

Sequences required for expression of *Bordetella pertussis* virulence factors share homology with prokaryotic signal transduction proteins

(*vir/bvgA*, *bvgB*, and *bvgC*/two-component regulatory systems/virulence regulation)

BEATRICE ARICÓ*, JEFF F. MILLER†, CRAIG ROY†, SCOTT STIBITZ‡, DENISE MONACK†, STANLEY FALKOW†, ROY GROSS*, AND RINO RAPPUOLI*§

*Sclavo Research Center, Via Fiorentina 1, 53100 Siena, Italy; †Department of Microbiology and Immunology, Stanford University, Stanford, CA 94305; and ‡Center for Biologics Evaluation and Research, Food and Drug Administration, Bethesda, MD 20892

Contributed by Stanley Falkow, May 9, 1989

ABSTRACT The *bvg* locus of *Bordetella pertussis* is required for coordinate regulation of several factors associated with virulence. The control system is modulated by various environmental signals, including low temperature, MgSO₄, and nicotinic acid. The nucleotide sequence of the *bvg* region has been determined and three open reading frames, *bvgA*, *bvgB*, and *bvgC*, are present. Twelve-base-pair linker insertion mutations in any of these open reading frames result in a Bvg⁻ phenotype. The predicted protein products of *bvgA* and *bvgC* share homology with a family of prokaryotic regulatory proteins that respond to environmental stimuli and are members of two-component sensory transduction systems. We propose a model in which BvgB and the N-terminal portion of BvgC are localized in the periplasm. Environmental signals are recognized, transduced to the cytoplasmic portion of BvgC, and then transmitted to BvgA, a positive regulator of transcription.

The host-microbe interaction that occurs during infection and disease is a dynamic process in which the infecting organism encounters a variety of external conditions. A need for efficient adaptation to changing environments has resulted in the evolution of specialized systems that control the expression of bacterial virulence (1).

Bordetella pertussis is a human pathogen that causes whooping cough. This organism produces a number of factors that appear to be involved in pathogenesis (2). Filamentous hemagglutinin (3) and pili (4) probably play a role in adherence of the bacteria to ciliated epithelial cells in the upper respiratory tract. Other products, such as tracheal cytotoxin (5), dermonecrotic toxin (6), hemolysin (6, 7), adenylate cyclase toxin (8), and pertussis toxin (7, 9), may be involved in local damage, evasion of host defenses, and systemic disease. With the exception of tracheal cytotoxin, expression of these virulence factors is coordinately regulated and requires the products of the *bvg* (*Bordetella* virulence gene) locus, which was initially designated *vir* (10-12). Positive control of the pertussis toxin operon by *bvg* has been shown to occur at the level of transcription (13).

Inactivation of the *bvg* locus by transposon mutagenesis results in a loss of virulence (7). Alternative virulence states of *B. pertussis* also occur by two mechanisms, phase variation and phenotypic modulation. Phase variation (11) is due to a low-frequency reversible alteration in the *bvg* locus that, for the Tohama III strain, has recently been shown to involve a frameshift mutation in *bvg* (14). In contrast, phenotypic modulation is a reversible response to environmental conditions. Growth at low temperature (28°C vs. 37°C), or in the presence of MgSO₄ or nicotinic acid, results in a lack of

expression of *bvg*-regulated genes and an avirulent phenotype (15, 16).

In this paper we report the complete nucleotide sequence of the *B. pertussis* *bvg* region. The predicted products of two *bvg* loci are homologous to a family of regulatory proteins that transmit sensory signals using a conserved two-component motif (1, 17). A model for the role of the *bvg* products in sensory transduction and gene regulation is also presented.

MATERIALS AND METHODS

Bacterial Strains. *B. pertussis* strains BP165 (Bvg⁺; ref. 9), BP536 (Bvg⁺; ref. 10), and the Tn5 insertion derivatives BP347 (*vir1*::Tn5; ref. 10) and BP359 (*vir2*::Tn5; ref. 11) have been described. BP370 is a virulent phase derivative of BP326 (11). *Escherichia coli* strain Sm10 (18) was used for IncP-mediated conjugative transfer from *E. coli* to *B. pertussis* as described below.

Chloramphenicol Acetyltransferase (CAT) Fusions. A promoterless CAT gene (19) was isolated as a *Bam*HI fragment and inserted into the *Bgl* II site of *bvgC* in both orientations. CAT fusions were returned to the chromosome of BP370 by homologous recombination (20). The orientation of the fusions within the *B. pertussis* chromosome was confirmed by Southern analysis. The resulting CAT fusion strains CF1470 and CF1770 carry the CAT gene in the orientations shown in Fig. 1. CAT assays were performed as described (9).

Nucleotide Sequence Determination. The nucleotide sequence of the 5995-base-pair (bp) *Eco*RI-*Xho* I fragment shown in Fig. 1 was determined by the standard dideoxy chain-termination method (21). The sequence of both strands was obtained by using synthetic oligonucleotide primers.

Linker Insertion Mutagenesis. Twelve-base-pair linker insertions were introduced into the *bvg* region at the locations shown in Fig. 1. Insertions at *Sal* I, *Pst* I, or *Mlu* I sites consisted of the sequence 5'-GACCTGCAGGTC-3', 5'-GCTCGAGCTGCA-3', or 5'-CGCGTCTCGAGA-3', respectively. Mutations were constructed in *E. coli* and transferred to *B. pertussis* by shuttle mutagenesis (20). The structures of the *bvg* loci in the resulting strains were verified by Southern blot analysis.

Sequence Analysis. The Sequence Analysis Software Package distributed by the University of Wisconsin Genetics Computer Group (22), running on a VAX/VMS computer, PC Gene (Intelligenetics), and Genepro (Riverside Scientific, Seattle) were used for protein and nucleic acid sequence analysis, database searching, and homology assessment.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: CAT, chloramphenicol acetyltransferase; ORF, open reading frame.

§To whom reprint requests should be addressed.

RESULTS

***B. pertussis* *bvg* Locus.** The *bvg* locus was initially identified by transposon Tn5 insertions in the chromosomes of *B. pertussis* strains BP347 and BP359 (10, 11). The *bvg* locus has been cloned (12) and a restriction map is shown in Fig. 1. To investigate the transcriptional activity of *bvg*, a promoterless CAT gene was inserted in both orientations at the unique *Bgl* II site within the *bvg* sequences of strain BP370. The orientations and insertion point of the *bvg*-CAT fusions carried by CF1470 and CF1770 were confirmed by Southern blot analysis and are shown in Fig. 1. The *Bgl* II site in *bvg* lies between the sites of Tn5 insertion in BP347 and BP359. Insertion of the CAT cassette at this site disrupts the *bvg* locus, CF1470 and CF1770 have a Bvg⁻ phenotype.

CAT activity was detected only in CF1770, in which the CAT gene is present in the left to right orientation as depicted in Fig. 1 (data not shown). Introduction of a plasmid carrying the wild-type *bvg* locus into these strains results in a Bvg⁺ phenotype and does not alter the relative pattern of CAT expression. We conclude from these results that the *bvg* locus is transcribed in Bvg⁺ and Bvg⁻ strains and that the direction of transcription is from left to right as shown in Fig. 1.

Nucleotide Sequence of the *bvg* Locus. The boundaries of the *bvg* locus are contained within a 6-kb DNA fragment delimited by the *Eco*RI-*Xho* I restriction sites shown in Fig. 1. In *B. pertussis* (10, 11) and *E. coli* (12), Bvg⁻ Tn5 insertions map within this region. In addition, this fragment is sufficient for transcriptional activation of an *phaB-lacZ* operon fusion in *E. coli* in a manner that responds to environmental signals that modulate virulence expression by *B. pertussis* (J.F.M., C.R., and S.F., unpublished data).

We have determined the complete nucleotide sequence of the 5995-bp *Eco*RI-*Xho* I fragment containing the *bvg* locus and have deposited this sequence in the EMBL/GenBank data base (accession no. M25401). The overall G+C content of the sequence is 65%, a value similar to that of other *B. pertussis* sequences (4, 8, 23). As summarized in Fig. 1, three tandem ORFs with codon usage patterns typical of *B. pertussis* genes are detected. These ORFs, which may form an operon, are designated *bvgA* (*Bordetella* virulence gene A), *bvgB*, and *bvgC*. Putative ribosomal binding sites for each of these loci are located 7–14 nucleotides upstream from AUG sequences. The proposed direction of *bvg* expression corresponds to the direction of transcription detected with the CAT fusions described above.

bvgA, *bvgB*, and *bvgC* are predicted to encode proteins of 209, 275, and 936 amino acids, with molecular weights of

23,000, 30,000, and 102,000, respectively. Characteristic features of a prokaryotic signal peptide are found at the N terminus of BvgB. Two basic amino acids at positions 6 and 10 are followed by 23 uncharged, predominantly hydrophobic residues. This pattern is similar to that present at the N termini of the *E. coli* Tsr, Tar, and Tap chemoreceptors (24). BvgC contains two hydrophobic segments. The first, which may also have a role in export, contains a lysine residue at position 8 followed by a relatively short 13-amino acid hydrophobic core. In addition, a sequence of 22 hydrophobic residues beginning at position 240 is predicted to form a transmembrane helix [transfer free energy = +44 kcal/mol (1 kcal = 4.18 kJ); ref. 25] and is followed by a highly basic stretch of 9 amino acids. This putative transmembrane region is similar to those found in several bacterial proteins that span the inner membrane, such as the *Agrobacterium* VirA protein (26) and the M13 procoat protein (27).

Although sequences upstream of *bvgA* are unusually rich in A+T content, putative bacterial promoters (28) could not be found. In addition, structures resembling constitutive terminators (Rho independent; ref. 29) do not appear to be present. Sequences downstream of *bvgC* show a low coding probability in the left to right orientation. The probability of coding in this region is high in the opposite strand; however, no obvious ORFs have been identified. We have determined the sites of transposon Tn5 insertion in the Bvg⁻ *B. pertussis* mutants BP347 and BP359 in relation to the *bvg* ORFs and found that these insertions map in *bvgC* and *bvgA*, respectively (Fig. 1).

BvgA and BvgC Belong to a Family of Regulatory Proteins Involved in Signal Transduction. The predicted amino acid sequences of BvgA, BvgB, and BvgC were compared to sequence entries in the National Biomedical Research Foundation database, and to sequences translated from the GenBank database. Localized regions of homology were found between BvgA, BvgC, and members of a family of bacterial proteins that mediate signal transduction using a conserved "two-component" motif (Fig. 1; refs. 17 and 30). In these systems a "sensor" protein directly or indirectly perceives sensory stimuli and transmits a signal to a "regulator" protein that functions to control transcription or some other cellular function. Sensor proteins share a homologous C-terminal domain of ≈250 amino acids that has been termed the "transmitter," and regulator proteins typically share an N-terminal region of about 115 residues, which comprises the "receiver" (30).

The central portion of BvgC (residues 423–655) shares homology with the transmitter domains of EnvZ (31), FixL

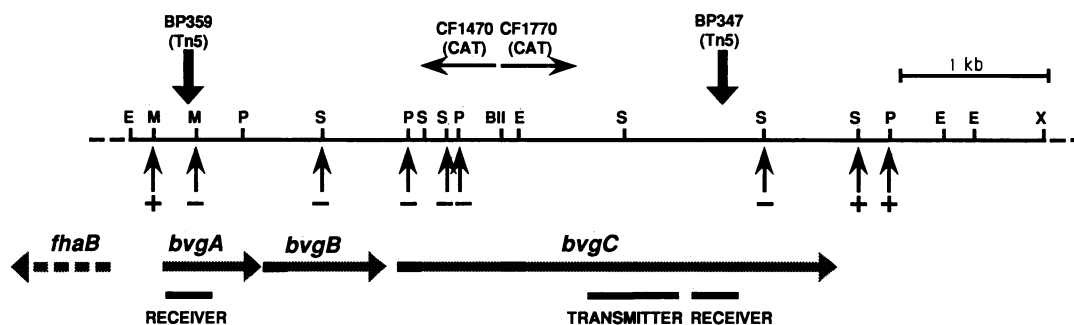


FIG. 1. Restriction map and features of the *bvg* locus. The 5995-bp *Eco*RI-*Xho* I fragment that contains the BP165 *bvg* locus is shown. The restriction map is derived from the nucleotide sequence. Arrows denote the position of transposon Tn5 insertions in strains BP359 and BP347. The location and orientation of the CAT fusions in CF1470 and CF1770 are also indicated. *bvgA*, *bvgB*, and *bvgC* are open reading frames (ORFs) predicted from the sequence. Transmitter and receiver indicate the relative locations of homologous sequences in the predicted protein products. Solid portions of the arrows indicate putative N-terminal signal sequences in BvgB and BvgC and the predicted transmembrane domain of BvgC. The location and orientation of the gene encoding filamentous hemagglutinin (*phaB*) are also shown. Positions of 12-bp linker insertion mutations at *Mlu* I, *Sal* I, and *Pst* I sites in the *bvg* region are indicated by upward vertical arrows. +, Insertions resulting in a Bvg⁺ phenotype; -, insertions giving a Bvg⁻ phenotype. E, *Eco*RI; M, *Mlu* I; P, *Pst* I; S, *Sal* I; BII, *Bgl* II; X, *Xho* I. kb, Kilobase.

(32), VirA (26), NtrB (33), PhoR (34), PhoM (35), and other transmitter proteins (Fig. 2). The similarity to FixL, a factor required for positive control of nitrogen fixation by *Rhizobium* (32), is observed over the entire length of FixL. In addition, regions of similarity with VirA, a sensor protein involved in *Agrobacterium* virulence control (26), can be found extending from the transmitter to the N terminus of BvgC. FixL and VirA contain predicted transmembrane domains similar to the one identified in BvgC, and VirA has been shown to be located in the *Agrobacterium* cytoplasmic membrane (26).

BvgA and BvgC also contain receiver domains (Fig. 3). These are located at the N terminus of BvgA (residues 1-115) and, surprisingly, at the C terminus of BvgC (residues 672-789). The homology between BvgA, FixJ (32), and UhpA (40) extends over the entire length of these proteins, whereas most other similarities are confined to N-terminal sequences. FixJ is a transcriptional activator that acts in conjunction with FixL (see above), and UhpA positively regulates expression of the sugar-phosphate transporter gene *uhpT* in *E. coli* (40). Many proteins that contain receiver modules are transcriptional activators that fall into different categories according to their C-terminal sequences (17). BvgA appears to belong to a group that previously has been shown to contain FixJ and UhpA (32). A measure of the significance of this relationship was obtained by using the RDF program of Lipman and Pearson (41). When BvgA was compared with randomly permuted versions of the FixJ or UhpA sequences, the optimized alignments scored 14.7 (FixJ) and 16.5 (UhpA) standard deviations above the mean for 20 randomized com-

parisons. Although FixJ and UhpA have been suggested to contain helix-turn-helix DNA binding motifs (32, 40), we find no such evidence of this motif in BvgA, FixJ, or UhpA using the refined method for identification of helix-turn-helix sequences recently described by Brennan and Matthews (42).

BvgA, BvgB, and BvgC Are Required for the Expression of Virulence Factors. Previous identification of the *bvg* locus in *B. pertussis* has relied on transposon Tn5 insertions that are capable of causing polar effects on distal transcription (10, 11). To determine if each of the three *bvg* ORFs is required for expression of virulence factors, 12-bp linker insertion mutations were introduced into the *Sal* I, *Pst* I, and *Mlu* I sites shown in Fig. 1. Since these mutations result in an in-frame insertion of 4 amino acids, polar effects would not be generated. Insertions at the *Mlu* I site in *bvgA*, the *Sal* I site in *bvgB*, and several *Pst* I and *Sal* I sites in *bvgC* cause a Bvg⁻ phenotype as shown by the loss of hemolytic activity. Mutations at the *Mlu* I site preceding *bvgA*, or the *Sal* I and *Pst* I sites located downstream from *bvgC*, had no effect on hemolytic activity. These results indicate that *bvgA*, *bvgB*, and *bvgC* are required for expression of virulence factors by *B. pertussis*.

DISCUSSION

The *bvg* regulon of *B. pertussis* includes several unlinked genes and operons that encode virulence factors. Expression of these loci, and their response to environmental signals, require the products of the *bvg* locus in trans (1, 10, 12, 43). An additional set of genes that are repressed by *bvg* and

BvgC	5	PVKVAVLNLFAFPFLFRTEDEQFGGISAAVLQQLQRTGLDFEIIIGVD. TVEELIAKLRSGEADNAGALFVNSAR. ESF. . LFSRSPYVRNGMVIIVTRQDPDAPVDADHLDGR. TV	115
VirA	32	AVLAIGPKMNEKSEIAELTLEQSIDVDCALLQRNVLRA. HAGLLRNYRPLMVRPLGRVRSIIANQQ. LFKK. ARVEDVUGLSELLARLKSSINT. TDAVASFGAQNWVDFSL	141
FixL			46
BvgC		MIRAEPICEGLLLSFIPAILVVALIGGRNP. ILFAAGLSLVAAVSH	
VirA			228
FixL		ALVRNSAAIPLLQRRYPQ. AKVVTADNPSEAMLMVANGQADAVVQTIASASYVVRNYFAGKLRIASALDLPAPAEIALATTRGQTELMSILNKALYSINDE. LAS. IISRMRGSDG	228
BvgC		ATFFQSS. ISSLLRTSDSRDLNAKV. PELGYL. HLQFS. FRPWTELALQITQSLDQLQMSINADKVAIQEIVRNRGRVILGVLPLNETVKLVQ. ASGTFENTKQLQRAVLEADSLARV	254
VirA			254
FixL		QQIS. SADGPSVVELLVFG. SAVLLVALGVELEARRADRTRDVEDVVRADAHLSRIIDTVPDAMVVSATDGTIVSFNAAVRQFGVAEEVIGQNLK. ILMPEYRHEHDGVLQR	159
BvgC		DPRTWYAYRNEIYLLIGLGLLSALLFLSHVYLLRQIRQRKRAERLNQDLEFMRVLDIGTNPMPYVRDKEGRHLLCNDAYLDTFGVITADAVLTKTIPKAVVGDPAALAREMHEFLTR	347
VirA			368
FixL		YMATGEKRIIGIDRVVSGQRKDG. TFPMLAVGEMRSGGERFFTCFIRDLTEREESAARLEQIQELARLARL. NEMGEMASTLAHELNQPLSA	252
BvgC		VAAERPEPFEDRIVLHGRTRHVVQWTIPYGDLSGELKXGIIIGWIDITERAELRRLHDAKESADANRAKTTFL. ATMSHEIRTPHMA	435
VirA		VSLVQADENGSIIFRDYFARKASCNFEDAPGRNALVAFKVSRLVAVFGLRFRDQVQASSEVQMLMELAGCVSHVVVIRCKQTRDRLERLKHAEERLEAVGLLGGIAHEFNHILGV	487
EnvZ			251
NtrB			147
PhoR			221
PhoM			273
FixL		IANYSHGCTRLLRDMDD. AVATRIREALEEVASQSLRAGQIIKHLREFVTKGETEKAPEDIRKLVESAAALALVGSREQQVTVFEY. LPGAEHVLVD. RIQQVQVILNMR	361
BvgC		IIGHLELA. LLRPTDQEPDRQSIQVAYDSARSLLELIGDILDIKIEAGKFDLAVPRTALRVLPDGAIRVDFGLARQKIELVLTDTIVGVDDVLD. PLRMMQVLSNLVG	544
VirA		ILGYA. MAQNILHRR. TYARHYIDIRINAESNRRLIIDQILALSRRRRTARFPNLSALVREIAPSLRVALPSEVEVDVFIQ. SAQMIVCGNPLEIEQILQNLCK	588
EnvZ		I. RLATEMSEDDQY. LAESINKDIEECNAIIIEQPIDVLRTOQE. MPMEMADLNAVLEGEVIAAESGYERELETALY. PGSEVEMPLESIKRAVANMVV	346
NtrB		L. RGAALQLSKALP. DPSLLEYT. KVIIEQADRLRNLVDRLLGQQL. PGRVTVESIHKVAE. RVVTLVSMELPDNVRLI. RDVDPSPLELAHPDQIEQVLLNIVR	247
PhoR		L. QGVLEHMEQPL. EGAVREKALHTMREQTQRMEGVLVKQLLTL. KIEAAPTHLNEKVVVPHML. RVVEREAQTLISQNKQTF. TF. EIDNGLKVSQMEQRLRSAINLVY	327
PhoM		I. RGAAILREGPPPE. VVA. RFTDNLITQARMQALVETLRLQA. RLENRQEVVLT. VDVAALFRVRSSEARTVQLAEKIKITLV. TPTEV. NVAAEPALLEQ. ALGNLLD	377
FixL		NAIEAM. RHVDRELITIRTPADPGEVAVVVEDTGGGIPPEVAGLQFKPFV. TTK. ASGMCIGLSISKRIVEAHGEMTVSKNEAGGATFRFT. LPAYLDERIVAND#	464
BvgC		NAIKFT. TEGQVVLAVTARPDGDAHVQFVSVDTCGCISEADQQLFKPFSVQGSAEAGP. APGTGLGLSISRRVELMCGTIVMRSAPCVGTIVSVDRLTM. VEKSVQAP.	655
VirA		NAAESV. YRSFVMKHVLANGTIPAGDYILLSFEDNGGGIQAALPHIPEFPFRTR. AQCQGTGLGLSTVGHVGSAMAGFVDVISTVGRGTRFDIYLP. TSAK. KPVNS.	704
EnvZ		NAARY. G. NG. NI. K. VSSGTEPNRAMEQVEDDGGPIAPEQRKHLGQPFVGRD. SA. RTISGTLGLAIVQRIVDNHNGMELGTSERCGLSIRAWLPPVPTRAQGTTEK#	450
NtrB		NALQALGPEGLLRTTRTAPQLTHGERYLRARIDVDMGPGIIPHLDLTFYFVSG. R. EGGTGLGLSIAARNLIDQHSCKIEFTSWP. GHTEFSVYLPTRK#	349
PhoR		NAVN. HPELTHITVRMQRVPHGAESVEDNGPIAEPHILPRLTERFYRDKARSQT. GSGGLGLAIVGHAVMHESRLNIESVCKGTRFSPVIFERLIANKNSD#	431
PhoM		NAID. FTPPESCITLSAEVDQEHVTLKVLDTGSGIPDVALSRIFERYSLPRANG. Q. KSSGLGLAFVSEVALRFGNEVTLRNVQEGCVLASLRL. HRHFT#	474

FIG. 2. Amino acid sequence comparisons between BvgC and members of bacterial sensory transduction systems that contain transmitter domains. Identical amino acids (*) and similar amino acids (:) between the adjacent sequences are indicated. The groups of similar amino acids are C, S, T, P, A, G; N, D, Q, E; M, I, L, V; F, Y, W. Proposed transmembrane domains and signal sequences are indicated by dashed lines. References for sequences are given in the Results.

BvgA	MYNKVLIIDHPVLRFAVRVLMKEGFEVIGET .DNGIDGLKIARE .KIPNLVLDIGIKPLDGLVIARLOSGLPLRVLVLTGQPPSLFARRCLNSGAAGFVCKHE	106
FixJ	MDYTYHIVDDEEPVRSKSLAFVLMYMGFAVKMH .QSAEAFAPDVR .NGVLVTDLRFMDNSGVELRNIGDLKINIPSVITGHGDVPMHMAKAGAVDFIEKFF	106
UhpA	MITVALIDHILVRSQFAGLLGLEPDLQVVAEFGSGREALAGLP .GRGVQVCICDISMPDISGLELLSGLPKG .MATIMLSVHSDSPALVEQALNAGARGFLSKRC	103
BvgC	672 . .RVLVVDHDKPNMLMLLRQDLYLQGVIA . .ADSGEAALALNREH .AFDVVITDCNMPGISGYELARRIRAAEAAPGYGRTRCILFGFTASAQNDQAACRAAGMDDCLFKPI	780
SpoOF	MMNEKLLIVDQYGRIRILLNEVFNKEGYQTF .QAANGLQALDIVT .KERPDVLLDMKIPGMDGIEILKRMKVIDENIRVIIMTAYGELDMIQESKELGALTHFAKPF	106
CheY	MADKELKFLVVDVSTMRIRVRNLLKELGFNNV . .EEAEDVDALNKLQAGGYFVSDNMPNMDGLELLKTRADGAMSALPVLMVTAEAKENI IAAQAAGASGYVVKPF	111
OmpR	HQENTYKMLVVDMDMLRALLERYLTDQGFQ .VRSVANAEQMDRL .TRESFH .LNVLDLHLPGEDGLSICRRLRSQSNPMPIIMVTAKGEEVDRIVGLGADDDYIPKFF	106
CheB	MSKIRVLSVDDSLMRQIMTEIINSHSDMEMVATAPDPLVARDL IKKFN .PDVLTLDVEMRMDGLDFLEKLMRLR .PMPVVMVSSLTGKGVSEV .TLRPLELGAIDFVTKPQ	109
PhoB	MARRILVDEDEAPIRENVCFVLEQNGFQFVEA .EDYSAVNQLN . .EPNPDILLDMHLPGGSGIQFINKLRE . .SMTRDIPVVMLTARGEEDVRGLETGADDDYITKFF	107
Dye	MQTPHLLIVDELVTRNFKLSIFEAEQYDVFEATD .GAEMHQIL . .SEYDINLVIMDINLPGKNGLLAREL .REQANVALMFLTGRDNEVDKILGLEIGADDDYITKFF	105
SpoOA	MEKIKVCVADNRELVSLLSEYIEGQEDMEVIGVAQYQCELSL .KEKDPDVLVLDIIMPHLDGLAVLERL .RESDLKQKPNVIMLTAFGQEDVTKKAVDLGASVYLLKPF	110
NtrC	MQRGIAIIVDDSSIRRWLERALTAGLSCTTFES . .GNEVLDALTT .KTPDVLVLDIRMPGMDGLALLKQI .KQRHPMLPVIIMTAHSDLDAVASAYQGFADYLLKPF	106

BvgA	NLHEVINAAKAVMAGTYFFPSTLSEMRMGDNAKSDSTLISVLSNRELIVLQ .LLA .QGMNKDIADSMFLSNKTYSTYKTRLLQKLNATSLEVELDLAKRNLA*	209
FixJ	EDTVIIEATERASEHLVAEAEDVDANDIRARLQTLSEERQVLSAVVA .GLPNKSIAYLDISPRTEVHRANVAMKAKSLPHLVRMALAGFGPSP	204
UhpA	SPDELIAAVHTVATGGCYLTPDIAIKLASGRDPLTKREROVAEK .LA .OGMAVKEIAEELGSPKTVHVRAMLMEKLGVSNDVELARRMFDG**	196

BvgC	GVDALRQRLN	790
SpoOF	DIDEIRDAVK	116
CheY	TAATLEEKLN	121
OmpR	NPRELLARIR	116
CheB	LGIREGMLAY	119
PhoB	SPKELVARIK	117
Dye	NPRELTRAR	115
SpoOA	DMENLVGHIR	120
NtrC	DIDEAVALVD	116

FIG. 3. Amino acid sequence comparisons between BvgA, BvgC, and members of bacterial sensory transduction systems that contain receiver domains. Similarities and identities are as described in the legend to Fig. 3. References for members of two-component systems that are not described in the text are as follows: SpoOF (36), SpoOA (37), CheY and CheB (38), and Dye (39).

derepressed by modulation signals has recently been identified (43); however, their functions are unknown.

Our analysis of the nucleotide sequence has identified three ORFs, designated *bvgA*, *bvgB*, and *bvgC*, which reside within the *bvg* region. In *B. pertussis*, transposon insertions in *bvgA* or *bvgC* disrupt *bvg* function (10), and spontaneous phase variation in the Tohamall strain results from the insertion (and deletion) of a cytosine residue at position 4129 in *bvgC* (amino acid 800; ref. 14). In this report we show that insertion of a promoterless CAT gene within *bvgC* eliminates *bvg* activity and that nonpolar linker insertion mutations in *bvgA*, *bvgB*, and *bvgC* result in a Bvg⁻ phenotype in *B. pertussis*. The *bvgC*-CAT fusions also confirm that transcription occurs from left to right as shown in Fig. 1.

Sequences within the *EcoRI*-*Xho* I fragment (Fig. 1) are sufficient for trans activation and modulation of an *fhaB-lacZ* transcriptional fusion in *E. coli* (J.F.M., C.R., and S.F., unpublished data). This suggests that *bvg*-specific components involved in signal transduction are present within these boundaries. Sequences downstream from *bvgC* do not show any obvious ORFs, and linker insertions within this region had no effect on *bvg* activity. Knapp and Mekalanos (43) have recently described the isolation of mutations in the *bvg* region that constitutively express virulence factors in the presence of modulating conditions. The nature and location of these mutations have not yet been reported.

Amino acid sequence similarities suggest that BvgA and BvgC belong to a large family of bacterial regulatory proteins that use conserved transmitter and receiver domains to transduce environmental signals. A growing body of evidence indicates that the mechanism of signal transduction involves autophosphorylation of transmitter proteins, followed by phosphorylation of regulator proteins that contain receiver domains (44-47). In relation to this "two-component motif," the *bvg* locus is interesting in several respects. Three components, one of which (BvgB) does not contain a transmitter or receiver domain, are needed for regulation. Three linked loci (*uhpABC*) are also responsible

for exogenous induction of sugar-phosphate transport in *E. coli* (48), and the products of these genes share functional similarities with the proposed Bvg proteins. Another unusual aspect of the *B. pertussis* system is the presence of a transmitter and a receiver domain in the same polypeptide sequence (see below).

A model describing the proposed cellular locations, functions, and interactions of the *bvg* products is shown in Fig. 4. It is suggested that BvgB and the N-terminal domain of BvgC (residues 1-240) are localized in the periplasm. Residues 60-276 of BvgB share striking similarities with the periplasmic region of BvgC, indicating that these proteins may interact with a common factor, or with each other. According

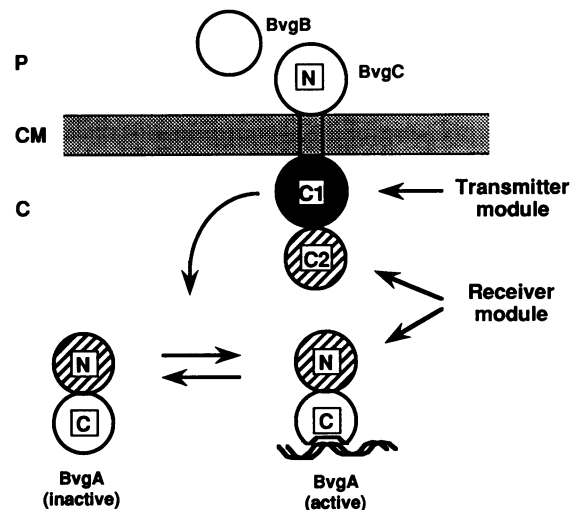


FIG. 4. Model for the regulation of virulence factors in *B. pertussis*. Details are presented in Discussion. P, CM, and C designate periplasm, cytoplasmic membrane, and cytoplasm, respectively.

to our model, the transmembrane signal in BvgC (residues 240–261) anchors this protein in the cytoplasmic membrane, and the C-terminal transmitter and receiver modules are located in the cytoplasm along with BvgA.

Several properties of the *bvgA* sequence suggest that it encodes a transcriptional activator. The presence of a receiver domain and extensive homology with FixJ and UhpA support this contention. We have recently observed that the *phaB* promoter is activated in *E. coli* by hyperexpression of BvgA alone (C.R., J.F.M., and S.F., unpublished data), although both BvgB and BvgC are normally required for BvgA activity in both *E. coli* and *B. pertussis*.

In the absence of modulating signals, we propose that BvgC activates BvgA. It is likely that this activation results from phosphorylation of BvgA by a mechanism involving the transmitter domain of BvgC. BvgB could exert its effect by directly interacting with BvgC, or by inactivating an unidentified inhibitor of BvgC. Although the modulating effect of temperature could act at any level, the inhibitory effects of signals like MgSO₄ or nicotinic acid may occur in the periplasm through an interaction with BvgB and/or BvgC.

The role of the additional receiver module at the C terminus of BvgC is unknown; however, it may be involved in modulating the proposed kinase activity of the transmitter domain. The frameshift mutation described above, which results in a Bvg⁻ phenotype, lies 10 residues downstream from this receiver (14). Deletion of sequences that encode the final 26 amino acids of BvgC also disrupts *bvg* activity (J.F.M., C.R., and S.F., unpublished data). A role for the C terminus of BvgC in DNA binding can also not be ruled out.

We thank D. Relman for advice, critical reading of the manuscript, and transatlantic communications and S. Ricci for oligonucleotide synthesis. This work was supported by Damon Runyon-Walter Winchell Postdoctoral Fellowship DRG-911 (to J.F.M.), by a National Institutes of Health Public Health Service grant and a National Institutes of Health National Research Service Award (to S.S.), and by National Institutes of Health Grant AI29345 (to S.F.). B.A., R.G., and R.R. were supported by a grant from ENI (Ente Nazionale Idrocarburi).

1. Miller, J. M., Mekalanos, J. J. & Falkow, S. (1989) *Science* **243**, 916–922.
2. Weiss, A. A. & Hewlett, E. L. (1986) *Annu. Rev. Microbiol.* **40**, 661–686.
3. Urisu, A., Cowell, J. L. & Manclark, C. R. (1986) *Infect. Immun.* **52**, 695–701.
4. Livey, I., Duggleby, C. J. & Robinson, A. (1987) *Mol. Microbiol.* **1**, 203–209.
5. Goldman, W. E., Klapper, D. G. & Baseman, J. B. (1982) *Infect. Immun.* **36**, 782–794.
6. Wardlaw, A. C. & Parton, R. (1983) *Pharmacol. Ther.* **19**, 1–53.
7. Weiss, A. A., Hewlett, E. L., Myers, G. A. & Falkow, S. (1984) *J. Infect. Dis.* **150**, 219–222.
8. Glaser, P., Landant, D., Sezer, O., Pichot, F., Ullmann, A. & Danchin, A. (1988) *Mol. Microbiol.* **2**, 19–30.
9. Nicosia, A. & Rappuoli, R. (1987) *J. Bacteriol.* **169**, 2843–2846.
10. Weiss, A. A., Hewlett, E. K., Myers, G. A. & Falkow, S. (1983) *Infect. Immun.* **42**, 33–41.
11. Weiss, A. A. & Falkow, S. (1984) *Infect. Immun.* **43**, 263–269.
12. Stibitz, S., Weiss, A. A. & Falkow, S. (1988) *J. Bacteriol.* **170**, 2904–2913.
13. Gross, R. & Rappuoli, R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 3913–3917.
14. Stibitz, S., Aaronson, W., Monack, D. & Falkow, S. (1989) *Nature (London)* **338**, 266–269.
15. Lacey, B. W. (1960) *J. Hyg.* **58**, 57–93.
16. Idigbe, O. R., Parton, R. & Wardlaw, A. C. (1981) *J. Med. Microbiol.* **14**, 409–418.
17. Ronson, C. W., Nixon, B. T. & Ausubel, F. M. (1987) *Cell* **49**, 579–581.
18. Simon, R., Priefer, U. & Puhler, A. (1983) *Biotechnology* **1**, 784–791.
19. Close, T. J. & Rodriguez, R. L. (1982) *Gene* **20**, 305–316.
20. Stibitz, S., Black, W. & Falkow, S. (1986) *Gene* **50**, 133–140.
21. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
22. Devereux, J., Haerberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
23. Relman, D. A., Domenighini, M., Tuomanen, E., Rappuoli, R. & Falkow, S. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 2637–2641.
24. Krikos, A., Mutoh, N., Boyd, A. & Simon, M. I. (1983) *Cell* **33**, 615–622.
25. Engelman, D. M., Steitz, T. A. & Goldman, A. (1986) *Annu. Rev. Biophys. Chem.* **15**, 321–353.
26. Leroux, B., Yanofsky, F. M., Winans, C. S., Ward, E. J., Ziegler, F. S. & Nester, E. W. (1987) *EMBO J.* **6**, 849–856.
27. Kuhn, A., Wickner, W. & Kreil, G. (1986) *Nature (London)* **322**, 335–339.
28. Harley, C. B. & Reynolds, R. P. (1987) *Nucleic Acids Res.* **15**, 2343–2361.
29. Brendel, V., Hamm, G. H. & Trifonov, E. N. (1986) *J. Biomol. Struct. Dyn.* **3**, 705–723.
30. Kofoid, E. C. & Parkinson, J. S. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4981–4985.
31. Mizuno, T., Wurtzel, T. & Inouye, M. (1982) *J. Biol. Chem.* **257**, 13692–13698.
32. David, M., Daveran, M., Batut, J., Dedieu, A., Domergue, O., Ghai, J., Hertig, C., Boistard, P. & Kahn, D. (1988) *Cell* **54**, 671–683.
33. MacFarlane, A. A. & Merrik, M. (1985) *Nucleic Acids Res.* **13**, 7591–7606.
34. Makino, K., Shinagawa, H. H., Amemura, M. & Nakata, A. (1986) *J. Mol. Biol.* **192**, 549–556.
35. Amemura, M., Makino, K., Shinagawa, H. & Nakata, A. (1986) *J. Bacteriol.* **168**, 294–302.
36. Trach, K. A., Chapman, J. W., Piggot, P. J. & Hoch, J. A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7260–7264.
37. Ferrari, F. A., Trach, K., LeCoq, D., Spence, J., Ferrari, E. & Hoch, J. A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2647–2651.
38. Stock, A., Koshland, D. E. & Stock, J. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7989–7993.
39. Drury, L. S. & Buxton, R. S. (1985) *J. Biol. Chem.* **260**, 4236–4242.
40. Friedrich, M. J. & Kadner, R. J. (1987) *J. Bacteriol.* **169**, 3556–3563.
41. Lipman, D. J. & Pearson, W. R. (1985) *Science* **227**, 1435–1441.
42. Brennan, R. G. & Matthews, B. W. (1989) *J. Biol. Chem.* **264**, 1903–1906.
43. Knapp, S. & Mekalanos, J. (1988) *J. Bacteriol.* **170**, 5059–5066.
44. Ninfa, A. J., Ninfa, E. G., Lupas, A. N., Stock, A., Magasanik, B. & Stock, J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5492–5496.
45. Hess, J. F., Oosawa, K., Kaplan, N. & Simon, M. I. (1988) *Cell* **53**, 79–87.
46. Igo, M. M. & Silhavy, T. J. (1988) *J. Bacteriol.* **170**, 5971–5973.
47. Keener, J. & Kustu, S. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4976–4980.
48. Weston, L. A. & Kadner, R. J. (1988) *J. Bacteriol.* **170**, 3375–3383.