

# Phase Variation of Gonococcal Protein II: Regulation of Gene Expression by Slipped-Strand Mispairing of a Repetitive DNA Sequence

George L. Murphy,\*† Terry D. Connell,\*  
Diana S. Barritt,\* Michael Koomey,‡ and  
Janne G. Cannon\*

\* Department of Microbiology and Immunology  
University of North Carolina  
Chapel Hill, North Carolina 27599

‡ Department of Microbiology and Immunology  
University of Michigan  
Ann Arbor, Michigan 48109

## Summary

**Expression of outer membrane protein II (P.II) of *Neisseria gonorrhoeae* is subject to reversible phase variation at a rate of  $10^{-3}$ – $10^{-4}$ /cell/generation. The signal peptide coding regions of P.II genes contain variable numbers of tandem repeats of the sequence CTCTT. Changes in the number of CTCTT units, leading to frameshifting within the gene, are responsible for changes in P.II expression. Phase variation mediated by the CTCTT repeat also occurred in *E. coli*, as assayed with a P.II–alkaline phosphatase (*phoA*) gene fusion. Phase variation in both the gonococcus and *E. coli* was *recA*-independent, occurred at similar rates, and involved insertions or deletions of one or more repeat units. The characteristics of the phase variation process were consistent with a model in which expression of P.II genes is regulated by slipped-strand mispairing of the DNA in the CTCTT repeat region.**

## Introduction

Many bacterial species that are obligate human pathogens have evolved novel genetic regulatory mechanisms for varying the array of surface components exposed to the defense systems of the infected host. Synthesis of such components may be subject to phase variation (reversible oscillation between different states of expression) and to antigenic variation (synthesis of antigenically different versions of a particular component). *Neisseria gonorrhoeae*, the causative agent of gonorrhea, demonstrates both phase and antigenic variation of several different surface constituents, including a family of outer membrane proteins designated Proteins II (P.II; also referred to as opacity proteins). The different P.II proteins made by one strain can be distinguished by electrophoretic and antigenic differences among them, although all are heat-modifiable proteins of similar apparent molecular weight (24–30 kd) (Diaz and Heckels, 1982; Swanson, 1982; Swanson and Barrera, 1983). Each P.II protein of a strain shows independent phase variation between on and off states of expression at a rate of  $10^{-3}$ – $10^{-4}$ /cell/generation (Mayer, 1982; Black et al., 1984). It is possible to isolate gonococ-

cal variants expressing from zero to at least three different P.II proteins in different combinations on the surface of individual organisms (Swanson, 1982; Black et al., 1984; Robinson et al., 1988). P.II phase variation occurs in vivo during gonococcal infection (Schwalbe et al., 1985) and is probably important in gonococcal pathogenesis (reviewed in Sparling et al., 1986; Blake and Gotschlich, 1987).

Analysis of three different gonococcal strains indicates that there are at least nine complete P.II structural genes in the gonococcal genome, although the actual number may be higher (Stern et al., 1984, 1986; Connell et al., 1988; van der Ley, 1988). Comparisons of DNA sequences of cloned P.II genes show that the genes share a conserved framework, with differences among them confined primarily to two short hypervariable regions (Stern et al., 1986; Connell et al., 1988; van der Ley, 1988). The hypervariable regions encode variable antigenic determinants recognized by P.II-specific monoclonal antibodies (Connell et al., 1988, and unpublished data). There is evidence that variability within the P.II gene family is increased by recombination among P.II genes that generates new assortments of hypervariable region sequences (Stern et al., 1986; Connell et al., 1988).

In all P.II genes sequenced, the region of the gene encoding the hydrophobic core of the signal peptide is composed of identical repeats of the pentameric unit CTCTT (Stern et al., 1986; Connell et al., 1988; van der Ley, 1988). The number of repeats differs in different P.II genes and can range from 7 to 27 repeat units (Stern et al., 1986). Stern et al. (1986) showed that P.II genes are constitutively transcribed and that changes in the number of repeats occur in P.II transcripts as P.II genes turn on or off. They proposed that the CTCTT repeats are involved in a translational control mechanism of phase variation. According to this model, changes in the number of CTCTT repeats place the P.II gene in-frame with the ATG initiation codon (full-sized P.II protein synthesized) or out-of-frame (truncated polypeptide synthesized).

There are several mechanisms that might be responsible for the high frequency changes in number of CTCTT repeats in P.II mRNA. Some or all of the changes could result from slippage or stuttering of RNA polymerase as the genes are transcribed. Also, unequal crossing-over among the signal peptide coding regions of different P.II genes in the chromosome could generate insertions or deletions of CTCTT repeat units. Such a *recA*-dependent recombinational mechanism is involved in antigenic variation of pilin protein in the gonococcus (Koomey et al., 1987).

An alternative model involves slipped-strand mispairing (SSM), which can occur during DNA replication or in non-replicating duplex DNA. Local denaturation of a DNA duplex can be followed by mispairing of bases in different copies of a short tandem repeat, leading to either insertions or deletions of the repetitive sequences (reviewed in Levinson and Gutman, 1987a). SSM has been suggested to be a major factor in the expansion of simple repetitive

† Present address: Burroughs-Wellcome Company, 3030 Cornwallis Road, Research Triangle Park, North Carolina 27709.

DNA sequences in a wide variety of prokaryotes and eukaryotes. This mechanism has also been proposed to account for mutational hotspots for frameshifts at single-base runs (Streisinger et al., 1966; Streisinger and Owen, 1985; Pribnow et al., 1981; Schaaper and Dunn, 1987) and at tandem repeats of short sequence elements (Farabaugh et al., 1978; Albertini et al., 1982; Pribnow et al., 1981), although it cannot account for all frameshift mutations (deBoer and Ripley, 1988). It has been suggested that transposon excision, and perhaps a number of other illegitimate recombination events, proceed by a similar mechanism (Albertini et al., 1982; Egner and Berg, 1981; Brunier et al., 1988). The SSM model makes a number of testable predictions: all size changes should involve addition or deletion of an integral number of discrete repeat units; the most common events should be those involving loss or gain of a single repeat unit, since these events will require minimal displacement of the two DNA strands; and the frequency of frameshift should increase with increasing length of the repeat region (Levinson and Gutman, 1987b). In this study, we have tested some of these predictions concerning the mechanism of P.II phase variation.

## Results and Discussion

### Analysis of the CTCTT Repeats in the DNA of a Specific P.II Gene as It Changes Expression State

The CTCTT repeat region in P.II genes is located in the signal peptide coding region, between the ATG initiation codon and the codons for mature P.II protein. Depending on the number of CTCTT repeats, the codons for mature protein may be either in-frame or out-of-frame with the ATG codon. All of the P.II genes that have been cloned to date are out-of-frame genes that were turned off in the gonococcal chromosome (Stern et al., 1986; Connell et al., 1988; van der Ley, 1988). We have been unable to clone in-frame P.II genes, even on single-copy plasmid vectors, which has made it impossible to compare the DNA of on and off versions of a single P.II gene. To confirm that phase variation of P.II genes involves changes at the DNA level, we developed a means of identifying specific P.II genes in a preparation of gonococcal chromosomal DNA and measuring the length of their signal peptide coding regions.

Gonococcal strain FA1090 produces six different identified P.II proteins, designated P.IIa through P.IIf in order of increasing apparent molecular weight. We have produced a panel of monoclonal antibodies (MAbs) specific for unique surface exposed determinants on the P.II proteins of this strain. Four MAbs bind individually to proteins P.IIa, P.IIc, P.IIe, and P.IIf. A fifth MAb binds to both P.IIb and P.IId (Black et al., 1984; Barritt et al., 1987). In previous studies, we used the MAbs to detect genetic transformation of P.II genes into a different gonococcal strain and showed that the transformed population binds only the MAbs specific for the P.II proteins that were expressed in the donor variant (Schwalbe and Cannon, 1986). Thus, a fragment of FA1090 DNA will yield MAb-binding transformants only if

it contains an intact gene expressing the P.II protein recognized by the MAb.

To isolate a fraction of FA1090 DNA containing the P.IIa and P.IIe genes, we first determined that digestion of DNA from a P.IIa- or P.IIe-expressing variant with *Cl*I did not inactivate P.IIa or P.IIe transforming activity. *Cl*I-digested chromosomal DNA was subjected to agarose gel electrophoresis, and different size fractions were screened by the transformation assay to determine which contained the intact P.IIa and P.IIe genes. Both were contained in the 8–10 kb size fraction of *Cl*I fragments, although not on the same fragment. No other intact P.II genes were contained in this size fraction of DNA. We isolated a series of colony variants differing in expression of P.IIa or P.IIe (Figure 1), isolated the 8–10 kb size fraction of *Cl*I fragments from each variant, and confirmed the expression state of the P.IIa and P.IIe genes in the isolated DNA fragments with the transformation assay. This size fraction was then digested with enzymes that cut in conserved sequences immediately flanking the CTCTT repeat region, electrophoresed on polyacrylamide gels, Southern blotted, and probed with an oligonucleotide homologous to the CTCTT repeats. This procedure allowed us to measure the precise lengths of the CTCTT regions of the P.IIa and P.IIe genes as they switched expression in different variants.

Two restriction fragments, containing the signal peptide coding DNAs of the P.IIa and P.IIe genes, hybridized to the probe (Figure 1). Changes in the length of the fragments occurred as the two genes switched on or off in different colony variants; changes in the large fragment correlated with switches in expression of P.IIa, and changes in the smaller fragment correlated with switches in P.IIe expression. All of the changes in length of the fragments involved addition or deletion of 5 bp or 10 bp, which corresponds to the size of one or two CTCTT repeats. In each case, the length of a fragment in a gene that was turned on was consistent with an in-frame gene, and its length in an off gene was consistent with an out-of-frame gene, based on comparisons to the DNA sequence of cloned FA1090 P.II genes (Connell et al., 1988). This experiment demonstrated that changes in length of the CTCTT region, in increments corresponding to the size of the repeat unit, accompany phase variation of a specific P.II gene. It is possible that transcriptional slippage also contributes to phase variation, but changes at the DNA level occurred in all of the phase switches we examined.

### *recA* Independence of P.II Phase Variation

To determine if P.II phase variation depends on gonococcal *recA* function, we compared P.II switching in a *recA* derivative of strain FA1090 to that in the wild-type parent. FA1090*recA* produced colony variants capable of binding each of the five P.II-specific MAbs. We isolated a P.IIa-expressing variant of FA1090*recA* and three P.II<sup>-</sup>switchers derived from it. The frequency with which P.II<sup>-</sup> variants appeared in the P.IIa-expressing population was the same as for wild-type FA1090 (data not shown). Western blot analysis showed that the P.IIa protein of FA1090*recA* had the same electrophoretic mobility as P.IIa of FA1090

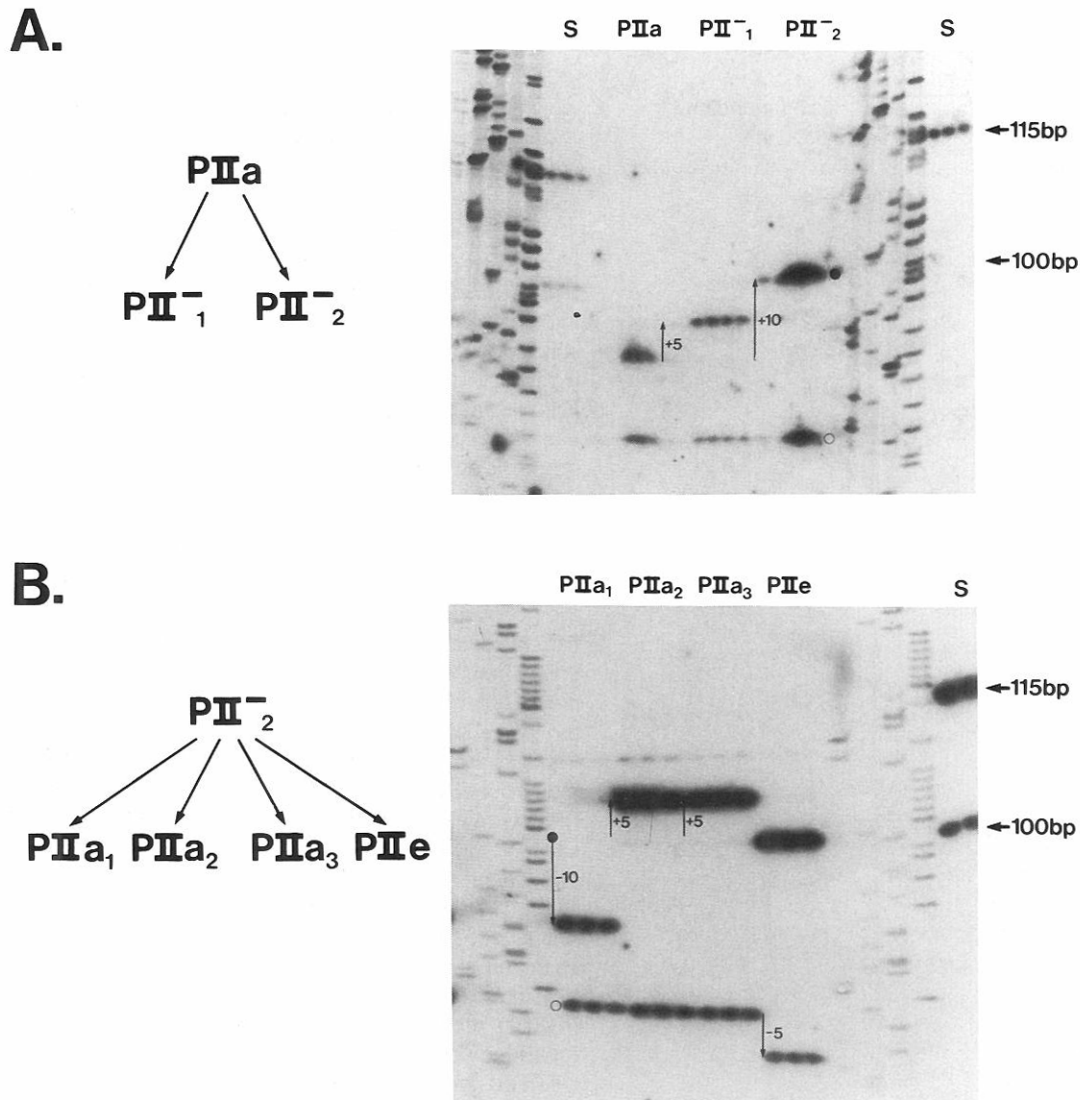
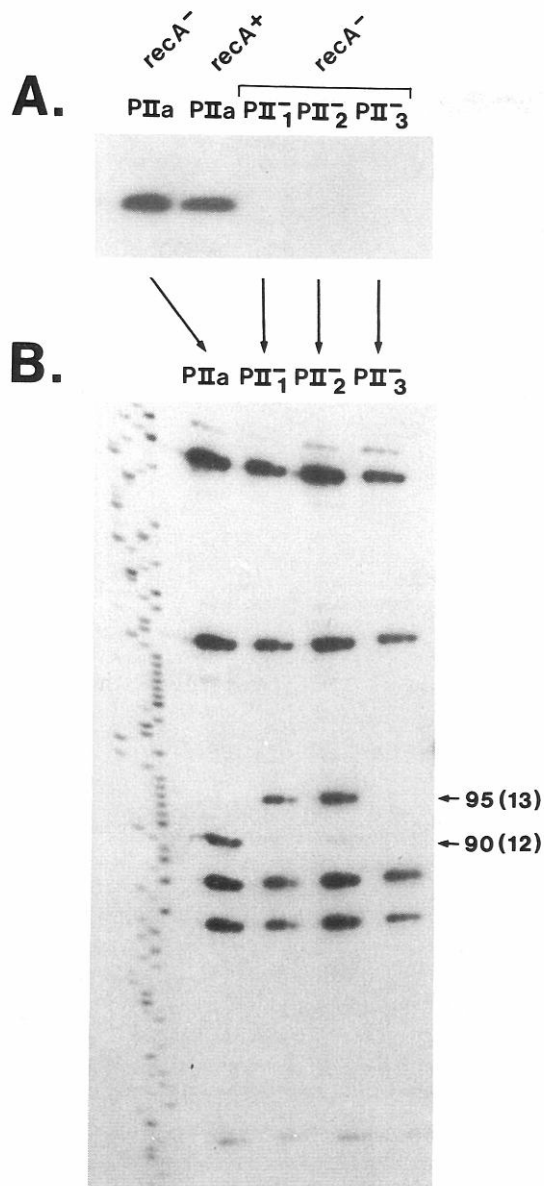


Figure 1. Southern Blot Analysis of the CTCTT-Containing Regions of a P.IIa Gene and a P.IIe Gene as They Change Expression State

The diagrams on the left show the lineage of variants that were derived as single-step switchers in P.II expression, and the photographs on the right show the Southern blot analysis of those variants. Variants were characterized as to P.II expression with the panel of P.II-specific MABs. The designation P.II<sup>-</sup> indicates that no P.II proteins were expressed; the designation P.IIa indicates that only P.IIa was expressed in a variant. In (A), two independent P.II<sup>-</sup> variants were derived from one P.IIa variant. In (B), one of those P.II<sup>-</sup> variants (P.II<sup>-</sup><sub>2</sub>) was used to derive four independent switchers in expression of either P.IIa or P.IIe. Chromosomal DNA of each variant was digested with *Clal*. The *Clal* fraction containing the intact P.IIa and P.IIe genes was identified by the genetic transformation assay, then digested with *HinfI* and *HhaI*, which cut in conserved sequences immediately flanking the CTCTT repeats. The *HinfI* site is 22 bp upstream of the first CTCTT repeat; the *HhaI* site is 8 bp downstream of the last CTCTT repeat. Fragments were electrophoresed on polyacrylamide gels, Southern blotted, and probed with a CTCTT-specific oligonucleotide. To measure the changes in length of the fragments, they were compared with DNA sequencing reactions run on M13 clones of known DNA sequence and to size standards (100 bp and 115 bp) consisting of the same digests of cloned P.II genes with known numbers of CTCTT repeats (lanes marked S). The fragment corresponding to the P.IIa gene is indicated with a closed circle; the fragment corresponding to the P.IIe gene is indicated with an open circle. The faint bands visible in the Southern in (B) either represent partial digestion products, or are due to the unavoidable heterogeneity in P.II expression that occurs within a colony variant.

(Figure 2). To examine the CTCTT repeat regions of P.II genes in the *recA* strain, total chromosomal DNA of the variants was digested with the enzymes that cut on either side of the CTCTT repeat region, then electrophoresed, Southern blotted, and probed with the CTCTT-homologous oligonucleotide (Figure 2). Five major bands and

several minor bands hybridized to the probe. A 90 bp band corresponding to 12 CTCTT repeats was present in the P.IIa-expressing variant but absent in the three variants in which P.IIa was turned off. In two of the P.II<sup>-</sup> switchers, a new band of 95 bp corresponding to 13 CTCTT repeats was present. In the third P.II<sup>-</sup> variant, no new band was



**Figure 2.** Analysis of P.II Variants of FA1090 *recA*  
(A) Western blot. Whole cells of different variants of FA1090 or FA1090*recA* were solubilized at 100°C, electrophoresed in a 4%–30% polyacrylamide gradient gel, transferred to nitrocellulose, and probed with MA b H.138 (P.IIa-specific) and <sup>125</sup>I-Protein A. The P.IIa protein of FA1090 and FA1090*recA* migrates with an apparent molecular weight of 28K (solubilized at 100°C), determined by comparison to molecular weight standards. The three P.II<sup>-</sup> variants of FA1090*recA* were independent switchers derived from the P.IIa variant of FA1090*recA*. Only the relevant portion of the gel is shown. (B) Southern blot. Chromosomal DNA of the four FA1090*recA* variants shown in (A) was digested with *Hin*I and *Hha*I, which cut in conserved sequences immediately flanking the CTCTT repeat regions of P.II genes. The *Hin*I site is 22 bp upstream of the first CTCTT repeat; the *Hha*I site is 8 bp downstream of the last CTCTT repeat. The DNA fragments were electrophoresed, Southern blotted, and probed with the CTCTT-specific oligonucleotide as in Figure 1. The numbers at the right of the figure indicate the sizes of the fragments in bp (determined from the DNA sequencing reaction at the left side of the blot). The number of CTCTT repeats corresponding to the fragment lengths are indicated in parentheses.

visible, presumably because the fragment containing the CTCTT repeat region of the P.IIa gene was now the same length as one of the other fragments. Thus, P.II phase variation was indistinguishable in strain FA1090 and its *recA* derivative, in terms of the array of P.II proteins produced, the frequency of variation, and the molecular events accompanying the process.

#### Phase Variation of a P.II-*phoA* Gene Fusion in *E. coli*

To determine if P.II phase variation was dependent only on the region of the gene containing the CTCTT repeats, we constructed a gene fusion in which P.II signal peptide coding DNA was fused to the structural gene for *Escherichia coli* alkaline phosphatase (*phoA*) with its own promoter and signal peptide coding region deleted, so that PhoA synthesis and export were under the control of the P.II sequences (Hoffman and Wright, 1985). The gene fusion was constructed from a cloned out-of-frame P.IIa gene, with the signal peptide coding DNA, the promoter, and approximately 350 bp of upstream sequence derived from the gonococcal clone (Figure 3). In the fusion construction, the codons for mature PhoA are in the same translational frame as those for mature P.II protein, so that production of PhoA protein requires a frame-shift placing the *phoA* coding sequence in the same translational frame as the P.II ATG initiation codon. If the ATG codon remains out-of-frame with the rest of the gene, translation of PhoA terminates prematurely. The activity of PhoA, which can be detected with the chromogenic substrate XP, is dependent on secretion of the protein (Hoffman and Wright, 1985). Therefore, detection of PhoA production from an in-frame gene by blue colony color also requires appropriate secretion of the protein across the inner membrane.

*E. coli* HB101 containing the gene fusion cloned in pBR322 formed blue colonies of varying color intensity on media containing XP, but no white colonies. However, when the fusion was cloned into plasmid pOU71 (Larsen et al., 1984), which replicates as a single copy per chromosome at 30°C, there was reversible phase variation of blue to white and white to blue colonies (Figure 4). The inability to detect blue–white phase variation when the gene fusion was present on a multicopy vector may have been due to independent phase variation of the gene copies in a cell, resulting in production of some PhoA by cells in each colony. All subsequent experiments with the P.II-*phoA* gene fusion were done with pFLOB901 (the derivative of single-copy vector pOU71) in cultures maintained at 30°C.

The rate of phase variation of the P.II-*phoA* fusion in *E. coli* HB101 (*recA*) containing pFLOB901 was  $1.2 \pm 0.2 \times 10^{-3}$ /cell/generation for blue to white transitions and  $2.6 \pm 0.3 \times 10^{-4}$ /cell/generation for white to blue transitions (mean  $\pm$  sem;  $n = 10$  independent colonies for each). In *E. coli* RR1, which is isogenic with HB101 except *recA*<sup>+</sup>, the rate for blue to white switching was  $2.2 \pm 0.4 \times 10^{-3}$ /cell/generation ( $n = 6$ ). Thus, the rate of blue–white phase variation in these strains of *E. coli* was similar to the rate of P.II phase variation in the gonococcus, and was *recA*-independent.

We also cloned the gene fusion into pLES2, a gonococal-*E. coli* shuttle vector, and mobilized it into the gono-

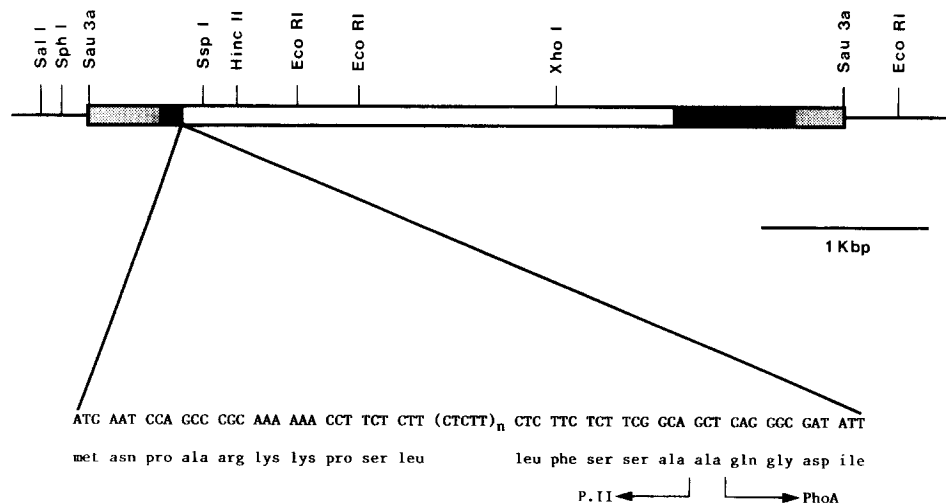


Figure 3. Diagram of the P.II-*phoA* Gene Fusion (pFLOB900)

The gene fusion was constructed by ligating a fragment containing the *phoA* gene (lacking its promoter and signal peptide coding region) into a unique site in an out-of-frame FA1090 P.IIa gene cloned in pBR322. The solid line indicates vector sequence, and the box indicates the fusion gene construction. The open area represents the inserted *phoA*-containing fragment; the solid black and dotted areas represent the P.IIa coding sequence and flanking sequence, respectively, from the gonococcal clone. The upstream flanking sequence derived from the gonococcal clone (350 bp) does not contain any open reading frames with appropriate signals for initiation of transcription and translation (Connell et al., 1988). The DNA sequence at the fusion junction is shown, starting with the ATG initiation codon for the first amino acid of the P.II signal peptide. The codons for PhoA, starting at Gln<sub>14</sub>, are in the same frame as the codons for mature P.II protein. The Sal-XhoI fragment from pFLOB900 was cloned into the single copy vector pOU71, resulting in plasmid pFLOB901.

coccus, hoping to determine if phase variation of the fusion would be similar in the gonococcus and in *E. coli*. Gonococci normally form white colonies on medium containing XP. Gonococci containing the gene fusion formed blue colonies but did not show detectable blue-white phase variation, probably because of the multicopy nature of vector pLES2 (Stein et al., 1983). No single-copy plasmids capable of replicating in the gonococcus are currently available.

#### Mechanism of Phase Variation of the Gene Fusion

If phase variation of the gene fusion in *E. coli* involves SSM of the repeated CTCTT element, the process should show the following features: all changes will involve integral numbers of discrete repeat units; there will be a bias towards loss or gain of single units; and the rate of frameshifting will increase as the length of the repetitive region increases (Levinson and Gutman, 1987a, 1987b). To characterize the events accompanying blue-white phase variation of the fusion construction, we sequenced the signal peptide coding regions of 17 variants of HB101(pFLOB901) that had undergone PhoA phase transitions. In each case, phase transitions between blue and white colony types were accompanied by changes in the number of CTCTT repeats. Sixteen of the switches involved deletions (six) or insertions (10) of a single repeat; one switch was accompanied by insertion of four CTCTT units (data not shown). All of the blue colonies contained plasmid with a number of CTCTT repeats placing the ATG initiation codon and the codons for mature PhoA in the same frame; the fusion gene in all of the white colonies was out-of-frame. No other DNA sequence alterations oc-

curred in the signal peptide coding region of these variants.

Once the sequence analysis indicated that the mechanism of phase variation did involve changes in CTCTT number, we analyzed a larger set of phase variants using the Southern blotting method described earlier (Figure 1) to measure the length of the CTCTT region in each variant. Fifty-one variants of HB101(pFLOB901) were analyzed. Every switch in PhoA expression was accompanied by a change in the length of the CTCTT repeat region, always by 5 bp or a multiple of 5 bp (data not shown). Forty-five were accompanied by insertion or deletion of 5 bp, 4 by changes of 10 bp, 1 by insertion of 20 bp, and 1 by a deletion of 35 bp. We also analyzed six PhoA phase transitions in strain RR1 (*recA*<sup>+</sup>) containing pFLOB901; all involved changes of 5 bp in the length of the CTCTT repeat region (one deletion and five insertions).

In each of three sequential cycles of isolating phase variants, we chose the ones with the longest and shortest CTCTT regions to use as the source for another cycle of variants (Figure 5). The longest signal peptide coding regions we obtained contained 24 CTCTT repeats in a PhoA<sup>+</sup> colony and 25 repeats in a PhoA<sup>-</sup> colony. The shortest were 12 (PhoA<sup>+</sup>) and 11 (PhoA<sup>-</sup>) repeats. We do not know if these numbers represent the actual limits of length of the CTCTT repeats, or if greater extremes would be found by screening additional variants. The phase variation rate (white to blue) for the variant with 25 repeats (PhoA<sup>-</sup>25) was  $4.6 \pm 1.2 \times 10^{-4}$ /cell/generation; the rate for the variant with 11 repeats (PhoA<sup>-</sup>11) was  $1.9 \pm 0.3 \times 10^{-4}$ /cell/generation (five separate determinations for each). Southern blot analysis of plasmid from several of

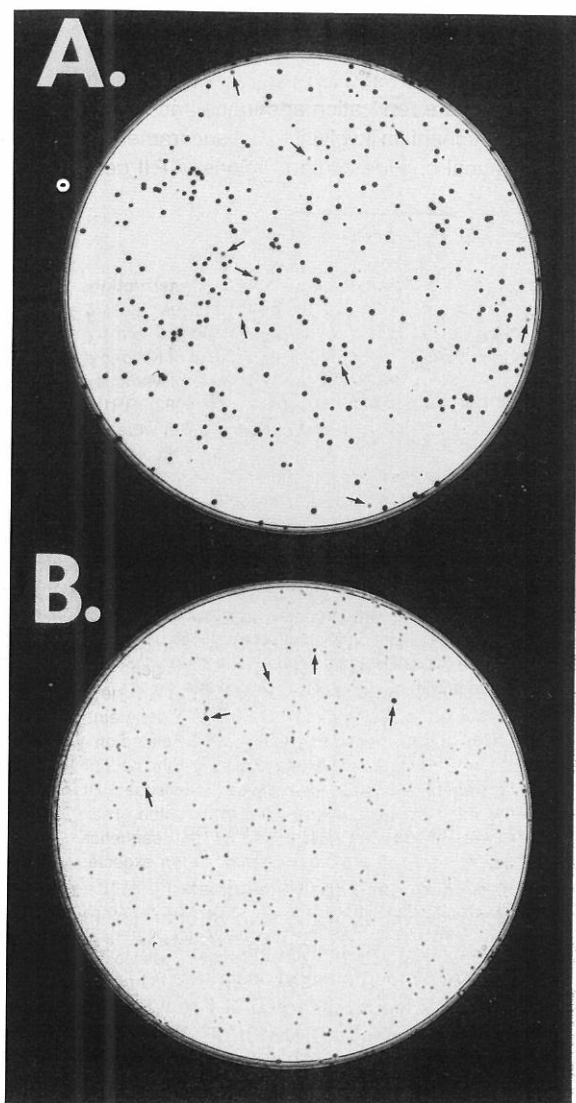


Figure 4. Phase Variation between PhoA<sup>+</sup> (Blue on Indicator Plates) and PhoA<sup>-</sup> (White on Indicator Plates) of the P<sub>II</sub>-*phoA* Fusion on a Single-Copy Vector (pFLOB901) in *E. coli* HB101

In (A), a single blue colony was diluted and plated on indicator medium; in (B), a single white colony was diluted and plated. The arrows indicate some of the colonies of the other color that appear on each plate.

the PhoA<sup>+</sup> switchers derived from PhoA<sup>-</sup>25 showed that they generally lost large numbers of repeats (>6), whereas the PhoA<sup>+</sup> derivatives of PhoA<sup>-</sup>11 gained single CTCTT repeats. It is possible that the mechanism for loss of CTCTTs from a longer repeating stretch might be different than from a shorter one. However, because every colony will actually contain a mixture of cells with different numbers of CTCTT repeats, it is impossible to conclude with certainty that those PhoA<sup>+</sup> progeny arose in single steps from PhoA<sup>-</sup>25.

Because this gene fusion will not detect all possible frameshifts, it cannot be used to compare relative frequencies of deletion and insertion. For example, starting

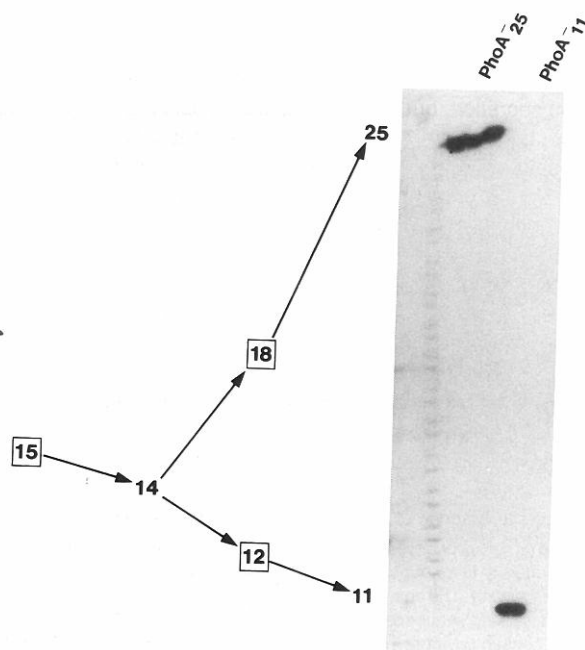


Figure 5. Isolation of Variants of HB101(pFLOB901) with the Longest and Shortest CTCTT Repetitive Region

In the diagram on the left of the figure, the arrows indicate the derivation of the variants, and the numbers indicate the number of CTCTT repeats in each variant (determined by Southern blot analysis). PhoA<sup>+</sup> variants have the number boxed; the number is not boxed for PhoA<sup>-</sup> variants. The starting variant was a blue colony containing 15 CTCTT repeats in the fusion gene. The Southern blot shows the length of the repetitive region for the longest and shortest ones isolated (PhoA<sup>-</sup>25 and PhoA<sup>-</sup>11). Plasmid DNAs from the two variants were digested with *Hin*I and *Dde*I, which cut on either side of the CTCTT region, subjected to electrophoresis on a polyacrylamide gel, and probed with a CTCTT-specific oligonucleotide as in Figure 1. A DNA sequencing reaction on a known DNA sequence provided size standards for determining the size of the fragments. The four lanes of the sequence reaction are not all clearly visible in this exposure of the autoradiograph but were visible in a longer exposure.

with a blue (PhoA<sup>+</sup>) colony, loss or gain of one CTCTT would result in a white (PhoA<sup>-</sup>) phenotype. Starting with a white colony with the gene in the -1 frame, addition of one CTCTT would cause a detectable phase shift to blue, but loss of one CTCTT would not change the phenotype of the colony. Since loss or gain of single CTCTT units was the most common event, the rate of blue to white phase variation should be greater than that of white to blue. Consistent with this prediction, the blue to white rate in HB101 (pFLOB901) was 4-fold higher than the white to blue rate.

Detection of phase variation by blue colony formation also indicated that the unusual signal peptide encoded by the repeating CTCTT element functioned appropriately to direct secretion of the fusion protein in *E. coli*. The repeating sequence Leu-Phe-Ser-Ser-Leu encoded by the CTCTT repeats forms the hydrophobic core of the P<sub>II</sub> signal peptide. In a gene with 12 CTCTT, the hydrophobic core of the encoded signal peptide would be 20 amino acids long; that of a gene with 24 repeats would be 40 amino acids long. All are longer than the hydrophobic

cores of most prokaryotic signal peptides that have been studied (von Heijne, 1985; Pollitt and Inouye, 1987). Production of active alkaline phosphatase is dependent on secretion, but may not require processing of the signal peptide, so we do not know if the unusually long signal peptides were processed. The range of variation in CTCTT number will undoubtedly be constrained by the requirement for a translated signal peptide that will function in the secretion and localization of P.II.

In these experiments, we analyzed a total of 74 phase transitions in *E. coli* containing the P.II-*phoA* gene fusion. All of the transitions involved loss or gain of CTCTT units, as shown directly in variants that were sequenced and indirectly in variants analyzed by Southern blotting. There was no evidence for changes in other sequences in the signal peptide coding region or in partial CTCTT units. The majority of the transitions (67 of 74) involved loss or gain of single CTCTT units. The phase variation rate (white to blue) increased 2.5-fold with an increase from 11 to 25 CTCTT repeats. Although this magnitude of increase in rate with increasing repetitive region length is not as dramatic as that observed for single-base runs (Streisinger and Owen, 1985; Pribnow et al., 1981; Kunkel, 1986), it is similar to that obtained for a repeating 2 bp element (Levinson and Gutman, 1987b).

### Regulation of Gene Expression by SSM

The process of SSM has been implicated in evolutionary expansion of repeated sequences and in mutational hotspots for frameshifts but has not heretofore been shown to be a mechanism for regulation of gene expression. The features of CTCTT-mediated frameshifting in the gonococcus and in *E. coli* were consistent with the model in which P.II phase variation involves SSM. Phase variation was an inherent property of the repeated CTCTT sequence elements and did not depend on specific gonococcal factors or on mature P.II coding sequences. Phase variation of the P.II-*phoA* gene fusion in *E. coli* involved the same events (loss and gain of CTCTTs) and occurred at the same high rate ( $10^{-3}$ – $10^{-4}$ /cell/generation) as P.II phase variation in the gonococcus. The process was *recA*-independent in both the gonococcus and *E. coli*, indicating that homologous recombination is probably not involved in these events. Although *recA* mutations do not abolish all recombinational events (Matfield et al., 1985), the high rate of frameshifting of the single cellular copy of the P.II-*phoA* gene in the *recA* *E. coli* strain argues against the likelihood of a recombination-based mechanism. Phase variation of the gene fusion in *E. coli* showed the predicted characteristics of loss or gain of integral numbers of CTCTT units, bias toward single unit changes, and increased frameshifting rate with increased length of the repetitive region.

SSM requires transient formation of single-stranded DNA, prior to the misalignment of copies of a repeated sequence. It is interesting to note that polypyrimidine/polypurine stretches of DNA, such as the CTCTT repeat studied here, are susceptible to local denaturation and the formation of unusual DNA structures (Johnston, 1988; Htun and Dahlberg, 1988; Wells, 1988). The data pre-

sented here support the hypothesis that phase variation of gonococcal P.II proteins is dependent only on the special properties of the CTCTT repeat sequence in its interaction with the replication apparatus and is regulated by a novel mechanism involving SSM and frameshifting within the signal peptide coding regions of P.II genes.

### Experimental Procedures

#### Bacterial Strains, Plasmids, and Plasmid Constructions

*N. gonorrhoeae* strain FA1090 has been previously described (Black et al., 1984; Connell et al., 1988). FA1090 colony variants were grown on GC Base agar (Difco) with the supplements of Kellogg et al. (1963) at 37°C in a 5% CO<sub>2</sub> atmosphere. All FA1090 variants were non-piliated. *E. coli* strains HB101 (Maniatis et al., 1982), RR1 (Maniatis et al., 1982), and JM83 (Vieira and Messing, 1982) were grown on LB agar (Maniatis et al., 1982) or MacConkey agar (Difco). When appropriate, ampicillin (50 µg/ml) or XP (5-bromo-4-chloro-3-indolyl phosphate, 40 µg/ml; Sigma) were added. Although all *E. coli* strains used possess a chromosomal *phoA* locus, the gene is not expressed except under low phosphate conditions (Torriani, 1960). Colonies were white on plates of LB plus XP. *E. coli* strains were made competent with CaCl<sub>2</sub> and transformed as described (Maniatis et al., 1982).

Plasmid pFLOB709 contains the out-of-frame P.IIa gene of FA1090 on a 1.6 kb Sau3AI fragment cloned into the BamHI site of pBR322 (Connell et al., 1988). To construct the P.II-*phoA* gene fusion (designated pFLOB900), pFLOB709 was linearized at a unique NotI site located 15 bp into the coding region of mature P.IIa protein. The *phoA* gene, beginning with the codon for Gln<sub>14</sub>, was isolated on a 3 kb PstI fragment from plasmid pCH2 (Hoffman and Wright, 1985). Both DNA fragments were blunt-ended with mung bean nuclease (BRL), according to the manufacturer's instructions, and ligated. The sequence across the fusion joint was determined by DNA sequence analysis, which showed that a few more nucleotides than expected were removed by the mung bean nuclease treatment.

The gene fusion was cloned into the EcoRI site of vector pOU71 (Larsen et al., 1984), which replicates as a single copy per chromosome when cells are grown at 30°C. The gene fusion was isolated on the 3 kb Sall-XhoI fragment of pFLOB900 (diagrammed in Figure 3). Both fragments were blunt-ended with mung bean nuclease and ligated, resulting in plasmid pFLOB901. The 3 kb Sall-XhoI fragment was also cloned into the unique Sall site of the *E. coli*-*N. gonorrhoeae* shuttle vector pLES2 (Stein et al., 1983), resulting in plasmid pFLOB902.

#### Immunochemical Techniques

The five MAbs specific for the six identified P.II proteins of FA1090 have been previously described (Black et al., 1984; Barritt et al., 1987). Procedures for their use in ELISA, colony blot RIA, SDS-PAGE, and Western blotting have been described (Black et al., 1984).

#### Gonococcal Genetic Techniques

Procedures for genetic transformation of P.II genes into gonococcal strain FA19 and identification of transformants with the P.II-specific MAbs were as previously described (Schwalbe and Cannon, 1986). The *recA* derivative of strain FA1090 was constructed by transformation-mediated marker rescue, using strain VD302 (Koomey and Falkow, 1987) as the donor. VD302 is a *recA* mutant of *N. gonorrhoeae* MS11mk, in which the *recA* gene is insertionally inactivated with a selectable β-lactamase marker. VD302 DNA was digested with ClaI, and the 6–7 kb size fraction of ClaI fragments, which contained the *recA* gene, was used to transform FA1090 under nonsaturating conditions. The *RecA* phenotype of an ampicillin-resistant FA1090 transformant was confirmed by testing for UV sensitivity, inability to be transformed by chromosomal DNA, and stabilization of pilus phenotype (Koomey and Falkow, 1987; Koomey et al., 1987).

Mobilization of pFLOB902 from *E. coli* JM83 to a spectinomycin-resistant derivative of gonococcal strain F62 was performed by the filter mating procedure previously described (Connell et al., 1988). Spectinomycin (200 µg/ml) selected against the donor, and penicillin G (1 µg/ml) selected for gonococcal transconjugants containing pFLOB902.

### Southern Hybridizations

Gonococcal chromosomal DNA was isolated by the method of Stern et al. (1984). Plasmid DNA was isolated by the alkaline lysis method (Maniatis et al., 1982). DNA was digested with restriction enzymes (BRL, New England Biolabs) according to the manufacturer's specifications.

For analysis of chromosomal fragments, *Cla*I-digested FA1090 DNA was electrophoresed through 0.7% agarose gels; DNA in the 8–10 kb size range was electroeluted and further digested with *Hin*FI and *Hha*I, which cut on either side of the CTCTT repeat region in the signal peptide coding region of P.II genes. The *Hin*FI site is 22 bp upstream of the first CTCTT repeat; the *Hha*I site is 8 bp downstream of the last CTCTT repeat. The sites for these enzymes are conserved in all published P.II gene sequences (Stern et al., 1986; Connell et al., 1988; van der Ley, 1988). The enzymes *Hin*FI and *Dde*I were used to isolate the corresponding region of the P.II-*phoA* gene fusion from pFLOB901; the *Dde*I site is 7 bp downstream of the last CTCTT repeat in the gene fusion construction. Small chromosomal and plasmid DNA fragments were electrophoresed through 6% polyacrylamide/8 M urea gels and electroblotted to GeneScreen (New England Nuclear) or Zetabind (AMF) in 1 × TAE (40 mM Tris-acetate, 2 mM EDTA) for 1.5 hr at a constant current of 90 mA. Filters were hybridized with 5–10 pmol of a synthetic oligonucleotide (5'-AAGAG<sub>5</sub>-3') (Connell et al., 1988) that was end-labeled with [ $\gamma$ -<sup>32</sup>P]ATP (ICN) (Maniatis et al., 1982). Oligonucleotides were synthesized and purified as previously described (Connell et al., 1988). Hybridizations were done in 0.5 × SSPE (1 × SSPE: 150 mM NaCl, 10 mM NaH<sub>2</sub>PO<sub>4</sub>, 1 mM EDTA [pH 7.4]), 0.5% SDS, 5 × Denhardt's solution, and 20 mM NaPPi at 42°C (Maniatis et al., 1982). Filters were washed twice in 1 × SSPE, 1% SDS, 5 mM NaPPi at 42°C–45°C for 45 min under high stringency conditions. Size standards were sequencing reactions done on M13 clones of known DNA sequence by the method of Sanger et al. (1977).

### Measurement of Phase Transition Rates

Rates of phase transition were determined as described by Eisenstein (1981), using the formula (M/N)/g, where M/N is the ratio of the minority population (PhoA<sup>+</sup> or PhoA<sup>-</sup>) to the total population of colony-forming units, and g is the number of generations of growth from a single cell to the total number of cells within a colony. Control phase transition rates (PhoA<sup>-</sup> to PhoA<sup>+</sup>) for E. coli HB101(pOU71) and RR1(pOU71) were less than 1 × 10<sup>-6</sup>/cell/generation.

### DNA Sequencing

Plasmids used for DNA sequencing were prepared as described above. An 820 bp *Sph*I-*Ssp*I fragment of pFLOB901 from PhoA expression switchers was cloned into M13mp18 (New England Biolabs) that had been digested with *Sma*I and *Sph*I, ligating appropriate fragments isolated from 1% gels of low gelling temperature agarose (Sea-Plaque, FMC) as described (Struhl, 1985), and transforming into E. coli DH5aF' (BRL). Sequencing was done by the method of Sanger et al. (1977), and priming was done with an oligonucleotide complementary to a region approximately 40 bp upstream from the ATG initiation codon of the P.II gene (5'-CCGCCTTGAAGCATCAGTG-3'). DNA from at least four independent M13 clones was sequenced for each of the PhoA variants analyzed, to exclude the possibility that changes in the length of the CTCTT region were occurring during the M13 transfection (Levinson and Gutman, 1987b).

### Acknowledgments

We thank Gour Biswas for performing the gonococcal conjugation experiment, John Puziss for preliminary gonococcal transformation experiments, Clyde Hutchison III and Bruce Wisely for synthesis of oligonucleotides, JoAnn Dempsey for preparation of the figures, and Lynn Brooks for expert preparation of the manuscript. This work was supported by Public Health Service grants AI15036 (to J. G. C.) and AI10615 (to M. K.) and National Research Service Awards AI07001 and AI07151, all from the National Institute of Allergy and Infectious Diseases.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby

marked "advertisement" in accordance with 18 U.S.C. Section 34 solely to indicate this fact.

Received October 11, 1988; revised December 2, 1988.

### References

- Albertini, A. M., Hofer, M., Calos, M. P., and Miller, J. H. (1982). On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. *Cell* 29, 319–328.
- Barritt, D. S., Schwalbe, R. S., Klapper, D. G., and Cannon, J. G. (1987). Antigenic and structural differences among six Proteins II expressed by a single strain of *Neisseria gonorrhoeae*. *Infect. Immunol.* 55, 2026–2031.
- Black, W. J., Schwalbe, R. S., Nachamkin, I., and Cannon, J. G. (1984). Characterization of *Neisseria gonorrhoeae* protein II phase variation by use of monoclonal antibodies. *Infect. Immunol.* 45, 453–457.
- Blake, M. S., and Gotschlich, E. C. (1987). Functional and immunologic properties of pathogenic *Neisseria* surface proteins. In *Bacterial Outer Membranes as Model Systems*, M. Inouye, ed. (New York: John Wiley and Sons), pp. 377–400.
- Brunier, D., Michel, B., and Ehrlich, S. D. (1988). Copy choice illegitimate DNA recombination. *Cell* 52, 883–892.
- Connell, T. D., Black, W. J., Kawula, T. H., Barritt, D. S., Dempsey, J. A., Kverneland, K., Jr., Stephenson, A., Schepart, B., Murphy, G. L., and Cannon, J. G. (1988). Recombination among protein II genes of *Neisseria gonorrhoeae* generates new coding sequences and increases structural variability in the Protein II family. *Mol. Microbiol.* 2, 227–236.
- de Boer, J. G., and Ripley, L. S. (1988). An *in vitro* assay for frameshift mutations: hotspots for deletions of 1 bp by Klenow-fragment polymerase share a consensus DNA sequence. *Genetics* 118, 181–191.
- Diaz, J.-L., and Heckels, J. E. (1982). Antigenic variation of outer membrane protein II in colonial variants of *Neisseria gonorrhoeae* P9. *J. Gen. Microbiol.* 128, 585–591.
- Egner, C., and Berg, D. E. (1981). Excision of transposon Tn5 is dependent on the inverted repeats but not on the transposase function of Tn5. *Proc. Natl. Acad. Sci. USA* 78, 459–463.
- Eisenstein, B. (1981). Phase variation of type 1 fimbriae in *Escherichia coli* is under transcriptional control. *Science* 214, 337–338.
- Farabaugh, P. J., Schmeissner, U., Hofer, M., and Miller, J. H. (1978). Genetic studies of the *lac* repressor. VII. On the molecular nature of spontaneous hotspots in the *lacI* gene of *Escherichia coli*. *J. Mol. Biol.* 126, 847–863.
- Hoffman, C. S., and Wright, A. (1985). Fusions of secreted proteins to alkaline phosphatase: an approach for studying protein secretion. *Proc. Natl. Acad. Sci. USA* 82, 5107–5111.
- Htun, H., and Dahlberg, J. E. (1988). Single strands, triple strands, and kinks in H-DNA. *Science* 241, 1791–1796.
- Johnston, B. H. (1988). The S1-sensitive form of d(C-T)<sub>n</sub>-d(A-G)<sub>n</sub>: chemical evidence for a three-stranded structure in plasmids. *Science* 241, 1800–1804.
- Kellogg, D. S., Peacock, W. L., Deacon, W. E., Brown, L., and Pirkle, C. I. (1963). *Neisseria gonorrhoeae*. I. Virulence genetically linked to clonal variation. *J. Bacteriol.* 85, 1274–1279.
- Koomey, J. M., and Falkow, S. (1987). Cloning of the *recA* gene of *Neisseria gonorrhoeae* and construction of gonococcal *recA* mutants. *J. Bacteriol.* 169, 790–795.
- Koomey, M., Gotschlich, E. C., Robbins, K., Bergstrom, S., and Swanson, J. (1987). Effects of *recA* mutations on pilus antigenic variation and phase transitions in *Neisseria gonorrhoeae*. *Genetics* 117, 391–398.
- Kunkel, T. A. (1986). Frameshift mutagenesis by eucaryotic DNA polymerases *in vitro*. *J. Biol. Chem.* 261, 13581–13587.
- Larsen, J. E. L., Gerdes, K., Light, J., and Molin, S. (1984). Low-copy-number plasmid-cloning vectors amplifiable by derepression of an inserted foreign promoter. *Gene* 28, 45–54.
- Levinson, G., and Gutman, G. A. (1987a). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* 4, 203–221.



- Levinson, G., and Gutman, G. A. (1987b). High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucl. Acids Res.* 15, 5323–5338.
- Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).
- Matfield, M., Badawi, R., and Brammar, W. J. (1985). *Rec*-dependent and *Rec*-independent recombination of plasmid-borne duplications in *Escherichia coli* K12. *Mol. Gen. Genet.* 199, 518–523.
- Mayer, L. W. (1982). Rates of *in vitro* changes of gonococcal colony opacity phenotypes. *Infect. Immunol.* 37, 481–485.
- Pollitt, S., and Inouye, M. (1987). Structure and functions of the signal peptide. In *Bacterial Outer Membranes as Model Systems*, M. Inouye, ed. (New York: John Wiley and Sons), pp. 117–139.
- Pribnow, D., Sigurdson, D. C., Gold, L., Singer, B. S., and Napoli, C. (1981). rII cistrons of bacteriophage T4: DNA sequence around the intercistronic divide and positions of genetic landmarks. *J. Mol. Biol.* 149, 337–376.
- Robinson, E. N., Jr., Clemens, C. M., McGee, Z. A., and Cannon, J. G. (1988). Immunoelectron microscopic localization of outer membrane proteins II on the surface of *Neisseria gonorrhoeae*. *Infect. Immunol.* 56, 1003–1006.
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
- Schaaper, R. M., and Dunn, R. L. (1987). Spectra of spontaneous mutations in *Escherichia coli* strains defective in mismatch correction: the nature of *in vivo* DNA replication errors. *Proc. Natl. Acad. Sci. USA* 84, 6220–6224.
- Schwalbe, R. S., and Cannon, J. G. (1986). Genetic transformation of genes for protein II in *Neisseria gonorrhoeae*. *J. Bacteriol.* 167, 186–190.
- Schwalbe, R. S., Sparling, P. F., and Cannon, J. G. (1985). Variation of *Neisseria gonorrhoeae* protein II among isolates from an outbreak caused by a single gonococcal strain. *Infect. Immunol.* 49, 250–252.
- Sparling, P. F., Cannon, J. G., and So, M. (1986). Phase and antigenic variation of pili and outer membrane protein II of *Neisseria gonorrhoeae*. *J. Infect. Dis.* 153, 196–201.
- Stein, D. C., Silver, L. E., Clark, V. L., and Young, F. E. (1983). Construction and characterization of a new shuttle vector, pLES2, capable of functioning in *Escherichia coli* and *Neisseria gonorrhoeae*. *Gene* 25, 241–247.
- Stern, A., Nickel, P., Meyer, T. F., and So, M. (1984). Opacity determinants of *Neisseria gonorrhoeae*: gene expression and chromosomal linkage to the gonococcal pilus gene. *Cell* 37, 447–456.
- Stern, A., Brown, M., Nickel, P., and Meyer, T. F. (1986). Opacity genes in *Neisseria gonorrhoeae*: control of phase and antigenic variation. *Cell* 47, 61–71.
- Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E., and Inouye, M. (1966). Frameshift mutations and the genetic code. *Cold Spring Harbor Symp. Quant. Biol.* 31, 77–84.
- Streisinger, G., and Owen, J. (1985). Mechanisms of spontaneous and induced frameshift mutation in bacteriophage T4. *Genetics* 109, 633–659.
- Struhl, K. (1985). A rapid method for creating recombinant DNA molecules. *Biotechniques* 3, 425.
- Swanson, J. (1982). Colony opacity and protein II compositions of gonococci. *Infect. Immunol.* 37, 359–368.
- Swanson, J., and Barrera, O. (1983). Immunological characteristics of gonococcal outer membrane protein II assessed by immunoprecipitation, immuno-blotting and coagglutination. *J. Exp. Med.* 157, 1405–1420.
- Torriani, A. (1960). Influence of inorganic phosphate in the formation of phosphatases by *Escherichia coli*. *Biochim. Biophys. Acta.* 38, 460–469.
- van der Ley, P. (1988). Three copies of a single protein II-encoding sequence in the genome of *Neisseria gonorrhoeae* JS3: evidence for gene conversion and gene duplication. *Mol. Microbiol.* 2, 797–806.
- Vieira, J., and Messing, J. (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* 19, 259–268.
- von Heijne, G. (1985). Signal sequences. The limits of variation. *J. Mol. Biol.* 184, 99–105.
- Wells, R. D. (1988). Unusual DNA structures. *J. Biol. Chem.* 263, 1095–1098.