JOURNAL OF BACTERIOLOGY, May 1998, p. 2373–2378 Vol. 180, No. 9
0021-9193/98/$04.00+0

Cloning and Characterization of a Gene Encoding the Major Surface Protein of the Bacterial Endosymbiont *Wolbachia pipientis*

HENK R. BRAIG,[1] WEIGUO ZHOU,[1,2] STEPHEN L. DOBSON,[1] AND SCOTT L. O'NEILL[1]*

Section of Vector Biology, Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, Connecticut 06520,[1] and Institute of Genetics, Fudan University, Shanghai 200433, People's Republic of China[2]

Received 7 November 1997/Accepted 24 February 1998

The maternally inherited intracellular symbiont *Wolbachia pipientis* is well known for inducing a variety of reproductive abnormalities in the diverse arthropod hosts it infects. It has been implicated in causing cytoplasmic incompatibility, parthenogenesis, and the feminization of genetic males in different hosts. The molecular mechanisms by which this fastidious intracellular bacterium causes these reproductive and developmental abnormalities have not yet been determined. In this paper, we report on (i) the purification of one of the most abundantly expressed *Wolbachia* proteins from infected *Drosophila* eggs and (ii) the subsequent cloning and characterization of the gene (*wsp*) that encodes it. The functionality of the *wsp* promoter region was also successfully tested in *Escherichia coli*. Comparison of sequences of this gene from different strains of *Wolbachia* revealed a high level of variability. This sequence variation correlated with the ability of certain *Wolbachia* strains to induce or rescue the cytoplasmic incompatibility phenotype in infected insects. As such, this gene will be a very useful tool for *Wolbachia* strain typing and phylogenetic analysis, as well as understanding the molecular basis of the interaction of *Wolbachia* with its host.

The genus *Wolbachia* comprises a group of maternally inherited intracellular bacteria that have been identified in a wide range of arthropod hosts. Indeed, some surveys suggest that around 16% of all insect species may be naturally infected with this agent (35). *Wolbachia* strains are best known for the reproductive distortions they generate in the arthropods they infect. These phenotypes include the feminization of genetic males, induction of parthenogenetic development, and most commonly the expression of cytoplasmic incompatibility (CI) (34). CI expression usually results in embryonic death in crosses in which the male insect parent is infected with a *Wolbachia* strain and the female parent is either uninfected or infected with a different *Wolbachia* strain. It appears that a *Wolbachia* strain is able to imprint the sperm of insects it infects through an unknown mechanism and that this imprint is rescued only in eggs that are infected with the same *Wolbachia* strain.

Through the action of these various reproductive manipulations, *Wolbachia* is able to efficiently invade host populations without being infectious or moving horizontally between individuals at an appreciable rate. It has been suggested that the ability of *Wolbachia* to actively invade populations could be used as a vehicle to drive desirable genotypes into wild insect populations, e.g., genes that prevent insect disease vectors from transmitting pathogens to humans, livestock, or plants (1, 30).

While much is known about the phenomenology and population genetics of *Wolbachia* infections, very little is known about the molecular mechanisms that underlie the interaction of this agent and the insect. In previous studies, we have identified by metabolic labeling the major proteins that *Wolbachia* synthesizes in vivo (27). One of these proteins showed size polymorphism between *Wolbachia* strains that correlated with the ability of a given *Wolbachia* strain to induce the CI phenotype in *Drosophila* species. In this paper, we report the purification of this protein and the cloning and characterization of its gene from a number of different *Wolbachia* strains. In addition, we show that the 5′ noncoding region can serve as a functional promoter in *Escherichia coli.*

* Corresponding author. Mailing address: Yale University, 60 College St., New Haven, CT 06520-8034. Phone: (203) 785 3285. Fax: (203) 785 4782. E-mail: scott.oneill@yale.edu.

MATERIALS AND METHODS

Insect strains. The following strains of *Drosophila* species which harbor *Wolbachia* infections were used: *D. melanogaster* CantonS (DMCS) (13), *D. melanogaster* Harwich (DMHarwich) (13), *D. simulans* Coffs Harbour (DSCoffs) (10), *D. simulans* Hawaii (DSH) (23), *D. simulans* Riverside (DSR) (11), tetracycline-treated DSR cured of *Wolbachia* (DSRT), and *D. simulans* Watsonville(Mauritiana) (DSW/Mau) carrying the transinfected *Wolbachia* strain from *D. mauritiana* (9). In addition to these *Wolbachia* strains known to naturally infect *Drosophila* spp., the strain from the *Wolbachia*-infected almond moth *Cadra* (*Ephestia*) *cautella* was used (4). Of the various *Wolbachia* strains, DMCS, DMHarwich, DSCoffs, and DSW/Mau have been reported previously as being incapable of expressing the CI phenotype (9, 10, 13), while DSR, DSH, and the *Wolbachia* strain *C. cautella* have all been reported as strong CI expressors (4, 11, 23).

Purification of *Wolbachia*. *Drosophila* eggs from 2- to 4-h collections were dechorionated for 2 min with 2.6% sodium hypochlorite, washed with water, and packed under water in a 1.5-ml microcentrifuge tube with repeated spins for 10 s so as to accumulate 50 μl of packed eggs. Excess water was removed, and the eggs were homogenized by hand with a tight-fitting pestle (Kontes Co., Vineland, N.J.) in 100 μl of homogenization buffer (250 mM sucrose, 90 mM potassium chloride, 30 mM sodium chloride, 15 mM magnesium sulfate, 5.5 mM calcium chloride, 0.1% [wt/vol] Lubrol; ICN Inc., Costa Mesa, Calif.). After homogenization, an additional 1 ml of homogenization buffer was added and the tube was vortexed for 3 s. Cellular debris was pelleted for 5 min with $80 \times g_{max}$ at 20°C. The supernatant was centrifuged for 5 min at $4,000 \times g_{max}$. The resulting pellet was carefully resuspended with 100 μl of homogenization buffer, an additional 1 ml of homogenization buffer was then added, and the tube was vortexed for 3 s. After a 5-min spin at $300 \times g_{max}$, the supernatant was loaded onto a 13-mm-diameter filter cassette containing a 0.8- to 8-μm-pore-size glass fiber prefilter (AP 20, 13 mm; Millipore Corp., Bedford, Mass.) and a strong protein binding 3-μm-pore-size mixed cellulose membrane (SSWP; Millipore) and filtered under unit gravity. The filter cassette was washed with homogenization buffer until 1.5 ml of filtrate was obtained; the filtrate was then spun for 5 min at $5,000 \times g_{max}$, and the pellet was saved.

Purification of *Wolbachia* was monitored by comparing the protein profiles of

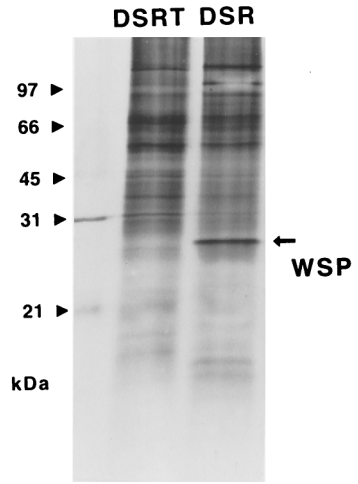

FIG. 1. WSP is the dominant protein in silver-stained SDS-polyacrylamide gels of *Wolbachia* fractions from DSR eggs; DSRT, uninfected control; left lane, marker proteins.

DSR and DSRT preparations by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis followed by silver staining. If no difference could be detected in protein profiles, the amount of *Wolbachia* was estimated by DAPI (4′,6-diamidine-2-phenylindole dihydrochloride) staining and/or PCR. DAPI staining for DNA was performed overnight by mixing equal volumes of sample preparation and a 1-μg/ml solution of DAPI in 100% methanol so that mitochondrial DNA could also be detected. Because the size of the mitochondrial DNA is only a fraction of that of *Wolbachia*, the two could easily be differentiated.

Wolbachia DNA quantities in preparations were also estimated by PCR of serial dilutions with *Wolbachia*-specific primers for the 16S rRNA gene (99F and 994R) (22) and for the *ftsZ* gene (29). To estimate whether nuclear DNA of the host was present in preparations, the single-copy gene *suppressor of sable* [*su(s)*] (33) was PCR amplified. For the *su(s)* gene, forward (5′-TCA GTA CCG CGA ACG CAG CAA ATA-3′) and reverse (5′-GCC GCC ACG TAC GTT CAT CAT CTC-3′) primers were designed; for mitochondria, 12S rRNA primers (12SAI and 12SBI) were used (22).

Protein purification and protein sequencing. Protein samples were desalted and concentrated by the method of Wessel and Flügge (36) except that acetone replaced methanol in the final step. For screening purposes, SDS–12% gels were used (15). Proteins were blotted from 10% (percentage of cross-linking monomer over total amount of monomer C = 3.3%) tricine gels (28), without a spacing gel, onto polyvinylidene difluoride membranes (Immobilon-P; Millipore) under semidry conditions, using the buffer described by Bjerrum and Schafer-Nielsen (2) with 20% methanol and 0.004% SDS. The molecular weight of the protein was estimated in 10% (C = 2.6%) tricine gels without a spacing gel. The isoelectric point was estimated in the matrix-free Rotofor system (Bio-Rad Laboratories, Hercules, Calif.) with samples prepared as described by O'Farrell (20). The protein was N terminally sequenced and, after trypsin digestion, internally sequenced at the Keck Foundation, Yale University, New Haven, Conn.).

Gene cloning and Southern blotting. Degenerate oligonucleotide primers were designed from regions of the N-terminal and internal peptide sequences (forward, 5′-TAY GTI GTI YTI CAR TAY AAY GGI GAR AT-3′; reverse, 5′-GTA IAG ICC ITC IAC ATC NAC-3′), and PCR amplification with an annealing temperature of 45°C was done on total DNA extracted from *Drosophila* DSR and DSRT eggs. A PCR product was obtained from strain DSR but not strain DSRT. This product was directly cloned into the *Eco*RV site of the pBluescript vector (Stratagene, La Jolla, Calif.) after tailing the vector with ddTTP, using terminal transferase (Boehringer Mannheim, Indianapolis, Ind.) (13a). Sequencing confirmed that the cloned fragment was derived from the gene encoding the protein, since putative translation of the cloned DNA yielded a sequence identical to the peptide sequence internal to the forward PCR primer.

The following strategies were used to clone the 5′ and 3′ coding regions of the gene as well as flanking DNA. For the 3′ end, total DNA of DSR was digested with *Eco*RI and ligated to similarly digested pBluescript. PCR was done with the primer *wsp* 115F (5′-GTG GTG CTG CAA TAC AAC-3′) and either the T3 or T7 primer which recognizes pBluescript. Then nested PCR was performed with the internal primer *wsp* 169F (5′-ATT GAA TAT AAA AAG GCC ACA GAC A-3′). The resulting PCR product of 900 bp was cloned and sequenced to confirm that it represented the 3′ end of the gene.

The 5′ flanking fragment was obtained by means of ligation-mediated PCR (LM-PCR) (17). Primer extension was performed with *wsp* 247R (5′-TGT AAC CAA ATG CAC CAC CAC CAG-3′) on *Pst*I-digested total DNA from DSR and DSRT, using *Pfu* polymerase to produce blunt ends. The products were ligated to partially double-stranded oligonucleotides formed by annealing LM-PCR 1 (5′-GCG GTG ACC CGG GAG ATC TGA ATT C-3′) and LM-PCR 2 (5′-GAA TTC AGA TC-3′). The PCR product generated with the primers LM-PCR 1 and *wsp* 247R was subjected to a nested amplification with LM-PCR 1 and the internal primer *wsp* 198R (5′-ATG AAT GTC TGT GGC CTT TTT AT-3′), of which the product (~1,200 bp) was cloned and partially sequenced (500 bp). Once the 5′ and 3′ ends were sequenced, PCR primers were designed to amplify the entire gene from infected DSR insects. Several primers were designed on the basis of the gene sequence from DSR and tested on different strains of *Wolbachia*. The primer combination *wsp* 81F (5′-TGG TCC AAT AAG TGA TGA AGA AAC-3′) and *wsp* 691R (5′-AAA AAT TAA ACG CTA CTC CA-3′) was found to be able to amplify *wsp* gene fragments from different strains of *Wolbachia*.

To determine copy number of the gene, 5 μg of total fly DNAs from DSR,

```
aagatggtac ttggataaga ggcattgccc tcctggtagg cccaaattgc aatgttcgta cagttgtgga tttctcatcc gctaacaagc agcgagatga

                                                                                                  minus 35
cgaggtcttt cgatagtaaa tcttacctgt gttagctata ccaaaaaatc aaaagtcgta tttattcatt aataaaatct ttgtct tctc ctaaaaatta

 minus 10      plus 1
gttaaaatgt agaatatttt gtgttattaa gttagcactt gacttaactt tagttgttac ttagactgct aaagtggaat tttaattttt tttacgagga

    signal sequence
     M  H    Y  K  K    F  F  S  A    A  A  L    A  T  L    L  S  L  S    N  S  A    F  S  D    P  V  G  P    I  S  D
caattatgca ttataaaaag ttttttcag cagccgcttt agcaacgttg ctaagtttat caaactctgc tttttcagat cctgttggtc caataagtga

 E  E  T    S  Y  Y  V    R  L  Q    Y  N  G    E  I  L  P    L  F  T    K  I  E    G  I  E  Y    K  K  A    T  D  I
tgaagaaact agctactacg ttcgtttgca atacaacggt gaaattttac ctcttttcac aaaaattgaa ggtattgaat ataaaaaggc cacagacatt

 H  N  P  L    K  A  S    F  I  A    G  G  G  A    F  G  Y    K  M  D    D  I  R  V    D  V  E    G  L  Y    S  Q  L  N
cataatccat taaaagcatc ttttatagct ggtggtggtg catttggtta caaaatggac gacatcaggg ttgatgttga agggctttat tcacagctaa

                                                1st putative transmembrane domain
   K  N  D    V  T  G    A  A  F  N    P  D  T    V  A  D    S  L  T  A    I  S  G    L  V  N    V  Y  Y  D    I  A  I
acaaaaatga tgttacaggt gcagcattta acccagatac tgttgcagac agtttaacag caatttcagg gctagttaac gtttattacg atatagcaat

          2nd putative transmembrane domain
 E  D  M    P  I  T  P    Y  V  G    V  G  V    G  A  A  Y    I  S  T    P  L  K    D  A  V  N    D  Q  K    S  K  F
tgaagatatg cctatcactc catatgttgg tgttggtgtt ggtgcagcgt atattagcac tcctttgaaa gacgctgtga atgatcaaaa aagtaaattt

 G  F  A  G    Q  V  K    A  G  V    S  Y  D  V    T  P  E    V  K  L    Y  A  G  A    R  Y  F    G  S  F    G  A  H  F
ggttttgctg gtcaagtaaa agctggtgtt agttatgatg taactccaga agtcaaactt tatgctggag ctcgttattt cggttctttt ggtgctcatt

   D  K  D,    T  A  A    A  S  K  D    K  G  E    L  K  V    L  Y  S  T    V  G  A    E  A  G    V  A  F  N    F  .
ttgataaaga tactgctgca gcaagcaaag acaaggggga actcaaagtt ctttacagca ctgttggtgc agaagctgga gtagcgttta atttttagtc

tatctcagtc ttgttatccc agtggttttt tcttgtcatt caagtagcat gacactggga tctagaaaaa agaaa
```

FIG. 2. *wsp* gene sequence and deduced amino acid sequence from *Wolbachia* harbored by DSR. Several putative regulatory regions are indicated.

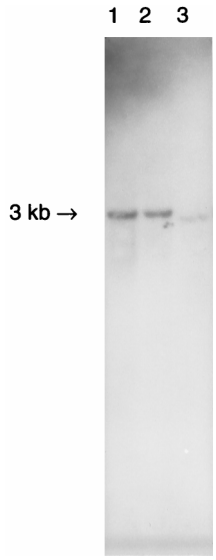

FIG. 3. Southern blot showing a single hybridizing fragment in total fly DNA digested with *Eco*RI and probed with 610 bp of *wsp* from DSR, indicating a single-copy gene. Lanes: 1, DSR; 2, DSH; 3, DSW/Mau.

DSRT, DSH, and DSW/Mau were digested with *Eco*RI. DNA was transferred to a Zeta-Probe membrane (Bio-Rad) by vacuum blotting after separation on a 1% agarose gel. The membrane was baked for 2 h at 80°C. Prehybridization was done overnight at 65°C with 0.5 M phosphate buffer (pH 7.2) containing 7% SDS, 1% bovine serum albumin, and 1 mM EDTA. The probe spanning the coding region of *wsp* was obtained by PCR with primers *wsp* 81F and *wsp* 691R and labeled with [α-^{32}P]dATP, using a random-primed DNA labeling kit (Boehringer Mannheim). The labeled probe was added directly into the prehybridization solution and left overnight for hybridization. The membrane was washed twice with low-stringency buffer (5% SDS, 40 mM phosphate) for 5 min, then twice in high-stringency buffer (1% SDS, 40 mM phosphate), both at 65°C. Additional Southern blots of total fly DNA from DSR were performed with *Ssp*I, *Taq*I, and *Tth*111I (all cutting in the 5′ noncoding region), with *Bsr*I and *Nla*III (both cutting in the 3′ noncoding region), and as a double digest with *Xba*I and *Bbs*I (cutting in the 5′ and 3′ regions, respectively).

Functional testing of the *wsp* promoter region in *E. coli*. The upstream *wsp* sequence (nucleotides −305 to −1) was PCR amplified by using specific primers with incorporated *Sac*I and *Sac*II restriction sites 5′-GAG CTC AAG ATG GTA CTT GGA TAA GA-3′ and 5′-CCG CGG AAT TGT CCT CGT AA-3′. To introduce additional restriction sites adjacent to the *wsp* sequence, this amplification product was double digested with *Sac*I and *Sac*II and ligated into the pEGFP-N1 vector (Clontech Laboratories Inc., Palo Alto, Calif.). To examine the *wsp* upstream region's promoting capability, it was cloned into the pKK232-8 vector (Pharmacia Biotech Inc., Piscataway, N.J.). For comparison, a *lac* promoter was similarly introduced into pKK232-8. To accomplish this, the *wsp* construct was digested with *Sac*II, blunted with T4 polymerase, and digested with *Bgl*II. This was then ligated to the pKK232-8 vector which had been digested with *Hin*dIII, blunted with T4 polymerase, and digested with *Bam*HI. The *lac* promoter was excised from pGFP (Clontech) by digestion with *Sap*I, blunting with T4 polymerase, and digestion with *Hin*dIII. It was then ligated to the pKK232-8 vector carrying an ampicillin resistance gene which had been digested with *Sal*I, blunted with T4 polymerase, and digested with *Hin*dIII. Transformants were first selected on ampicillin plates for the pKK232-8 vector-encoded resistance and then on chloramphenicol (100 μg/ml) plates for chloramphenicol acetyltransferase (CAT) activity. Each construct was verified by restriction analysis.

The promoter strength of the *wsp* upstream sequence was determined by using a Flash CAT 1-deoxyCAM assay (Stratagene) for CAT. As a comparison, the *lac* promoter was also examined. For the assay, a 1:160 dilution of an overnight culture was grown for 3 h under selection with ampicillin. In one culture, 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) was added. At the end of this growth, an equivalent concentration of each culture was verified by plating a serial dilution. The cells in 1 ml of each culture were pelleted and washed with phosphate-buffered saline. Each pellet was resuspended in 1 ml of 100 mM KCl–25 mM HEPES–0.1 mM EDTA–12.5 mM $MgCl_2$–10% glycerol, 0.1% Nonidet P-40–1 mM dithiothreitol, 0.1 mM phenylmethylsulfonyl fluoride–500 μg of lysozyme per ml (pH 8.0). This suspension was then shaken on ice for 30 min and sonicated. Following centrifugation for 5 min at 21,000 × g_{max}, 55-μl aliquots of the supernatants were assayed as recommended in the Flash CAT protocol. The results were quantified with the NIH Image software package (25). After the assay, the gene construct was reisolated and sequenced; it was identical to the original construct.

Nucleotide sequence accession numbers. The sequences reported in this paper have been deposited in the GenBank database under accession no. AF020070, AF020065, AF020066, AF020067, AF020068, AF020069, and AF020075.

RESULTS

Initial attempts to purify *Wolbachia* strains from *Drosophila* eggs were based on protocols for the purification of *Rickettsia* species from cultured cells, which rely on isopycnic gradient separations. These protocols did not result in a sufficiently complete separation of *Wolbachia* from host material. The main reason for this failure may lie in the fact that *Wolbachia* resides in a host vacuole that is intimately associated with host cytoskeletal elements. We chose instead to use a method utilizing a detergent and filtration membranes. Critical parameters in the purification were the detergent concentration and the pore size of the filtration membranes. The use of Lubrol at 0.1% (range, 0.01 to 1%) and a 3-μm-pore-size membrane provided an optimal trade-off between yield and purity. The major contaminants were mitochondria, as judged by DAPI staining and PCR analysis, and host membrane material. A higher purity with a lower yield could be achieved by omitting the detergent and using a 5-μm-pore-size membrane instead. The detergent removes host vacuolar material from *Wolbachia* and facilitates the passage of *Wolbachia* through the filtration membrane by preventing an interaction between *Wolbachia* and the filtration membrane. This makes the operational filtration size of *Wolbachia* smaller under detergent, but at the same time, the detergent generates membrane aggregates from host material that contaminate the *Wolbachia* fractions. In the absence of detergent, a larger pore size is needed. Since no aggregates of host membranes are generated in detergent-free samples, the larger-pore-size filtration results in preparations with less host contamination.

When purified samples of *Wolbachia* from DSR eggs were run on silver-stained SDS-gels, a dominant protein with an apparent molecular mass of around 28 kDa was resolved. In tricine-based gels, the same protein had an apparent size of around 22 kDa, which is closer to the size predicted from the gene sequence (Fig. 1). Under denaturing conditions, this protein focused at a pI between 4.5 and 4.7. After cell lysis with either water or treatment with detergents, the protein stayed associated with membrane fractions. Its behavior on treatment with the detergent Sarkosyl, which selectively solubilizes inner membrane proteins, suggested that the protein is located in the outer membrane of the bacterium (5, 7): the *Wolbachia* protein stays membrane associated with 1% Sarkosyl but is solubilized with 2.75%. These data and the homology to other bacterial

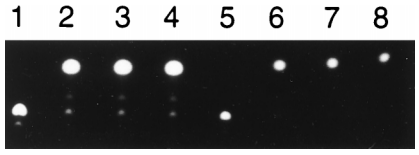

FIG. 4. *E. coli* strains transformed with different plasmid constructs were used to assay the *wsp* upstream sequence for its ability to drive expression of a *CAT* gene. Shown is a reverse negative of an autofluorogram of a thin-layer chromatogram of CAT activity revealed by the cleavage of a CAT substrate. Presence of CAT is demonstrated by cleavage of the fluorescently labeled 1-deoxyCAM substrate (lower band) to generate the higher band. Lanes 5 to 8 show 10-fold dilutions of lanes 1 to 4. The unmodified pKK232-8 vector (lanes 1 and 5) shows no CAT activity. The insertion of the *wsp* upstream sequence at the 5′ end of the *CAT* gene induced its expression (lanes 2 and 6). Similarly, the insertion of the *lac* promoter induced expression of the *CAT* gene in the presence (lanes 3 and 7) and in the absence (lanes 4 and 8) of IPTG.

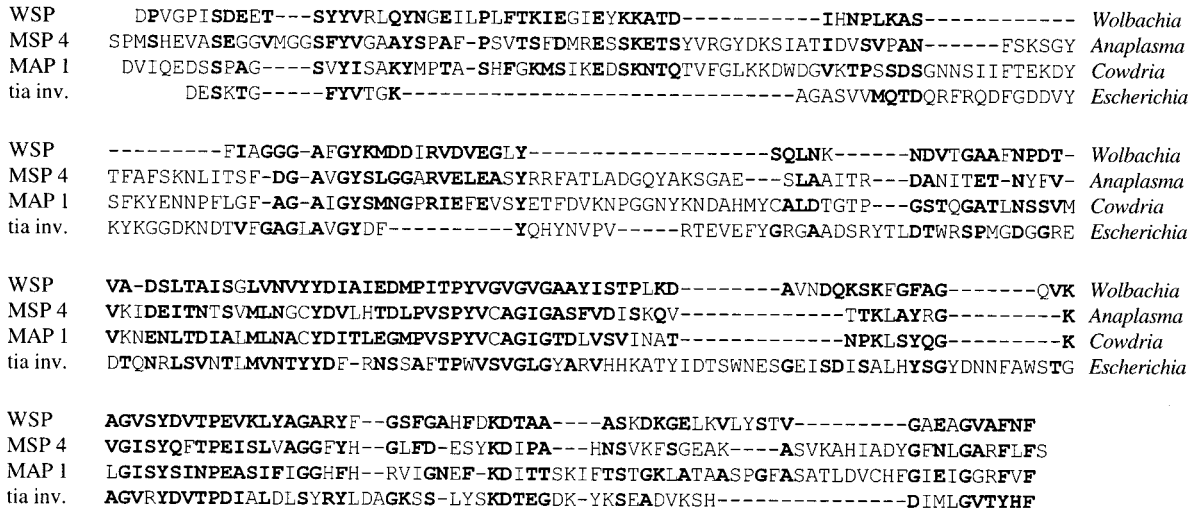

FIG. 5. Alignment of the deduced amino acid sequence of WSP from DSR with sequences of homologous outer surface proteins: MSP4 from *A. marginale* (L01987), MAP1 from *Cowdria ruminantium* (U50832), and the tia invasion (inv.) determinant from *E. coli* (U20318). Only the second putative transmembrane domain shown in Fig. 2 is shared by all homologs. Amino acids identical or similar (3 distance units out of 22, using the PAM250 table of the MegAlign software program, version 3.11 [DNASTAR Inc., Madison, Wis.]) to those in the WSP sequence are in boldface.

outer membrane proteins as shown later indicate that this *Wolbachia* protein is a surface protein; therefore, we named it *Wolbachia* surface protein (WSP).

Sequencing of the N terminus of WSP revealed the following 30-amino-acid sequence: N′-(A, S, G [V])-P-I (P)-S-D-E-E-T-X-Y-Y-V-(V)-L-Q-Y (Q)-N-G-Q-I-L-P-X-F-X-K-(I)-C′. After trypsin digestion, the sequence of an internal fragment was obtained: N′-X-P-V (I)-X-P (I, A, D)-(I)-(I)-(D)-C′. These sequences were sufficient to generate a nested PCR protocol that led to the cloning of the complete gene.

The cloned gene from DSR *Wolbachia* (*w*Ri) contains an open reading frame (ORF) that codes for a protein of 230 amino acids (690 bp, 24,633 Da) starting with an N-terminal methionine (Fig. 2). However, sequencing of the purified mature protein showed that the N-terminal amino acid was aspartic acid. The 24 additional coded amino acids appear to constitute a well-recognized gram-negative signal secretion sequence with a cleavage site in front of the aspartic acid as predicted by the von Heijne algorithm (18).

On Southern blots of total *D. simulans* DNA from DSR, DSH, and DSW/Mau, digested with *Eco*RI, and DSR, digested with either *Bsr*I, *Nla*III, *Ssp*I, *Taq*I, *Tth*111I, or *Xba*I-*Bbs*I, and probed with a 610-bp fragment (bp 81 to 691) of *wsp* revealed only one hybridizing fragment, indicating that *wsp* is a single-copy gene (Fig. 3). Southern blots of total fly DNA from uninfected DSRT, digested with *Eco*RI, did not show any hybridization when probed with the same *wsp* fragment.

Because WSP is an abundant *Wolbachia* protein, its expression is likely to be driven by a very strong promoter. To examine whether the DNA sequence immediately upstream of the *wsp* ORF could function as a promoter, 305 bp of 5′ flanking DNA was cloned into the pKK232-8 vector in front of the *CAT* gene. For comparison, the *lac* promoter was similarly cloned into this vector. Selection of transformants on the basis of their resistance to chloramphenicol indicated that a functional promoter was driving *CAT* gene expression. The level of CAT activity induced by the *wsp* promoter was quantified and compared with levels induced by the *lac* promoter. Calculations based on densitometry scores of the data in Fig. 4 revealed that the *wsp* sequence induced 33% more CAT activity than the *lac* promoter with IPTG induction and 116% more than the *lac* promoter without IPTG induction. The pKK232-8 vector without a promoter insert did not show any CAT activity. These results demonstrate that the upstream *wsp* sequence not only is recognized in the unrelated bacterium *E. coli*, but also contains sequences that function as a stronger promoter than the *lac* promoter.

DISCUSSION

WSP is a membrane protein, and its solubilization behavior in Sarkosyl suggests that it is an outer membrane protein. The sequence-based predicted protein localization site according to Klein et al. (14) is equally probable for either the outer membrane or periplasmic space. In addition, the protein contains a

FIG. 6. The amino acid sequence alignment of a segment of WSP of three *Wolbachia* strains capable of expressing CI (DSR, DSH, and *C. cautella*) compared to four strains incapable of expressing the phenotype (DSW/Mau, DS Coffs, DM Harwich, and DM CantonS). Differences are in boldface.

carboxy-terminal phenylalanine that is considered to be essential for the correct assembly of bacterial outer membrane proteins (31). The TMpred algorithm (12) predicts two membrane-spanning regions, surprisingly, both with the same strong preference for an outside-inside orientation: 88 to 107 (20 amino acids; score, 551) and 111 to 128 (18 amino acids; score, 575) (Fig. 2).

The gene encoding WSP shows homology with genes encoding outer membrane proteins of the closely related rickettsiae. The greatest homology is shared with major surface protein 4 (MSP4) of *Anaplasma marginale* (31% similarity to *WSP* from DSR) (19), major antigenic protein 1 (MAP1) of *Cowdria ruminantium* (28% similarity) (26), and the tia invasion determinant of an enterotoxigenic strain of *E. coli* (24% overall similarity; high similarity is found mainly in the middle third of the sequence) (Fig. 5). The tia invasion determinant is thought to be responsible for both epithelial adherence and invasion of enteropathogenic and enteroaggregative strains of *E. coli* and *Shigella sonnei* (8). In addition, fragments of the *wsp* gene show similarity with (i) the scrub typhus antigen (TSA, STA56) from *Orientia* (*Rickettsia*) *tsutsugamushi*, where the protein may function as an adherence factor potentiating rickettsial adsorption to the host cell surface and as a virulence determinant of individual rickettsial strains (21, 32), (ii) MSP2 from *A. marginale* (6), and (iii) a highly conserved outer membrane protein from *Neisseria gonorrhoeae* (16). While many bacterial outer surface proteins (e.g., MSP2 from *A. marginale* [24]) are encoded by multicopy gene families, Southern blot analyses indicate that *wsp*, like its closest known relative, MSP4, is a single-copy gene.

Alignment of the putative amino acid sequences of the *wsp* gene from different strains of *D. simulans*, *D. melanogaster*, and the moth *C. cautella* show a high similarity between these proteins. Nevertheless, there are small regions that are highly variable. Two strains of *D. melanogaster*, DMCS and DMHarwich, and one strain of *D. simulans*, DSCoff, carry *Wolbachia* strains that are incapable of inducing the phenotype of CI. These strains have been extensively studied to exclude epiphenomena as the reason for the lack of expression of CI (reference 10 and unpublished data). Interestingly, in the three strains of *Wolbachia* we investigated that were incapable of expressing the CI phenotype, all shared similar sequences that were not conserved in the strains that were capable of expressing the phenotype (Fig. 6). The conserved differences that we see in this alignment between expressing and nonexpressing strains might reflect a functional difference related to the CI phenotype. The protein composition in the outer membrane may reflect an adaptation and specialization to the intimate interaction of this intracellular bacterium with the insect host cell. Alternatively, the sequence variability may reflect strain differences which correlate with the phenotypic differences of these strains. In either case, the *wsp* gene is an excellent candidate for strain typing different *Wolbachia* strains as well as providing characters for a fine-scale phylogeny of *Wolbachia* strains (37). At the present time, no other *Wolbachia* gene which can be used to adequately resolve the evolutionary relationships between different *Wolbachia* strains has been cloned. In addition, the variability observed between *wsp* sequences of different *Wolbachia* strains suggests that this gene should be able to be used to predict reproductive phenotypes generated by different strains (3) as well as be used as a marker to track multiple *Wolbachia* infections within individual hosts. This, in turn, should greatly enhance our ability to study the biology of this fastidious microorganism.

ACKNOWLEDGMENTS

Weiguo Zhou and Henk R. Braig contributed equally to this report.

We gratefully acknowledge A. Hoffmann for supplying *D. simulans* Coffs Harbour S20, strain, R. Carde for supplying the strain of almond moth *C. cautella*, and M. Turelli for the *suppressor of sable* primers.

This work was supported by grants from the National Institutes of Health (AI 34355, AI 07404, and AI40620), the McKnight Foundation, and the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases.

REFERENCES

1. **Beard, C. B., S. L. O'Neill, R. B. Tesh, F. F. Richards, and S. Aksoy.** 1993. Modification of arthropod vector competence via symbiotic bacteria. Parasitol. Today **9:**179–183.
2. **Bjerrum, O. J., and C. Schafer-Nielsen.** 1986. Buffer systems and transfer parameters for semidry electroblotting with a horizontal apparatus, p. 315–327. *In* M. J. Dunn (ed.), Electrophoresis 86. VCH, Weinheim, Germany.
3. **Bourtzis, K., S. L. Dobson, H. R. Braig, and S. L. O'Neill.** 1998. Rescuing *Wolbachia* have been overlooked. Nature **391:**852–853.
4. **Brower, J. H.** 1976. Cytoplasmic incompatibility: occurrence in a stored product pest *Ephestia cautella*. Ann. Entomol. Soc. Am. **69:**1011–1015.
5. **Chart, H.** 1994. Bacterial fractionation and membrane protein characterization, p. 1–10. *In* H. Chart (ed.), Methods in practical laboratory bacteriology. CRC Press, Boca Raton, Fla.
6. **Eid, G., D. M. French, A. M. Lundren, A. F. Barbet, T. F. McElwain, and G. H. Palmer.** 1996. Expression of major surface protein 2 antigenic variants during acute *Anaplasma marginale* rickettsemia. Infect. Immun. **64:**836–841.
7. **Filip, C., G. Fletcher, J. L. Wulff, and C. F. Earhart.** 1973. Solubilization of the cytoplasmic membrane of *Escherichia coli* by the ionic detergent sodium-lauryl sarcosinate. J. Bacteriol. **115:**717–722.
8. **Fleckenstein, J. M., D. J. Kopecko, R. L. Warren, and E. A. Elsinghorst.** 1996. Molecular characterization of the tia invasion locus from enterotoxigenic *Escherichia coli*. Infect. Immun. **64:**2256–2265.
9. **Giordano, R., S. L. O'Neill, and H. M. Robertson.** 1995. *Wolbachia* infections and the expression of cytoplasmic incompatibility in *Drosophila sechellia* and *D. mauritiana*. Genetics **140:**1307–1317.
10. **Hoffmann, A. A., D. Clancy, and J. Duncan.** 1996. A naturally-occurring *Wolbachia* infection in *Drosophila simulans* that does not cause cytoplasmic incompatibility. Heredity **76:**1–8.
11. **Hoffmann, A. A., M. Turelli, and G. M. Simmons.** 1986. Unidirectional incompatibility between populations of *Drosophila simulans*. Evolution **40:** 692–701.
12. **Hofmann, K., and W. Stoffel.** 1993. TMbase—a database of membrane spanning proteins segments. Biol. Chem. Hoppe-Seyler **374:**166.
13. **Holden, P. R., P. Jones, and J. F. Brookfield.** 1993. Evidence for a *Wolbachia* symbiont in *Drosophila melanogaster*. Genet. Res. **62:**23–29.

13a. **Halton, T. A., and M. W. Graham.** 1990. A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. Nucleic Acids Res. **19:**1156.

14. **Klein, P., M. Kanehisa, and C. DeLisi.** 1985. The detection and classification of membrane-spanning proteins. Biochim. Biophys. Acta **815:**468–476.
15. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:**680–685.
16. **Martin, D., N. Cadieux, J. Hamel, and B. R. Brodeur.** 1997. Highly conserved *Neisseria meningitidis* surface protein confers protection against experimental infection. J. Exp. Med. **185:**1173–1183.
17. **Mueller, P. R., B. Wold, and P. A. Garrity.** 1992. Ligation-mediated PCR for genomic sequencing and footprinting, p. 15.5.1–15.5.26. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current protocols in molecular biology, vol. 2. John Wiley & Sons, New York, N.Y.
18. **Nielsen, H., J. Engelbrecht, S. Brunak, and G. von Heijne.** 1977. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. **10:**1–6.
19. **Oberle, S. M., and A. F. Barbet.** 1993. Derivation of the complete msp4 gene sequence of *Anaplasma marginale* without cloning. Gene **136:**291–294.
20. **O'Farrell, P. H.** 1975. High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. **250:**4007–4021.
21. **Ohashi, N., H. Nashimoto, H. Ikeda, and A. Tamura.** 1992. Diversity of immunodominant 56-kDa type-specific antigen (TSA) of *Rickettsia tsutsugamushi*. Sequence and comparative analyses of the genes encoding TSA homologues from four antigenic variants. J. Biol. Chem. **267:**12728–12735.
22. **O'Neill, S. L., R. Giordano, A. M. E. Colbert, T. L. Karr, and H. M. Robertson.** 1992. 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. Proc. Natl. Acad. Sci. USA **89:**2699–2702.
23. **O'Neill, S. L., and T. L. Karr.** 1990. Bidirectional incompatibility between conspecific populations of *Drosophila simulans*. Nature **348:**178–180.
24. **Palmer, G. H., G. Eid, A. F. Barbet, T. C. McGuire, and T. F. McElwain.** 1994. The immunoprotective *Anaplasma marginale* major surface protein 2 is

encoded by a polymorphic multigene family. Infect. Immun. **62:**3808–3816.
25. **Rasband, W.** 1997. NIH Image, V. 1.61. National Institutes of Health, Bethesda, Md.
26. **Reddy, G. R., C. R. Sulsona, R. H. Harrison, S. M. Mahan, M. J. Burridge, and A. F. Barbet.** 1996. Sequence heterogeneity of the major antigenic protein 1 genes from *Cowdria ruminantium* isolates from different geographical areas. Clin. Diagn. Lab. Immunol. **3:**417–422.
27. **Sasaki, T., H. R. Braig, and S. L. O'Neill.** Analysis of *Wolbachia* protein synthesis in *Drosophila in vivo*. Insect Mol. Biol., in press.
28. **Schägger, H.** 1994. Denaturing electrophoretic techniques, p. 59–80. *In* G. von Jagow and H. Schägger (ed.), A practical guide to membrane protein purification. Academic Press, San Diego, Calif.
29. **Sinkins, S. P., H. R. Braig, and S. L. O'Neill.** 1995. *Wolbachia* superinfections and the expression of cytoplasmic incompatibility. Proc. R. Soc. Lond. B Biol. Sci. **261:**325–330.
30. **Sinkins, S. P., C. F. Curtis, and S. L. O'Neill.** 1997. The potential application of inherited symbiont systems to pest control, p. 155–175. *In* S. L. O'Neill, A. A. Hoffmann, and J. H. Werren (ed.), Influential passengers: inherited microorganisms and arthropod reproduction. Oxford University Press, Oxford, England.
31. **Struyvé, M., M. Moons, and J. Tommassen.** 1991. Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein. J. Biol. Chem. **218:**141–148.
32. **Takahasi, M., H. Urakami, Y. Yoshida, Y. Furuya, H. Misumi, E. Hori, A. J. Kawamura, and H. Tanaka.** 1997. Occurrence of high ratio of males after introduction of minocycline in a colony of *Leptotrombidium fletcheri* infected with *Orientia tsutsugamushi*. Eur. J. Epidemiol. **13:**79–86.
33. **Voelker, R. A., W. Gibson, J. P. Graves, J. F. Sterlin, and M. T. Eisenberg.** 1991. The *Drosophila* suppressor of sable gene encodes a polypeptide with regions similar to those of RNA-binding proteins. Mol. Cell. Biol. **11:**894–905.
34. **Werren, J. H.** 1997. Biology of *Wolbachia*. Annu. Rev. Entomol. **42:**587–609.
35. **Werren, J. H., D. Windsor, and L. R. Guo.** 1995. Distribution of *Wolbachia* among neotropical arthropods. Proc. R. Soc. Lond. B **262:**197–204.
36. **Wessel, D., and U. I. Flügge.** 1984. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. Anal. Biochem. **138:**141–143.
37. **Zhou, W., F. Rousset, and S. L. O'Neill.** Phylogeny and PCR based classification of *Wolbachia* strains using *wsp* gene sequences. Proc. R. Soc. Lond. B, in press.