Ala Pro Leu Ser

1590 1620
ATA GGT GTC TTT AGG GTG CGC TAT AAT GCA ACT GGT ACC AGT GCT ACT GTA ACT GGT TGA
Ile Gly Val Phe Arg Val Arg Tyr Asn Ala Thr Gly Thr Ser Ala Thr Val Thr Gly Trp

1650 1680
CCA TAT GCC TTA CTG TTC TCA GGC ATG GTC AAC AAA CAA ACT GAC GGG TTA AAG GAT CTA
Pro Tyr Ala Leu Leu Phe Ser Gly Met Val Asn Lys Gln Thr Asp Gly Leu Lys Asp Leu

1710 1740
CCC TTT AAC AAT AAC CGC TGG TTT GAA TAT GTA CCA CGG ATG GCA GTT GCT GGC GCT AAG
Pro Phe Asn Asn Asn Arg Trp Phe Glu Tyr Val Pro Arg Met Ala Val Ala Gly Ala Lys

1770 1800
TTC GTT GGT AGG GAA CTC GTT TTA GCG GGT ACC ATT ACC ATG GGT GAT ACC GCT ACC GTA
Phe Val Gly Arg Glu Leu Val Leu Ala Gly Thr Ile Thr Met Gly Asp Thr Ala Thr Val

1830 1860
CCT CGC TTA CTG TAC GAT GAA CTT GAA AGC AAC CTG AAC TTA GTA GCG CAA GGC CAA GGT
Pro Arg Leu Leu Tyr Asp Glu Leu Glu Ser Asn Leu Asn Leu Val Ala Gln Gly Gln Gly

1890 1920
CTT TTA CGC GAA GAC TTG CAA CTC TTC ACA CCC TAC GGA TGA GCC AAT CGT CCG GAT TTA
Leu Leu Arg Glu Asp Leu Gln Leu Phe Thr Pro Tyr Gly Trp Ala Asn Arg Pro Asp Leu

1950 1980
CCA ATC GGG GCT TGA AGT AGT AGT AGT AGT AGT AGT CAC AAC GCA CCC TAC TAC TTC CAC
Pro Ile Gly Ala Trp Ser Ser Ser Ser Ser Ser Ser His Asn Ala Pro Tyr Tyr Phe His

2010 2040
AAT AAC CCC GAT TGA CAA GAC CGT CCA ATC CAA AAT GTG GTT GAT GCC TTT ATT AAG CCC
Asn Asn Pro Asp Trp Gln Asp Arg Pro Ile Gln Asn Val Val Asp Ala Phe Ile Lys Pro

2070 2100
TGA GAG GAC AAG AAC GGT AAG GAT GAT GCC AAA TAC ATC TAC CCT TAC CGT TAC AGT GGC
Trp Glu Asp Lys Asn Gly Lys Asp Asp Ala Lys Tyr Ile Tyr Pro Tyr Arg Tyr Ser Gly

2130 2160
ATG TGA GCT TGA CAG GTA TAC AAC TGG TCC AAT AAG CTC ACT GAC CAA CCA TTA AGT GCT
Met Trp Ala Trp Gln Val Tyr Asn Trp Ser Asn Lys Leu Thr Asp Gln Pro Leu Ser Ala

2190 2220
GAC TTT GTC AAT GAG AAT GCT TAC CAA CCA AAC TCC TTG TTT GCT GCT ATT CTC AAT CCG
Asp Phe Val Asn Glu Asn Ala Tyr Gln Pro Asn Ser Leu Phe Ala Ala Ile Leu Asn Pro

2250 2280
GAA TTG TTA GCA GCT CTT CCC GAC AAG GTT AAA TAC GGT AAG GAA AAC GAG TTT GCT GCT
Glu Leu Leu Ala Ala Leu Pro Asp Lys Val Lys Tyr Gly Lys Glu Asn Glu Phe Ala Ala

2310 2340
AAC GAG TAC GAG CGC TTT AAC CAG AAG TTA ACG GTA GCT CCT ACC CAA GGA ACA AAC TGA
Asn Glu Tyr Glu Arg Phe Asn Gln Lys Leu Thr Val Ala Pro Thr Gln Gly Thr Asn Trp

2370 2400
TCC CAC TTC TCC CCC ACG CTT TCC CGT TTC TCC ACC GGG TTC AAC CTT GTG GGG TCG GTG
Ser His Phe Ser Pro Thr Leu Ser Arg Phe Ser Thr Gly Phe Asn Leu Val Gly Ser Val

2430 2460
CTC GAC CAG GTG TTG GAT TAT GTG CCC TGG ATT GGG AAT GGG TAC AGG TAT GGC AAT AAC
Leu Asp Gln Val Leu Asp Tyr Val Pro Trp Ile Gly Asn Gly Tyr Arg Tyr Gly Asn Asn

2490 2520
CAC CGG GGC GTG GAT GAT ATA ACC GCG CCT CAA ACC AGC GCG GGG TCG TCC AGC GGA ATT
His Arg Gly Val Asp Asp Ile Thr Ala Pro Gln Thr Ser Ala Gly Ser Ser Ser Gly Ile

2550 2580
AGT ACG AAC ACA AGT GGT TCG CGT TCC TTT CTC CCG ACG TTT TCC AAC ATC GGC GTC GGC
Ser Thr Asn Thr Ser Gly Ser Arg Ser Phe Leu Pro Thr Phe Ser Asn Ile Gly Val Gly

2610 2640
CTC AAA GCG AAT GTC CAA GCC ACC CTC GGG GGC AGT CAG ACG ATG ATT ACA GGC GGT TCG
Leu Lys Ala Asn Val Gln Ala Thr Leu Gly Gly Ser Gln Thr Met Ile Thr Gly Gly Ser

2670 2700
CCT CGA AGA ACC CTC GAC CAA GCC AAC CTC CAG CTC TGA ACG GGG GCG GGG TGA AGG AAT
Pro Arg Arg Thr Leu Asp Gln Ala Asn Leu Gln Leu Trp Thr Gly Ala Gly Trp Arg Asn

2730 2760
GAT AAG GCT TCA AGT GGA CAA AGT GAC GAA AAC CAC ACC AAG TTC ACG AGC GCT ACG GGG
Asp Lys Ala Ser Ser Gly Gln Ser Asp Glu Asn His Thr Lys Phe Thr Ser Ala Thr Gly

2790 2820
ATG GAC CAG CAG GGA CAA TCA GGT ACC TCC GCG GGG AAT CCC GAC TCG TTA AAG CAG GAT
Met Asp Gln Gln Gly Gln Ser Gly Thr Ser Ala Gly Asn Pro Asp Ser Leu Lys Gln Asp

2850 2880
AAT ATT AGT AAG AGT GGG GAT AGT TTA ACC ACG CAG GAC GGC AAT GCG ATC GAT CAA CAA
Asn Ile Ser Lys Ser Gly Asp Ser Leu Thr Thr Gln Asp Gly Asn Ala Ile Asp Gln Gln

2910 2940
GAG GCC ACC AAC TAC ACC AAC CTC CCC CCC AAC CTC ACC CCC ACC GCT GAT TGA CCG AAC
Glu Ala Thr Asn Tyr Thr Asn Leu Pro Pro Asn Leu Thr Pro Thr Ala Asp Trp Pro Asn

2970 3000
GCG CTG TCA TTC ACC AAC AAG AAC AAC GCG CAG CGC GCC CAG CTC TTC CTC CGC GGC TTG
Ala Leu Ser Phe Thr Asn Lys Asn Asn Ala Gln Arg Ala Gln Leu Phe Leu Arg Gly Leu

3030 3060
TTG GGC AGC ATC CCG GTG TTG GTG AAT CGA AGT GGG TCC GAT TCC AAC AAA TTC CAA GCC
Leu Gly Ser Ile Pro Val Leu Val Asn Arg Ser Gly Ser Asp Ser Asn Lys Phe Gln Ala

3090 3120
ACC GAC CAA AAA TGG TCC TAC ACC GAC TTA CAT TCG GAC CAA ACC AAA CTG AAC CTC CCC
Thr Asp Gln Lys Trp Ser Tyr Thr Asp Leu His Ser Asp Gln Thr Lys Leu Asn Leu Pro

3150 3180
GCT TAC GGT GAG GTG AAT GGG TTG TTG AAT CCG GCG TTG GTG GAA ACC TAT TTT GGG AAC
Ala Tyr Gly Glu Val Asn Gly Leu Leu Asn Pro Ala Leu Val Glu Thr Tyr Phe Gly Asn

3210 3240
ACG CGA GCG GGT GGT TCG GGG TCC AAC ACG ACC AGT TCA CCC GGT ATC GGT TTT AAA ATT
Thr Arg Ala Gly Gly Ser Gly Ser Asn Thr Thr Ser Ser Pro Gly Ile Gly Phe Lys Ile

3270 3300
CCC GAA CAA AAT AAT GAT TCC AAA GCC ACC CTG ATC ACC CCC GGG TTG GCT TGA ACG CCC
Pro Glu Gln Asn Asn Asp Ser Lys Ala Thr Leu Ile Thr Pro Gly Leu Ala Trp Thr Pro

3330 3360
CAG GAC GTC GGT AAC CTC GTT GTC AGT GGC ACC ACG GTG AGC TTC CAG CTC GGC GGG TGG
Gln Asp Val Gly Asn Leu Val Val Ser Gly Thr Thr Val Ser Phe Gln Leu Gly Gly Trp

3390 3420
CTG GTC ACC TTC ACG GAC TTT GTC AAA CCC CGC GCG GGT TAC CTC GGT CTC CAG TTA ACG
Leu Val Thr Phe Thr Asp Phe Val Lys Pro Arg Ala Gly Tyr Leu Gly Leu Gln Leu Thr

3450 3480
GGC TTG GAT GCA AGT GAT GCG ACG CAG CGC GCC CTC ATT TGG GCC CCC CGG CCC TGA GCG
Gly Leu Asp Ala Ser Asp Ala Thr Gln Arg Ala Leu Ile Trp Ala Pro Arg Pro Trp Ala

3510 3540
GCC TTT CGT GGC AGT TGG GTC AAC CGG TTG GGC CGC GTG GAG AGT GTG TGG GAT TTG AAG
Ala Phe Arg Gly Ser Trp Val Asn Arg Leu Gly Arg Val Glu Ser Val Trp Asp Leu Lys

3570 3600
GGG GTG TGG GCG GAT CAA GCT CAG TCC GAC TCG CAA GGA TCT ACC ACC ACC GCA ACA AGG
Gly Val Trp Ala Asp Gln Ala Gln Ser Asp Ser Gln Gly Ser Thr Thr Thr Ala Thr Arg

3630 3660
AAC GCC TTA CCG GAG CAC CCG AAT GCT TTG GCC TTT CAG GTG AGT GTG GTG GAA GCG AGT
Asn Ala Leu Pro Glu His Pro Asn Ala Leu Ala Phe Gln Val Ser Val Val Glu Ala Ser

3690 3720
GCT TAC AAG CCA AAC ACG AGC TCC GGC CAA ACC CAA TCC ACT AAC AGT TCC CCC TAC CTG
Ala Tyr Lys Pro Asn Thr Ser Ser Gly Gln Thr Gln Ser Thr Asn Ser Ser Pro Tyr Leu

3750 3780
CAC TTG GTG AAG CCT AAG AAA GTT ACC CAA TCC GAC AAG TTA GAC GAC GAT CTT AAA AAC
His Leu Val Lys Pro Lys Lys Val Thr Gln Ser Asp Lys Leu Asp Asp Asp Leu Lys Asn

3810 3840
CTG TTG GAC CCC AAC CAG GTT CGC ACC AAG CTG CGC CAA AGC TTT GGT ACA GAC CAT TCC
Leu Leu Asp Pro Asn Gln Val Arg Thr Lys Leu Arg Gln Ser Phe Gly Thr Asp His Ser

3870 3900
ACC CAG CCC CAG CCC CAA TCG CTC AAA ACA ACG ACA CCG GTA TTT GGG ACG AGT AGT GGT
Thr Gln Pro Gln Pro Gln Ser Leu Lys Thr Thr Thr Pro Val Phe Gly Thr Ser Ser Gly

3930 3960
AAC CTC AGT AGT GTG CTT AGT GGT GGG GGT GCT GGA GGG GGT TCT TCA GGC TCA GGT CAA
Asn Leu Ser Ser Val Leu Ser Gly Gly Gly Ala Gly Gly Gly Ser Ser Gly Ser Gly Gln

3990 4020
TCT GGC GTG GAT CTC TCC CCC GTT GAA AAA GTG AGT GGG TGG CTT GTG GGG CAG TTA CCA
Ser Gly Val Asp Leu Ser Pro Val Glu Lys Val Ser Gly Trp Leu Val Gly Gln Leu Pro

4050 4080
AGC ACG AGT GAC GGA AAC ACC TCC TCC ACC AAC AAC CTC GCG CCT AAT ACT AAT ACG GGG
Ser Thr Ser Asp Gly Asn Thr Ser Ser Thr Asn Asn Leu Ala Pro Asn Thr Asn Thr Gly

4110 4140
AAT GAT GTG GTG GGG GTT GGT CGA CTT TCT GAA AGC AAC GCC GCA AAG ATG AAT GAC GAT
Asn Asp Val Val Gly Val Gly Arg Leu Ser Glu Ser Asn Ala Ala Lys Met Asn Asp Asp

4170 4200
GTT GAT GGT ATT GTA CGC ACC CCA CTC GCT GAA CTG TTA GAT GGG GAA GGA CAA ACA GCT
Val Asp Gly Ile Val Arg Thr Pro Leu Ala Glu Leu Leu Asp Gly Glu Gly Gln Thr Ala

4230 4260
GAC ACT GGT CCA CAA AGC GTG AAG TTC AAG TCT CCT GAC CAA ATT GAC TTC AAC CGC TTG
Asp Thr Gly Pro Gln Ser Val Lys Phe Lys Ser Pro Asp Gln Ile Asp Phe Asn Arg Leu

4290 4320
TTT ACC CAC CCA GTC ACC GAT CTG TTT GAT CCG GTA ACT ATG TTG GTG TAT GAC CAG TAC
Phe Thr His Pro Val Thr Asp Leu Phe Asp Pro Val Thr Met Leu Val Tyr Asp Gln Tyr

4350 4380
ATA CCG CTG TTT ATT GAT ATC CCA GCA AGT GTG AAC CCT AAA ATG GTT CGT TTA AAG GTC
Ile Pro Leu Phe Ile Asp Ile Pro Ala Ser Val Asn Pro Lys Met Val Arg Leu Lys Val

4410 4440
TTG AGC TTT GAC ACC AAC GAA CAG AGC TTA GGT CTC CGC TTA GAG TTC TTT AAA CCT GAT
Leu Ser Phe Asp Thr Asn Glu Gln Ser Leu Gly Leu Arg Leu Glu Phe Phe Lys Pro Asp

4470 4500
CAA GAT ACC CAA CCA AAC AAC AAC GTT CAG GTC AAT CCG AAT AAC GGT GAC TTC TTA CCA
Gln Asp Thr Gln Pro Asn Asn Asn Val Gln Val Asn Pro Asn Asn Gly Asp Phe Leu Pro

4530 4560
CTG TTA ACG GCC TCC AGT CAA GGT CCC CAA ACC TTG TTT AGT CCG TTT AAC CAG TGA CCT
Leu Leu Thr Ala Ser Ser Gln Gly Pro Gln Thr Leu Phe Ser Pro Phe Asn Gln Trp Pro

4590 4620
GAT TAC GTG TTG CCG TTA GCG ATC ACT GTA CCT ATT GTT GTG ATT GTG CTC AGT GTT ACC
Asp Tyr Val Leu Pro Leu Ala Ile Thr Val Pro Ile Val Val Ile Val Leu Ser Val Thr

4650 4680
TTA GGA CTT GCC ATT GGA ATC CCA ATG CAC AAG AAC AAA CAG GCC TTG AAG GCT GGG TTT
Leu Gly Leu Ala Ile Gly Ile Pro Met His Lys Asn Lys Gln Ala Leu Lys Ala Gly Phe

4710 4740
GCG CTA TCA AAC CAA AAG GTT GAT GTG TTG ACC AAA GCG GTT GGT AGT GTC TTT AAG GAA
Ala Leu Ser Asn Gln Lys Val Asp Val Leu Thr Lys Ala Val Gly Ser Val Phe Lys Glu

4770 4800
ATC ATT AAC CGC ACA GGT ATC AGT CAA GCG CCA AAA CGC TTG AAA CAA ACC AGT GCG GCT
Ile Ile Asn Arg Thr Gly Ile Ser Gln Ala Pro Lys Arg Leu Lys Gln Thr Ser Ala Ala

4830 4860
AAA CCA GGA GCA CCC CGC CCA CCA GTA CCA CCA AAG CCA GGG GCT CCT AAG CCA CCA GTG
Lys Pro Gly Ala Pro Arg Pro Pro Val Pro Pro Lys Pro Gly Ala Pro Lys Pro Pro Val

CAA CCA CCT AAA AAA CCC GCT TAGTATTTATGAAATCGAAGCTAAAGTTAAAACGTTATTTACTGTTTTTAC
Gln Pro Pro Lys Lys Pro Ala End

CACTTTTACCGCTAGGACGTTGTCACTAGCCAACACCTACCTCCTCCAAGACCACAACACCCTCACCCCCTACACGCCC

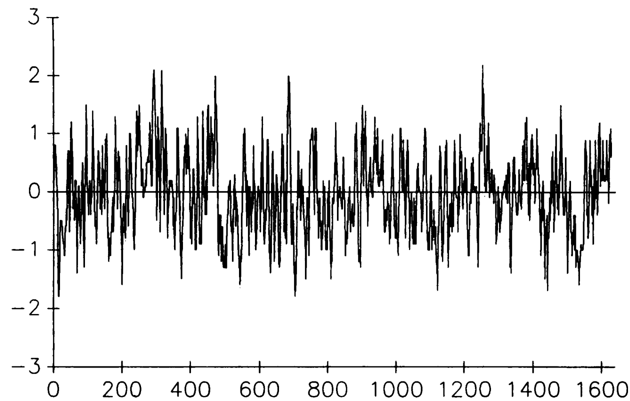


FIG. 7. Plot of hydrophilicity value versus sequence position of P1 according to the method of Hopp and Woods (9). Hydrophilicity values are averaged over six amino acids through the length of P1; highest positive values represent charged hydrophilic regions.

a partial amino acid sequence of protein P1 and then synthesizing two oligonucleotides (14-mer and 18-mer) complementary to the corresponding gene in order to identify appropriate clones. Because of the small size of the *M. pneumoniae* genome (approximately 800 kb), we calculated that sequences which hybridized to the 14-mer or 18-mer probes should occur only one or two times. If the same piece of *M. pneumoniae* DNA hybridized to both 14-mer and 18-mer probes, this DNA fragment should contain the sequence coding for the amino terminus of P1. The 4.3-kb *Hin*dIII piece of *M. pneumoniae* (Fig. 4 to 6) contains the appropriate P1 gene sequence not only because it hybridizes to both 14-mer and 18-mer probes, but also because it codes for all 18 amino acids determined from the N-terminal sequencing of the purified P1 protein. Of interest is the observation that the 10th amino acid that we sequenced from the N terminus is a tryptophan, which further suggests that if *E. coli* uses the P1 promoter sequence, the protein might be terminated prematurely.

From the nucleotide sequences, a possible translational initiation site ATG occurs in frame, which is 177 nucleotides from the P1 N-terminal sequence. There are conventional transcription initiation sites at −35 and −10 upstream with a distance of 14 nucleotides between these two consensus sequences (27), but no ribosome-binding site is observed between −10 and the initiation codon. This predicts a protein with an extension of 59 amino acids from the N terminus. Another possible translational initiation codon is the GTG (7) at position 91. Use of this initiation site would predict a 28-amino-acid precursor. Since all of the predictions are made on the basis of the study of *E. coli* and other procaryotic organisms and no information about mycoplasma promoter or initiation sites is available, further study is needed to determine the precise transcriptional and translational initiation sites.

The predicted P1 gene sequence is consistent with available information about protein P1: (i) the predicted molecular weight of P1 approximates the reported values; (ii) the predicted N-terminal amino acid sequence fits exactly with the gas phase sequence analysis of purified P1 protein; (iii) the predicted P1 sequence contains more basic amino acids (Arg + Lys + His = 169) than acidic (Asp + Glu = 143) (our unpublished isoelectric focusing data show that P1 has an isoelectric point at a basic pH); (iv) the predicted P1 has no intramolecular disulfide bonding, which correlates with our previous observation that the P1 position in polyacrylamide gels is not changed after exposure to sample buffer containing reducing agents.

Since cytadhesin P1 is strongly immunogenic based upon humoral immune response data in humans and hamsters (19) and the appearance of anti-P1 antibodies correlates with resolution of the infection (19), we are interested in identifying the antigenic sites which may serve as effective vaccine candidates. The hydrophilicity plot (Fig. 7) implicates several potential sites. It is fascinating that parts of the P1 sequence are homologous to specific viral coat proteins and mammalian cytoskeletal keratin as well as the human fibrinogen alpha chain precursor, which may relate to previous observations of autoimmune-like mechanisms of pathophysiology associated with mycoplasma disease (4, 33).

Now that the gene sequence of *M. pneumoniae* cytadhesin P1 has been delineated and its protein sequence has been deduced, we can begin to address other relevant issues. Do gene duplications or rearrangements of the P1 gene occur? Why does P1 lose functionality as an adhesin after brief protease treatment? What factors regulate the mobilization and clustering of P1 at the tip organelle, and what is the basis for the cooperative interactions between P1 and other membrane proteins (1, 17)? Will synthetic peptides which substitute for the immunologically and functionally relevant domains of P1 serve as effective vaccine candidates and diagnostic probes? Can the available P1 gene and protein data assist in characterizing the immunologically related 140-kDa protein (23) of the newly discovered human pathogen, *Mycoplasma genitalium* (32)? In addition, identification of the host cell receptor(s) which mediates *M. pneumoniae* cytadherence in the respiratory tract has been elusive, and information provided by the P1 sequence may assist in defining tissue tropism at the molecular level.

ACKNOWLEDGMENTS

We thank Rose Garza for her expert secretarial assistance.

This research was supported in part by Public Health Service grant AI 18540 from The National Institute of Allergy and Infectious Diseases.

LITERATURE CITED

1. **Baseman, J. B., R. M. Cole, D. C. Krause, and D. K. Leith.** 1982. Molecular basis for cytadsorption of *Mycoplasma pneumoniae*. J. Bacteriol. **151:**1514–1522.
2. **Baseman, J. B., D. L. Drouillard, D. K. Leith, and J. Morrison-Plummer.** 1985. Role of *Mycoplasma pneumoniae* adhesin P1 and accessory proteins in cytadsorption, p. 18–23. *In* S. E. Mergenhagen and B. Rosan (ed.), Molecular basis of oral microbial adhesion. American Society for Microbiology, Washington, D.C.
3. **Berent, S. L., M. Mahmoudi, R. M. Torczynski, P. W. Bragg, and A. P. Bollon.** 1985. Comparison of oligonucleotide and long DNA fragments as probes in DNA and RNA dot, Southern, Northern, colony and plaque hybridizations. Biotechnology **3:**208–220.
4. **Biberfeld, G.** 1971. Antibodies to brain and other tissues in cases of *Mycoplasma pneumoniae* infection. Clin. Exp. Immunol. **8:**319–333.
5. **Collier, A. M., and J. B. Baseman.** 1973. Organ culture techniques with mycoplasmas. Ann. N.Y. Acad. Sci. **225:**277–289.
6. **Feldner, J., U. Gobel, and W. Bredt.** 1982. *Mycoplasma pneumoniae* adhesin localized to tip structure by monoclonal antibody. Nature (London) **298:**765–767.
7. **Gold, L., D. Pribnow, T. Schneider, S. Shinedling, B. S. Singer, and G. Stormo.** 1981. Translational initiation in prokaryotes. Annu. Rev. Microbiol. **35:**365–403.
8. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III

creates targeted breakpoints for DNA sequencing. Gene **28**:351–359.

9. **Hopp, T. P., and K. R. Woods.** 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA **78**:3824–3828.
10. **Hu, P. C., R. M. Cole, Y. S. Huang, J. A. Graham, D. E. Gardner, A. M. Collier, and W. A. Clyde.** 1982. *Mycoplasma pneumoniae* infection: role of a surface protein in the attachment organelle. Science **216**:313–315.
11. **Hu, P. C., A. M. Collier, and J. B. Baseman.** 1977. Surface parasitism by *Mycoplasma pneumoniae* of respiratory epithelium. J. Exp. Med. **145**:1328–1343.
12. **Hunkapiller, M. W., E. Lujan, F. Ostrander, and L. E. Hood.** 1983. Isolation of microgram quantities of proteins from polyacrylamide gels for amino acid sequence analysis. Methods Enzymol. **91**:227–236.
13. **Ish-Horowicz, D., and J. F. Burke.** 1981. Rapid and efficient cosmid vector cloning. Nucleic Acids Res. **9**:2989–2998.
14. **Jacobs, E., A. Bennewitz, and W. Bredt.** 1986. Reaction pattern of human anti-*Mycoplasma pneumoniae* antibodies in enzyme-linked immunosorbent assays and immunoblotting. J. Clin. Microbiol. **23**:517–522.
15. **Kahane, I., S. Tucker, and J. B. Baseman.** 1985. Detection of *Mycoplasma pneumoniae* adhesin (P1) in the nonhemadsorbing population of virulent *Mycoplasma pneumoniae*. Infect. Immun. **49**:457–458.
16. **Krause, D. C., D. K. Leith, and J. B. Baseman.** 1983. Reacquisition of specific proteins confers virulence in *Mycoplasma pneumoniae*. Infect. Immun. **39**:830–836.
17. **Krause, D. C., D. K. Leith, R. M. Wilson, and J. B. Baseman.** 1982. Identification of *Mycoplasma pneumoniae* proteins associated with hemadsorption and virulence. Infect. Immun. **35**:809–817.
18. **Leith, D. K., and J. B. Baseman.** 1984. Purification of a *Mycoplasma pneumoniae* adhesin by monoclonal antibody affinity chromatography. J. Bacteriol. **157**:678–680.
19. **Leith, D. K., L. B. Trevino, J. G. Tully, L. B. Senterfit, and J. B. Baseman.** 1983. Host discrimination of *Mycoplasma pneumoniae* proteinaceous immunogens. J. Exp. Med. **157**:502–514.
20. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
21. **Messing, J., R. Crea, and P. H. Seeburg.** 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9**:309–321.
22. **Mizusawa, S., S. Nishimura, and F. Seela.** 1986. Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. Nucleic Acids Res. **14**:1319–1324.
23. **Morrison-Plummer, J., A. Lazzell, and J. B. Baseman.** 1987. Shared epitopes between *Mycoplasma pneumoniae* major adhesin protein P1 and a 140-kilodalton protein of *Mycoplasma genitalium*. Infect. Immun. **55**:49–56.
24. **Morrison-Plummer, J., D. K. Leith, and J. B. Baseman.** 1986. Biological effects of anti-lipid and anti-protein monoclonal antibodies on *Mycoplasma pneumoniae*. Infect. Immun. **53**:398–403.
25. **Oliver, D.** 1985. Protein secretion in *Escherichia coli*. Annu. Rev. Microbiol. **39**:615–648.
26. **Powell, D. A., P. C. Hu, M. Wilson, A. M. Collier, and J. B. Baseman.** 1976. Attachment of *Mycoplasma pneumoniae* to respiratory epithelium. Infect. Immun. **13**:959–966.
27. **Reznikoff, W. S., D. A. Siegele, D. W. Cowing, and C. A. Gross.** 1985. The regulation of transcription initiation in bacteria. Annu. Rev. Genet. **19**:355–387.
28. **Rosenberg, M., and P. Court.** 1979. Regulatory sequence involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
29. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
30. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
31. **Trevino, L. B., W. G. Haldenwang, and J. B. Baseman.** 1986. Expression of *Mycoplasma pneumoniae* antigens in *Escherichia coli*. Infect. Immun. **53**:129–134.
32. **Tully, J. G., D. Taylor-Robinson, R. M. Cole, and D. L. Rose.** 1981. A newly discovered mycoplasma in the human urogenital tract. Lancet **i**:1288–1291.
33. **Wise, K. S., and R. K. Watson.** 1985. Antigenic mimicry of mammalian intermediate filaments by mycoplasmas. Infect. Immun. **48**:587–591.
34. **Wood, W. I., J. Gitschier, L. A. Lasky, and R. M. Lawn.** 1985. Base composition-independent hybridization in tetramethylammonium chloride. A method for oligonucleotide screening of highly complex gene libraries. Proc. Natl. Acad. Sci. USA **82**:1585–1588.
35. **Yamao, F., A. Muto, Y. Kawauchi, M. Iwami, S. Iwagami, Y. Azumi, and S. Osawa.** 1985. UGA is read as tryptophan in *Mycoplasma capricolum*. Proc. Natl. Acad. Sci. USA **82**:2306–2309.