# The phasevarion: A genetic system controlling coordinated, random switching of expression of multiple genes

Yogitha N. Srikhanta*, Tina L. Maguire†, Katryn J. Stacey†, Sean M. Grimmond†, and Michael P. Jennings*‡

*School of Molecular and Microbial Science and †Institute for Molecular Bioscience, University of Queensland, St. Lucia, Brisbane, Queensland 4072, Australia

Communicated by Sankar Adhya, National Institutes of Health, Bethesda, MD, February 14, 2005 (received for review December 3, 2004)

**Several host-adapted bacterial pathogens contain methyltransferases associated with type III restriction-modification (R-M) systems that are subject to reversible, high-frequency on/off switching of expression (phase variation). To investigate the role of phase-variable expression of R-M systems, we made a mutant strain lacking the methyltransferase (*mod*) associated with a type III R-M system of *Haemophilus influenzae* and analyzed its phenotype. By microarray analysis, we identified a number of genes that were either up- or down-regulated in the *mod* mutant strain. This system reports the coordinated random switching of a set of genes in a bacterial pathogen and may represent a widely used mechanism.**

DNA methyltransferase | *Haemophilus influenzae* | phase variation

Restriction-modification (R-M) systems are frequently found in bacteria and generally are thought to confer protection to the bacterial host against infections by foreign DNA (1). Type III R-M systems are composed of two subunits: Restriction (Res) and Modification (Mod) enzymes, which are encoded by the *res* and *mod* genes, respectively. Res catalyzes the double-stranded cleavage of unmethylated foreign DNA at a specific recognition sequence. Mod catalyzes the methylation of DNA at the same sequence, protecting host DNA from cleavage (2, 3). DNA recognition sites of type III R-M systems are asymmetrical, and methylation is only on one strand (4).

Several studies have reported methyltransferases associated with type III R-M systems in pathogenic bacteria [*Pastuerella haemolytica* (5), *Haemophilus influenzae* (6), *Neisseria meningitidis*, *Neisseria gonorrhoae* (7), *Helicobacter pylori* (8), and *Morexella catarrahalis* (9)] that have sequence features that are consistent with phase-variable expression. Phase variation is the reversible, high-frequency on/off switching of expression, and it is usually mediated by mutations in simple DNA repeats located either within the ORF of genes encoding variant proteins or in their promoter region (10, 11).

Many genes encoding virulence factors in bacterial pathogens display phase-variable expression, such as pili (12), iron-binding proteins (13, 14), lipopolysaccharide biosynthesis genes (15, 16), and outer-membrane proteins (17). Phase variation results in genetically and phenotypically diverse populations, which is important in pathogenesis because it provides a strategy for rapid adaptation to changes within the host environment and immune response (18). The existence of such phase-variable methyltransferases raises the possibility of a role for phase variable type III R-M systems in pathogenesis, such as differential immune stimulation and gene regulation (9).

To examine the biological role of phase variation of methyltransferases from type III R-M systems, we chose the *mod* gene of *H. influenzae* strain Rd (HI1058/HI1056) as a model system. *H. influenzae* is an obligate, host-adapted bacterial pathogen that colonizes the upper respiratory tract. It is the second leading cause of community-acquired pneumonia and accounts for several thousand deaths annually worldwide, especially in children (19). Of the potentially phase-variable type III R-M systems that have been described (9), the *mod* gene of *H. influenzae* is the only example in which phase-variable expression, mediated by tetranucleotide repeats, has been confirmed experimentally (6). Also, only one type III R-M system is present in strain Rd, and mircoarrays for strain Rd are commercially available. The *mod* gene of *H. influenzae* contains tetranucleotide repeats (5′-AGTC-3′) within its ORF, and strains have been observed with a repeat-number range of 2–41 (6). The rate of phase variation depends on the number of repeats that are present in the gene; a greater number of repeats increases the frequency of phase variation. High phase variation rates may be significant in infections by *H. influenzae* (6).

## Experimental Procedures

**Bacterial Strains and Growth Conditions.** *H. influenzae* strains were grown at 37°C in brain–heart infusion (BHI) supplemented with either hemin (10 mg/liter) and NAD (2 mg/liter) in liquid medium or Levinthal supplement in solid medium with 5% $CO_2$. *Escherichia coli* strains DH5$\alpha$ and JM109 (Promega) were used to propagate cloned plasmids and were grown at 37°C in LB broth supplemented with either ampicillin (100 $\mu$g/ml) or kanamycin (50 $\mu$g/ml).

**DNA Manipulation and Analysis.** All enzymes were sourced from New England Biolabs. PCR was performed with oligonucleotides purchased from Proligo (Boulder, CO). Sequencing was performed on PCR products by using QiaQuick gel extraction kit (Qiagen, Valencia, CA) and Big-Dye (PerkinElmer) sequencing kits. Sequencing reactions were submitted to the Australian Genome Research Facility (University of Queensland) and analyzed by using an ABI 3700 automatic sequencer (Applied Biosystems International). Data were analyzed by using SEQED (version 1.0.3). The sequences of primers used in this study are given in Table 2, which is published as supporting information on the PNAS web site.

**Construction of a Knockout Mutant of the *mod* Gene and Insertion into *H. influenzae* Strain Rd.** The *mod* ORF was amplified by PCR with primers Him1 and Him2. The PCR product was cloned into the pGEM-Teasy (Promega) vector. This construct was digested with *Mfe*I and blunted by using Klenow polymerase (New England Biolabs). The Tn*903 kan* resistance gene from the pUC4K vector (Pfizer) was excised by using *Hinc*II and inserted into the blunt *Mfe*I site. The resulting plasmid, pGEM*mod*::*kan* was linearized by digestion with *Eco*RI and used to transform competent *H. influenzae* Rd (20). Rd*mod*::*kan* transformants were selected on BHI plates containing Levinthal supplement

Abbreviations: R-M, restriction modification; Res, Restriction; Mod, Modification; BHI, brain–heart infusion.

‡To whom correspondence should be addressed. E-mail: jennings@uq.edu.au.

and 10 μg/ml kanamycin (Fig. 3 and *Supporting Experimental Procedures*, which are published as supporting information on the PNAS web site). Rd*mod*::*kan* transformants were confirmed by PCR and sequence analysis using primers Him1 and Him2 and kanamycin-specific primers kanfor and kanrev (see Table 2). RNA midi-preps of both the WT (Rd) and mutant (Rd*mod*::*kan*) were made by using the RNeasy Midiprep kit (Qiagen). The WT *H. influenzae* Rd colonies that were used to make RNA for microarray analysis were sequenced to check that *mod* gene expression was from the Distal ATG.

**Construction of Translation Fusion Between the *opa* Gene and *lacZ* Gene and Insertion into *H. influenzae* Strain Rd.** An *opa*–*lacZ* fusion was constructed in *H. influenzae* Rd. The gene fusion was constructed initially in *E. coli*, with subsequent transformation into the *H. influenzae* chromosome. In the fusion construct, the codons for LacZ are in the same translational frame as Opa, resulting in an in-frame Opa–LacZ fusion protein. A 2.0-kb DNA fragment of the Opa ORF was amplified by PCR using the primer pair 1457F and 1457R and strain Rd as the template. The reaction was performed in 50 μl by using 1× *Taq* buffer, 1.5 mM $MgCl_2$, and 1 unit of *Taq*DNA polymerase (Promega) with the following cycling conditions: 30 cycles of 94°C for 30 s, 56°C for 30 s, and 72°C for 1 min, and one cycle of 72°C for 7 min. The fragment was then cloned into vector pGEM-Teasy (Promega). A 4-kb fragment of a promoterless *lacZ*::*kan* fragment was amplified by PCR using the primer pair LacZ*Stu*I1 and Kan*Stu*I. The plasmid pBluescript*lacZ*::*kan* (M. Dieckelmann, personal communication) was used as template. The PCR was performed in 50 μl by using 1× *Taq* buffer, 1.5 mM $MgCl_2$, and 1 unit of *Taq*DNA polymerase (Promega) with the following cycling conditions: 30 cycles of 94°C for 30 s, 50°C for 30 s, and 72°C for 4 min, and one cycle of 72°C for 7 min. After digestion with *Stu*I, the 4.0-kb *lacZ*::*kan* fragment was then ligated into the *Stu*I site of the o*pa* construct. The ligation mixture was transformed into *E. coli* JM109 and transformants were selected on LB agar plates supplemented with kanamycin (50 μg/ml; Sigma). The orientation and sequence of the insert were checked and found to be correct. The resulting construct was named pGEM*opa*::*lacZ*::*kan.* This plasmid was linearized with *Sac*II and used to transform competent *H. influenzae* (see *Supporting Experimental Procedures* and Fig. 4, which are published as supporting information on the PNAS web site). The Rd*opa*::*lacZ*::*kan* transformants were streaked on BHI plates containing Levinthal supplement and X-gal (5-bromo-4-chloro-3-indolyl-D-galactopyranoside; 40 μg/ml). Blue and white color transformants were picked and the *mod* repeat tract sequenced by using the primers Him1 and Him3 to confirm that the *mod* tract was out of frame (Off) or in frame (Distal).

**β-Galactosidase Assay.** The appropriate strains were grown on BHI plates at 37°C with 5% $CO_2$, and plate culture cells were taken after 16–18 h of growth. The cells were scraped into PBS and lysed by repeated freeze–thaw cycles. The cell debris was removed by centrifugation at 15,000 × *g* for 10 min. The amount of protein was calculated by using the BCA protein assay reagent kit (Pierce), and extracts were adjusted so that equivalent amounts of protein were used in the assay. The amount of β-galactosidase in the cell extracts was measured in Miller units, in triplicate, as described (21). Miller units were calculated as follows: Units$(1{,}000 \times A_{420})/(t \times v \times C)$, where $t$ is the time of assay (in min), $v$ is the volume of cell extract used in the assay (in μl), and $C$ is the total protein concentration (in μg/ml).

**Northern Blot Analysis.** RNA was separated by electrophoresis on 1% denaturing formaldehyde/agarose gels. Northern blotting onto GeneScreen membrane (DuPont) was accomplished by means of capillary action using the wick-transfer method in 10×

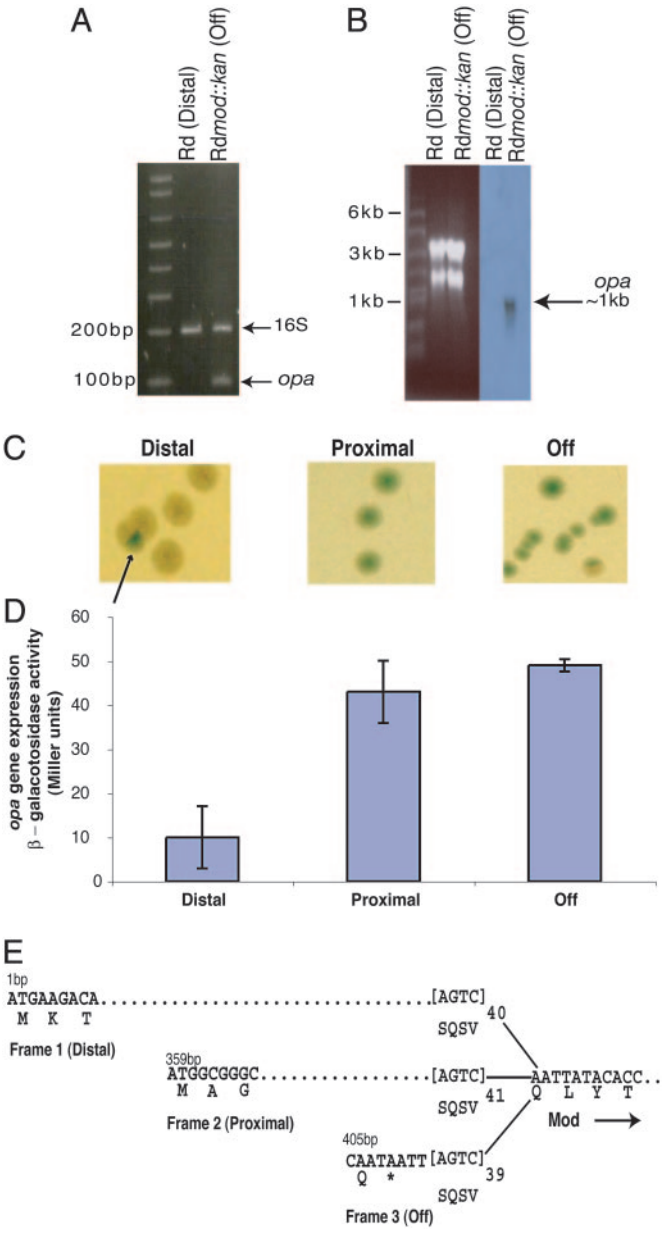


**Fig. 1.** Effect of *mod* phase variation on expression of the *opa* gene (HI1457). (*A*) RT-PCR shows expression of the *opa* gene in Rd*mod*::kan but not WT Rd (Distal). 16S RNA was amplified as a control. (*B*) Northern blot analysis showing a higher level of *opa* transcriptional expression in Rd*mod*::kan. A band of ≈1-Kb corresponds to the expected size of a putative operon of two genes HI1457 (*opa*) and HI1456. (*C*) Phenotypic validation that *opa*::*lacZ* gene expression depends on phase variation of the *mod* gene. Rd*opa*::*lacZ* colonies with the *mod* 5′-AGTC-3′ repeat tract in frame with the Distal ATG (resulting in active Mod) were white (40 or 37 repeats), indicating low *opa*::*lacZ* expression. Colonies that phase varied to a blue phenotype (arrow indicates an example) were observed and picked, and the *mod* repeat region was sequenced to determine whether change in *opa*::*lacZ* expression correlated with *mod* phase variation. Blue colonies were found to have switched from Distal (40 repeats) to be in frame with either the Proximal ATG (41 repeats) or Off (39 repeats). Colonies that switched back from blue (39 repeats) to white were found to be in frame with the Distal ATG (40 or 37 repeats). (*D*) β-galactosidase assays showing quantitative differences in the level of *opa*::*lacZ* gene expression resulting from *mod* repeat tract changes (Distal, Proximal, or Off). A 5-fold difference in expression was observed between Distal and Proximal/Off, consistent with the array and quantitative PCR data. (*E*) Schematic diagram showing that translation of the *mod* gene is initiated from one of three frames (Distal, Proximal, or Off), depending on the number of 5′-AGTC-3′ repeats.

**Table 1. Differentially expressed genes in WT (Distal) versus the *mod*::*kan* mutant (Off)**

| Accession no. | Description | Average ratio | *P* value | Validation |
|---|---|---|---|---|
| | Reduced expression in *mod* mutant | | | |
| HI0661 | Hemoglobin–haptoglobin binding protein | 0.197 | 0.145452 | 0.191 ± 0.026*† |
| HI0853 | Heme-binding protein A precursor | 0.396 | 7.46E-06 | * |
| HI1054 | Type III R-M system | 0.410 | 0.211941 | * |
| HI1055 | Type III R-M system | 0.330 | 0.285167 | 0.34 ± 0.048*† |
| HI1078 | Glutamate/aspartate transport ATP-binding protein | 0.355 | 0.044136 | * |
| HI1079 | Cysteine transport system permease protein | 0.327 | 0.048049 | * |
| HI1080 | Cysteine-binding protein | 0.381 | 3.08E-05 | 0.42 ± 0.202*† |
| HI1154 | Proton/sodium glutamate symport protein | 0.347 | 4.38E-07 | * |
| HI1227 | Uracil permease | 0.505 | 0.129607 | * |
| | Increased expression in *mod* mutant | | | |
| HI0104 | Heat-shock protein htpG | 2.22 | 0.073379 | 2.06 ± 0.65*† |
| HI0542 | 10-kDa chaperonin GROES | 2.01 | 0.036 | 1.59 ± 0.65† |
| HI0543 | 60-kDa chaperonin GROEL | 2.02 | 0.015129 | 2.89 ± 0.47*† |
| HI1237 | Chaperone protein dnaK | 2.08 | 0.039308 | 2.00 ± 0.57*† |
| *HI1238* | *Chaperone protein dnaJ* | 1.60 | *0.0263* | *1.73 ± 0.66*† |
| HI1456 | Hypothetical protein | 1.99 | 0.406473 | 2.66 ± 0.56*† |
| HI1457 | OPA protein | 3.86 | 0.014312 | 4.91 ± 0.86*†‡§ |

The genes listed are either down- or up-regulated in the Rd *mod*::*kan* mutant strain. The identity of each gene is indicated by the accession no. allocation in the annotation of the *H. influenzae* strain Rd genome. The average ratio presented is the mean of mutant/WT from six replicate spots on two independent microarrays, incorporating a dye swap. Only genes with an expression value of >2-fold were included in this study, except for HI1238, which is shown in italics (see text). All primary data, related metadata, and a detailed summary of the protocols used in this project are available in the BASE database (see text).

*Semiquantitative real-time PCR (e.g., see Fig. 1*A*).

†Quantitative real-time PCR.

‡Northern blot analysis (see Fig. 1*B*).

§*lacZ* reporter fusion (see Fig. 1 *C* and *D*).

SSC (1× SSC = 0.15 M sodium chloride/0.015 M sodium citrate, pH 7). Northern blots were hybridized at 65°C with DIG-labeled probe generated by using the DIG High Prime kit (Roche) and primers HI1457PF and HI1457PR. CDP-Star chemiluminescence was used for detection (Roche). The size marker was a 0.2- to 6-kb RNA ladder (Progen, Heidelberg).

**Microarray Analysis.** All microarray analysis was performed on *H. influenzae* Rd genome arrays obtained from Integrated Genomics (Chicago). Each microarray was composed of ≈1,880 elements corresponding to each predicted ORF in *H. influenzae* Rd and were spotted in triplicate.

We prepared ≈100 μg of total RNA from each sample by using the RNeasy Midiprep kit. The integrity and concentration of RNA was then determined by means of microfluidic analysis on a bioanalyzer (Agilent), and 40 μg of each total RNA sample was labeled by using random hexamers and direct incorporation of fluorescently labeled nucleotides, as described in ref. 22. The hybridizations were performed in duplicate, and they incorporated a dye swap to account for dye bias.

After 16 h of hybridization, the arrays were washed and scanned on an Agilent G2565BA microarray scanner at a 5-μm resolution. The resulting images of the hybridizations were analyzed by using IMAGENE 5.5 (BioDiscovery, El Segundo, CA) and the mean foreground, mean background, and spot/signal quality determined. These data were then exported into GENESPRING 6.1 (SiliconGenetics, Redwood City, CA), and the mean differential expression was observed. A further filtering of elements with a mean foreground intensity of ≥200 fluorescence units was included to improve the accuracy of predicting differential expression. Last, elements that were found to have a >2-fold mean change and $P < 0.4$ (one-sample Student's *t* test calculated to test whether the mean normalized expression level for the gene is statistically different from 1.0) in expression between WT and mutant samples were identified and used for further analysis.

All primary data, related metadata, and a detailed summary of the protocols used in this project are available from an instance of the MIAME (Minimum Information About a Microarray Experiment) supportive array database BASE 1.2.16 (BioArray Software Environment; available at http://kidney.scgap.org/base; username, reviewer; password, yogi123).

**Heat-Shock Killing Assay.** *H. influenzae* strains were grown for 16 h at 37°C with 5% $CO_2$, followed by inoculation of 2 ml of BHI broth (Oxoid, Basingstoke, U.K.) supplemented with NAD and hemin. On the next day, 1 ml of BHI broth was inoculated with either the Rd*mod*::*kan* mutant or WT Rd; supplemented with NAD and Hemin; and grown, with aeration, to logarithmic phase at 37°C. Heat-shock treatment was performed at time intervals of 0–90 min by incubating $3 \times 10^8$ cells per $ml^{-1}$ under aerobic conditions at 46°C. A sample was taken at each point, serial dilutions were carried out in sterile phosphate buffer, and 10 μl of each dilution was spotted onto BHI plates in triplicate. For WT Rd, 12 individual colonies from the 0- and 70-min time points were isolated, and the *mod* repeat region was amplified by PCR using primers Him1 and Him3 and sequenced to determine changes in repeat numbers. WT Rd colonies from the 70-min time point ("Rd survivors") were used for a subsequent assay to compare with WT Rd and Rd*mod*::*kan*.

## Results and Discussion

**Differences in *mod* Expression from Alternate Initiation Codons.** In strain Rd, the *mod* gene has two potential start codons that code for proteins of either 72- (Proximal) or 86-kDa (Distal), depending on the number of repeats that are present (6) (Fig. 1*E*). Because this study is focused on *mod* function and the characterization of *mod* expression is crucial, expression from both ATGs was examined by using a *mod*::*lacZ* reporter fusion (Fig.

1 and *Supporting Experimental Procedures* and Fig. 5, which are published as supporting information on the PNAS web site). The Distal start gave maximum expression, the Proximal start showed low expression (2.5% of maximum), and the third reading frame (which has no candidate ATG and a stop codon immediately before the 5′-AGTC-3′ repeats) was "Off."

**Analysis of Differentially Expressed Genes in WT (Distal) Versus the *mod*::*kan* Mutant (Off).** To determine the biological role of phase-variable expression of the *H. influenzae mod* gene, a *mod* knockout mutant was constructed by interrupting the *mod* gene with a kanamycin antibiotic resistance marker and transferring this inactivated allele to the chromosome of strain Rd to create strain Rd*mod*::*kan* (Fig. 3). Comparison of the phenotype of this mutant strain with strain Rd, in which the *mod* gene was maximally expressed (Distal start, 40 repeats; see Fig. 1*E*), formed the basis of subsequent studies.

DNA methylation has been shown to be involved in regulating bacterial gene expression and virulence by changing the affinity of regulatory proteins for DNA. Alternatively, DNA target sites can be protected from methylation by the binding of regulatory proteins to nonmethylated sites (23–25). To examine the possibility that phase-variable methyltransferase activity may randomly switch the expression of virulence factors, gene expression of WT Rd (Distal) and Rd*mod*::*kan* was compared by microarray analysis on *H. influenzae* strain Rd arrays (Integrated Genomics). Seven genes were up-regulated in Rd*mod*::*kan* relative to WT, and nine genes were down-regulated (Table 1), confirming that expression of the *mod* gene has a direct influence on gene expression. Two genes are surface-exposed proteins; HI0661 (*hgpB*) encodes a TonB-dependent outer-membrane protein mediating the binding and utilization of hemoglobin–haptoglobin (13, 14), and HI1457 (*opa*) encodes a homologue of NspA from *N. meningitidis* an outer-membrane protein that is a potential vaccine candidate (26). These proteins are typical examples of those expected to display phase-variable expression because of functional and/or immune selection. The selective pressure for phase variation is further supported by the observation that the *hgpB* gene also contains a 5′-CCAA-3′ repeat tract within its coding sequence, mediating phase variation at the translational level (13, 14). These data suggested that this set of genes was subject to phase-variable expression dependent on Mod activity.

In addition to the OMPs, there were two other groups of genes given in Table 1 that are subject to *mod*-dependent phase-variable expression. The first group is a series of genes encoding various transport functions, three with homology to various amino acid transport proteins. HI1080, HI1079, and HI1078 are potentially part of an operon, one with homology to a uracil permease (HI1227) and one that encodes *hbpA* (HI0853) involved in heme binding and transport (27). It is possible the coordinated phase-variable expression of this group of transporters represents a switch between two distinct microenvironments within the host (28).

The second group of genes with related function that are subject to *mod*-dependent phase-variable expression are five heat-shock genes: HI0104 (*htpG*), HI0542/3 (*dnaKJ*), and HI1237/8 (*groEL* and *groES*). These genes encode highly conserved proteins in bacteria, HtpG, DnaKJ, and GroELS, which have important roles in cell physiology as molecular chaperones. Transient induction of heat-shock proteins is a vital protective mechanism to cope with various sources of physiological and environmental stress at the cellular level. Evidence suggests that one of these chaperones, DnaK, may have a direct role in the pathogenesis of *H. influenzae* in cell adhesion (29, 30).

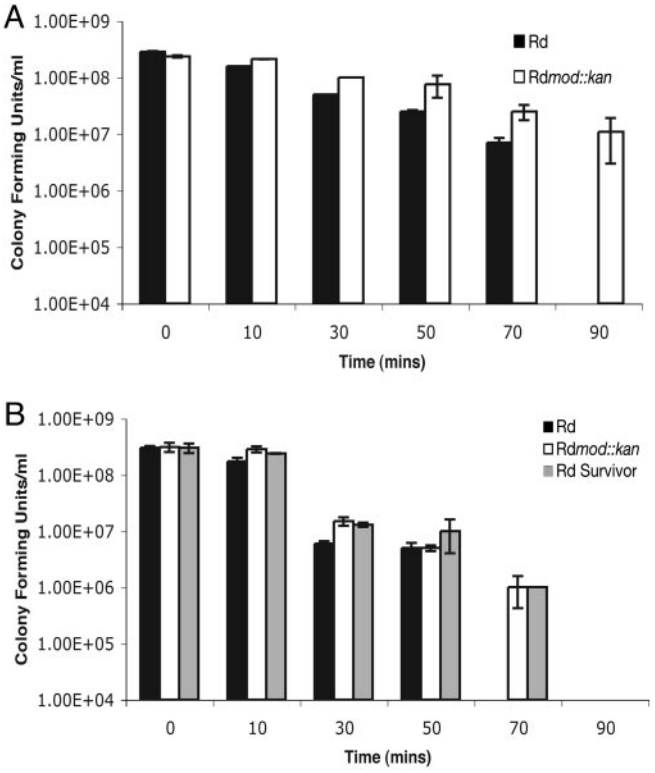


**Fig. 2.** Comparison of WT Rd (Distal) and Rd*mod*::*kan* in a heat-shock killing assay; selection of *mod* phase variants with increased heat-shock resistance. Cells were incubated at 46°C, and samples were taken at 10-min intervals, diluted, and plated onto BHI plates for determination of viable colony-forming units. (*A*) Comparison of WT Rd (Distal) and Rd*mod*::*kan* revealed that Rd*mod*::*kan*, in which the heat-shock proteins are up-regulated, is markedly more resistant to heat shock than WT Rd (Distal). A Student's *t* test showed a significant difference at time points of 10–90 min ($P \leq 0.00506$). Analysis of WT Rd (Distal) from the last time point containing viable cells (70-min time point), hence Rd survivors, showed that 33% were phase variants in which the *mod* gene had switched from Distal to Off (i.e., from 40 to 39 repeats). (*B*) Comparison of WT Rd (Distal), Rd*mod*::*kan*, and the Rd survivors pool (derived from the 70-min time point of the previous assay) (see *A*) showed a significant difference between Rd (Distal) and Rd survivors at time points 10–70 min ($P \leq 0.04712$), but not between Rd*mod*::*kan* and Rd survivors. Analysis of the Rd survivor group from the last time point containing viable cells (50-min time-point) revealed that 67% of the survivor colonies had the *mod* gene switched to Off (starting population was 33% Off).

***mod* Gene Phase Variation Linked to Phase Variation of the *opa* Gene.** To confirm that reversible, high-frequency on/off switching of *mod* expression resulted in phase-variable expression of this set of genes, a LacZ reporter fusion was made in HI1457 (*opa*) and transferred to the chromosome of strain Rd (Fig. 4). O*pa* gene expression was directly linked to phase variation of the *mod* gene. Colonies with a white phenotype (low *opa*::*lacZ* expression; Fig. 1 *C* and *D*) had a *mod* repeat region in frame with Distal start (40 repeats) (Fig. 1*E*), resulting in maximal Mod expression. Colonies that switched from white to a blue phenotype (high *opa*::*lacZ* expression; Fig. 1 *C* and *D*) were all *mod* phase variants that were in frame with Proximal (41 repeats) or Off (39 repeats) (Fig. 1).

***mod* Gene-Mediated Phase Variation of Heat-Shock Proteins.** The coordinated phase variation of the heat-shock proteins HI0104 (HtpG), HI0542/3 (DnaKJ), and HI1237/8 (GroELS) would result in the generation of two distinct populations, one presumably more fit to respond to environmental stress. To test this hypothesis, a 46°C heat-shock killing assay was conducted, which

compared the survival of WT Rd (Distal) and Rd*mod*::*kan*. Rd*mod*::*kan*, in which the heat-shock proteins are up-regulated, was markedly more resistant to heat shock than WT Rd. A typical result is shown in Fig. 2*A*. WT Rd survivors from the 70-min time-point were picked as isolated colonies, and the *mod* repeat region sequenced to determine whether phase variants inactivating *mod* expression (e.g., Distal to Off) had been selected for in the killing assay. This analysis revealed that 33% of colonies that had survived the heat shock (Rd survivors) had switched from Distal to Off. This Rd survivor group was pooled and subjected to a further heat-shock killing assay and its fitness compared with WT Rd and Rd*mod*::*kan* (Fig. 2*B*). The Rd survivor group displayed a more resistant phenotype than WT Rd in the killing assay, similar to Rd*mod*::*kan*. A sample of this group was taken from the 50-min time-point, and sequencing of the *mod* repeat tract showed that 67% of colonies from this second round of survivors had phase varied to the Off reading frame. These results indicate that *mod* phase variation randomises expression of heat-shock proteins and creates a subpopulation of individuals of increased fitness that can be selected for by physiological and environmental stress.

**Differential Methylation of *H. influenzae* DNA by *mod* and Immunostimulatory Activity.** Differential methylation of DNA is part of the basis for immunostimulation of macrophages by bacterial DNA but not by DNA from vertebrates (31). To address the further possibility that differential methylation of *H. influenzae* DNA by *mod* influenced the degree of immune stimulation, we purified DNA from WT Rd (Distal) and Rd*mod*::*kan* and conducted macrophage-activation assays (31) (see *Supporting Experimental Procedures* and Fig. 6, which are published as supporting information on the PNAS web site). No significant difference in stimulation by DNA from WT Rd (Distal) and Rd*mod*::*kan* was observed.

## Conclusion

Simple tandem repeats associated with a particular gene indicate the potential for phase-variable expression, have led to identification of key genes mediating host–pathogen interactions, and have also influenced the selection of potential vaccine candidates. In all reported examples, phase-variable expression mediated by simple tandem repeats has been limited to the gene associated with the repeats (i.e., present within the coding sequence or promoter region of the gene exhibiting phase-variable expression). Here, we report phase-variable expression of a gene whose activity influences expression of multiple genes. The effect of methylation on gene regulation in bacterial pathogens is well established (23) and may result in increased or decreased levels of gene expression depending on the site of methylation. A global regulatory role for a phase-variable type III R-M offers distinct evolutionary advantages over reported examples of phase-variation mechanisms that operate within individual genes. Simple tandem repeats controlling expression of a specific gene must arise and expand within the promoter regions or within the coding sequence of that gene, without affecting the proper function of the promoter or structure and function of the encoded protein. In the case of phase variation mediated by type III methyltransferases, genes may come under the influence of the methyltransferases by a few point mutations generating a recognition site in a key position effecting transcriptional control of the gene. Other important bacterial pathogens, including *N. meningitidis* and *H. pylori*, also contain potentially phase-variable type III R-M systems (9), suggesting the phase-variable regulon "phasevarion" may be a commonly used mechanism mediating coordinated phase variation of multiple genes in bacterial pathogens.

We thank Prof. E. R. Moxon for comments on the manuscript. This work was supported by National Health and Medical Research Council (Australia) Program Grant 284214.

1. Bickle, T. A. & Kruger, D. H. (1993) *Microbiol. Rev.* **57,** 434–450.
2. Boyer, H. W. (1971) *Annu. Rev. Microbial.* **25,** 153–176.
3. Iida S., M., J., Bachi, B., Stalhammar-Carlemalm, M., Schrickel, S., Bickel, T.A., and Arber, W. (1983) *J. Mol. Biol.* **165,** 19–34.
4. Dryden, D. T., Murray, N. E. & Rao, D. N. (2001) *Nucleic Acids Res.* **29,** 3728–3741.
5. Ryan, K. A. & Lo, R. Y. (1999) *Nucleic Acids Res.* **27,** 1505–1511.
6. De Bolle, X., Bayliss, C. D., Field, D., van de Ven, T., Saunders, N. J., Hood, D. W. & Moxon, E. R. (2000) *Mol. Microbiol.* **35,** 211–222.
7. Saunders, N. J., Jeffries, A. C., Peden, J. F., Hood, D. W., Tettelin, H., Rappuoli, R. & Moxon, E. R. (2000) *Mol. Microbiol.* **37,** 207–215.
8. de Vries, N., Duinsbergen, D., Kuipers, E. J., Pot, R. G., Wiesenekker, P., Penn, C. W., van Vliet, A. H., Vandenbroucke-Grauls, C. M. & Kusters, J. G. (2002) *J. Bacteriol.* **184,** 6615–6623.
9. Seib, K. L., Peak, I. R. & Jennings, M. P. (2002) *FEMS Immunol. Med. Microbiol.* **32,** 159–165.
10. Weiser, J. N., Williams, A. & Moxon, E. R. (1990) *Infect. Immun.* **58,** 3455–3457.
11. van Ham, S. M., van Alphen, L., Mooi, F. R. & van Putten, J. P. (1993) *Cell* **73,** 1187–1196.
12. Blyn, L. B., Braaten, B. A. & Low, D. A. (1990) *EMBO J.* **9,** 4045–4054.
13. Richardson, A. R. & Stojiljkovic, I. (1999) *J. Bacteriol.* **181,** 2067–2074.
14. Ren, Z., Jin, H., Whitby, P. W., Morton, D. J. & Stull, T. L. (1999) *J. Bacteriol.* **181,** 5865–5870.
15. Weiser, J. N. & Pan, N. (1998) *Mol. Microbiol.* **30,** 767–775.
16. Jennings, M. P., Hood, D. W., Peak, I. R., Virji, M. & Moxon, E. R. (1995) *Mol. Microbiol.* **18,** 729–740.
17. van der Ende, A., Hopman, C. T., Zaat, S., Essink, B. B., Berkhout, B. & Dankert, J. (1995) *J. Bacteriol.* **177,** 2475–2480.
18. Moxon, E. R. & Thaler, D. S. (1997) *Nature* **387, 659,** 661–662.
19. Herbert, M. A., Hood, D. W., Moxon, E. R. (2003) *Haemophilus influenzae Protocols* (Humana, Totowa, NJ).
20. Herriott, R. M., Meyer, E. M. & Vogt, M. (1970) *J. Bacteriol.* **101,** 517–524.
21. Miller, J. H. (1972) in *Experiments in Molecular Genetics*, ed. Miller, J. M. (Cold Spring Habor Lab. Press, Woodbury, NY).
22. Grimmond, S., Van Hateren, N., Siggers, P., Arkell, R., Larder, R., Soares, M. B., de Fatima Bonaldo, M., Smith, L., Tymowska-Lalanne, Z., Wells, C. & Greenfield, A. (2000) *Hum. Mol. Genet.* **9,** 1553–1560.
23. Julio, S. M., Heithoff, D. M., Provenzano, D., Klose, K. E., Sinsheimer, R. L., Low, D. A. & Mahan, M. J. (2001) *Infect. Immun.* **69,** 7610–7615.
24. Low, D. A., Weyand, N. J. & Mahan, M. J. (2001) *Infect. Immun.* **69,** 7197–204.
25. Heithoff, D. M., Enioutina, E. Y., Daynes, R. A., Sinsheimer, R. L., Low, D. A. & Mahan, M. J. (2001) *Infect. Immun.* **69,** 6725–6730.
26. Martin, D., Cadieux, N., Hamel, J. & Brodeur, B. R. (1997) *J. Exp. Med.* **185,** 1173–1183.
27. Hanson, M. S., Slaughter, C. & Hansen, E. J. (1992) *Infect. Immun.* **60,** 2257–2266.
28. Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. (2004) *Science* **305,** 1622–1625.
29. Hartmann, E., Lingwood, C. A. & Reidl, J. (2001) *Infect. Immun.* **69,** 3438–3441.
30. Hartmann, E. & Lingwood, C. (1997) *Infect. Immun.* **65,** 1729–1733.
31. Stacey, K. J., Young, G. R., Clark, F., Sester, D. P., Roberts, T. L., Naik, S., Sweet, M. J. & Hume, D. A. (2003) *J. Immunol.* **170,** 3614–3620.