Shared Themes of Antigenic Variation and Virulence in Bacterial, Protozoal, and Fungal Infections

KIRK W. DEITSCH,1 E. RICHARD MOXON,2 AND THOMAS E. WELLEMS1*

Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892-0425, and Institute of Molecular Medicine and University of Oxford Department of Paediatrics, John Radcliffe Hospital, Headington, Oxford OX3 9DU, United Kingdom

INTRODUCTION

The success of pathogenic microbes depends on their ability to colonize host tissues and counter host defense mechanisms. The microbes must also maintain infectivity through transmission, thereby ensuring their long-term survival. The occurrence of disease may therefore be an incidental or accidental matter, with morbidity and mortality reflecting the more aggressive strategies of microbes to perpetuate their genes within and between hosts. The pathogenicity or virulence of microbes reflects their dynamic interactions with host tissues and the adaptive responses that enable them to escape host defenses and maintain infection. All of us are carriers of organisms that reflect their dynamic interactions with host tissues and the environments and evasion of host clearance mechanisms.

PHENOTYPIC DIVERSITY, IMMUNE SYSTEM EVASION, AND VIRULENCE PROPERTIES OF PATHOGENS

Surface Glycoproteins of Trypanosomes Promote Persistent Infection, Resulting in Sleeping Sickness

P. falciparum Infection Causes Disease of Variable Severity and Provokes Only Partial Immunity in Infected Hosts

Malaria Diagnoses of Severe Malaria Often Reflect Strain Diversity

Surface Molecules of P. falciparum-Infected Erythrocytes Are Antigenically Variant and Promote Adherence in Different Tissues

Phenotypic Diversity Is Characteristic of Many Bacterial Pathogens

Antigenic Variation of N. meningitidis and N. gonorrhoeae Is Extensive and Defines Host Cell Tropism

LPS and CP Are Major Determinants of H. influenzae Virulence

Major Outer Membrane Proteins Play a Principal Role in the Antigenic Variation of Borrelia spp.

Differential Expression of Developmentally Regulated Genes Appears To Be Involved in Phase Variation by C. albicans

Changes in Host Receptors for Pathogens Can Influence Their Adherence and Tissue Invasion

The Capacity for Rapid Variation in Surface Molecules Is a Common Evolutionary Feature of a Wide Spectrum of Pathogens

GENETIC MECHANISMS OF VARIATION IN BACTERIAL AND PROTOZOAAL PATHOGENS

High Rates of Mutation Can Lead to Greater Genetic Variability and Increased Ability To Adjust to Unstable Environments

Mechanisms of Phase Variation Govern the Expression of Many Genes Encoding Surface Molecules

Gene Conversion and Intra- and Intergenomic Recombination Generate Extensive Repertoires of Antigenically Variant Molecules

Point Mutations Are Frequently Involved in Immune Evasion by Pathogens

Mechanisms of Antigenic Variation Include Other Active (and Uncharacterized) Mechanisms of Switching among Gene Expression Sites

PERSPECTIVE: VARIATION IN PATHOGENIC MICROBES IS MEDITATED BY HYPERMUTABLE LOCI

REFERENCES

* Corresponding author. Mailing address: Bldg. 4, Rm. B1-34, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892-0425. Phone: (301) 496-4021. Fax: (301) 402-0079. E-mail: tew@helix.nih.gov.

protozoal, bacterial, and fungal infections in humans. We review the molecular determinants of the corresponding host-parasite interactions and finally explore common characteristics of hypermutable genetic loci that mediate the high rates of phenotypic variation that facilitate adaptation to host microenvironments and evasion of host clearance mechanisms.

PHENOTYPIC DIVERSITY, IMMUNE SYSTEM EVASION, AND VIRULENCE PROPERTIES OF PATHOGENS

Surface Glycoproteins of Trypanosomes Promote Persistent Infection, Resulting in Sleeping Sickness

One of the best-studied examples of antigenic variation and its role in infection is that of pathogenic African trypanosomes, the causative organisms of sleeping sickness. Trypanosoma brucei produces rising and falling waves of parasitemia by generating subpopulations that have antigenically different forms of a major variant-specific glycoprotein (VSG) at the cell surface (29, 44, 123, 140). The waves of parasitemia are a principal feature of a chronic infection that persists until the individual is either treated or dies of the disease (109). In the terminal
stages of sleeping sickness, invasion and proliferation within tissue and organ systems overwhelm the host defenses. Invasion of the central nervous system produces an array of symptoms including mental disorders, slurred speech, epileptic fits, and paralysis (63, 109). Invasion of the cardiac system can cause congestive heart failure (63). The potential role of the antigenically variant forms in host cell recognition and tissue invasion remains to be determined.

The antigenic variation of African trypanosomes is mediated by VSGs that coat the organism and are the major antigens recognized by the immune system (43). Individual trypanosomes generally express a single VSG. In the course of infection, expanding numbers of parasites give rise to subpopulations that have switched to different VSGs and therefore are able to escape the immediate impact of the antibody response to the parental population. These subpopulations are subsequently recognized by the immune system, but continued generation of novel VSGs prevents clearance of the infection (reviewed in reference 123). This capacity for switching among different VSGs is thought to lead to eventual exhaustion of host defenses in the terminal phases of the disease. Although frequencies of switching in VSG expression probably vary greatly, several studies have measured rates varying from $10^{-2}$ to $10^{-7}$ (reviewed in references 44 and 45).

Individual trypanosome parasites are capable of expressing more than 100 different VSGs during the course of an infection (34). The genes that encode these proteins are members of a diverse multigene family and are thought to be transcribed exclusively from one of a limited number of telomeric expression sites in the genome (106, 118). The expression site comprises a single, large transcription unit (~40 kb) that includes the sequences from at least eight different expression site-associated genes (ESAGs) (125). The polycistrionic transcripts are rapidly processed into individual mRNAs through the processes of polyadenylation and trans-splicing (77, 85). One ESAG has been identified as encoding an adenylate cyclase of unknown function (122). Two other ESAGs (ESAG 6 and 7) encode subunits of a transferrin receptor (149, 163, 164), a protein thought to be necessary for uptake by the parasite of host transferrin. This receptor is found within the flagellar pocket of the parasite, a place largely sheltered from the host immune system but still vulnerable to antibody attack. Some evidence indicates that VSGs may be evolutionarily related to these transferrin subunits (35). Extensive conserved regions within these subunits are thought to be necessary to maintain the proper function of the transferrin receptor (28). However, there is a stretch of 32 nucleotides in which up to 12 nucleotides can differ, and most of these point mutations result in amino acid substitutions. This tightly contained region of heterogeneity is thought to be restricted to an immunodominant surface loop of the protein. It may thus serve to protect the invariant binding portion of the receptor from immune system attack or may provide superior adaptation of particular transferrin receptor forms to different mammalian species within the range of T. brucei subspecies (29, 164). The products of most of the other ESAGs remain uncharacterized (29, 45).

VSG sequences typically comprise 400 to 500 amino acids and show dramatic levels of polymorphism in primary structure. The three-dimensional structure, however, appears to be remarkably conserved in spite of this polymorphism (27). The primary sequences show greater conservation at the carboxy terminus, where the VSGs are anchored at the parasite membrane by glycosylphosphatidylinositol (GPI) linkage to ethanolamine (55). VSGs in solution are present as dimers, possibly attached through disulfide bonds (9). The VSG coat on the trypanosome surface is a closely packed monolayer, so that only a small part of the VSG molecule is accessible to antibodies on the cell surface (44). In addition to antigenic variation, two other functions have been ascribed to this coat. First, it probably shields other parasite membrane molecules and permeases from immune attack. Second, the VSGs of the coat are thought to interfere with host defenses by inhibiting phagocytosis (173).

**P. falciparum** Infection Causes Disease of Variable Severity and Provokes Only Partial Immunity in Infected Hosts

Infections caused by *Plasmodium falciparum* are persistent, recurrent, and characterized by highly variable patterns of disease. Immunity against malaria is slow to develop, and only after extensive and repeated infections does a form of immunity arise that is protective against severe disease, albeit incomplete and incapable of eradicating the parasites. In areas of high parasite transmission, severe complications of the disease and mortality are usually highest in children younger than 5 years (39, 59, 189). These severe forms of disease can include coma, severe anemia, and respiratory failure. Children in areas of high parasite transmission who live beyond the age of 5 years generally have adequate immunity to control the disease, a state that persists through adulthood in the presence of continual sporozoite inoculation from infective mosquitoes (11).

The immunological factors responsible for disease-controlling immunity remain uncertain. Specific antibodies directed against *Plasmodium* antigens appear to play a prominent role (30, 40), and various cytokines have been correlated with protection (88). Interestingly, AIDS, which can dramatically exacerbate the course of tuberculosis and opportunistic infections, does not seem to predispose individuals to more severe malaria. In a number of studies in Africa, no increase was seen in parasitemia levels, malaria frequency, or rates of cerebral malaria in human immunodeficiency virus (HIV)-positive individuals compared to those not carrying the virus (reviewed in reference 32). The one exception appears to be an increased risk of mortality among infants born to mothers coinfected with *Plasmodium* and HIV (26). Why individuals with AIDS are not in general predisposed to the severe complications of malaria is not clear; however, a number of speculative explanations have been proposed. These include earlier complications in and more rapid demise of individuals with AIDS from pathogens other than *P. falciparum*, induction by malarial antigens of a T-cell-independent immunoglobulin G response (150), and interferon-independent macrophage destruction of *P. falciparum*-infected erythrocytes (32). The depression of T-cell activation and reduced gamma interferon activation of macrophages in individuals with AIDS may thus not dramatically affect immunity to malaria. Whatever host factors determine the control of severe malaria, the humoral and cellular responses involved appear to have a balance contrasting dramatically to that required for the control of other pathogens so devastating in AIDS patients.

**Miniepidemics of Severe Malaria Often Reflect Strain Diversity**

Some clues to the complexities of malaria infections are provided by studies showing that the incidence of childhood malaria in regions of endemic infection is highly variable and that the incidence of severe disease is not evenly distributed but, rather, shows periodicity and space-time clustering (41, 156). Therefore, there can be marked year-to-year variation in the occurrence of severe cases, which tend to group into localized miniepidemics within areas of broad endemicity. Seasonal
increases in the mosquito vector populations are not thought to explain adequately the differences in rates of severe disease. Instead, the parasites responsible for severe illness themselves appear to differ in virulence from the parasites that cause the more common, less severe infections. An explanation for the occurrence of such miniepidemics is that they arise from changes in the cytoadherence and antigenic properties of the parasitized erythrocytes (50).

The adherence properties of parasites are implicated in a number of the severe manifestations of malaria, including cerebral malaria, while changes in antigenicity allow the parasites to persist and produce repeated infections in individuals. Marsh and Howard (99) showed that parasite strains in an African village can exhibit extremely diverse antigenic determinants on the surface of infected erythrocytes. In that study, convalescent-phase sera from infected children who received antimarial treatment always reacted with erythrocytes infected by the homologous \textit{P. falciparum} strains but usually did not react with erythrocytes infected by heterologous strains. In contrast, antisera from native adults generally reacted with all erythrocytes infected by the different parasite strains, indicating that the adults had developed immunity to a wide range of antigenically diverse erythrocyte surface antigens.

These data highlight the extreme levels of phenotypic variation characteristic of \textit{P. falciparum} populations and the key role that this diversity plays in the epidemiology, pathogenesis, and immunobiology of malaria. The observations also suggest an explanation for the susceptibility of children to fatal malaria, in contrast to the relative resistance of chronically infected adults to severe disease. The antigenic variation that enables the parasites to avoid elimination by the host immune response leads to the generation of highly diverse parasite populations in regions of high malaria transmission. Repeated exposure to these populations is necessary for children to develop the disease-controlling response and partial immunity against the many different strains. This is particularly important for strains that are especially virulent and responsible for miniepidemics of severe and fatal disease. Adults with a history of extensive exposure to phenotypically variant parasites will have acquired a sufficiently broad immune status to limit their susceptibility to such outbreaks of severe disease. These patterns of age-related susceptibility may also be influenced by different mosquito inoculation rates in malarial regions (172).

**Surface Molecules of \textit{P. falciparum}-Infected Erythrocytes Are Antigenically Variant and Promote Adherence in Different Tissues**

Mature-form \textit{P. falciparum}-infected erythrocytes avoid clearance from the circulation by being sequestered in tissue microvessels away from the action of the spleen (20, 21, 107). The presence of these sequestered forms in the brain, lungs, kidneys, liver, etc., is thought to be the major determinant of severe malaria. Antigenically variant populations of parasites may have different sequestration profiles and may thereby influence disease severity and organ involvement. Continuous generation of these diverse parasite populations leads to persistent infections that are characterized by waves of parasitemia and variable clinical courses that are the hallmark of malaria. Indeed, particular sequestration phenotypes are thought to be involved in outbreaks of severe disease and cerebral malaria (21, 184).

Variable forms of proteins at the surface of infected erythrocytes mediate both the adherence and antigenic properties of parasite-infected erythrocytes (16, 31, 75, 93). These proteins can change at rates reported to be as high as 2% per generation (138), and this can lead in a short time to parasite populations that are highly diverse both in their surface antigens and in their ability to bind to different cell surface molecules, including postcapillary endothelium and uninfected erythrocytes (24).

Molecules responsible for the adherence and antigenic variation of \textit{P. falciparum}-infected erythrocytes are termed \textit{PIEMP1} and are encoded by members of the \textit{var} family of genes (18, 155, 170). Individual parasites exclusively express a limited number of \textit{var} genes as 200- to 350-kDa proteins from a diverse repertoire of 50 to 150 copies (accounting for ca. 2 to 6% of the genomic DNA). Switches in \textit{var} gene expression are thus capable of generating a large variety of antigenic forms. This diversity of binding phenotypes appears to be an important determinant of the different disease-causing phenotypes of parasite isolates.

The products of the \textit{P. falciparum} \textit{var} genes are transported to the surface of infected erythrocytes and anchored in electron-dense membrane structures termed knobs (reviewed in reference 51). The predicted amino acid sequences of these proteins show a large, variable extracellular segment with domains having receptor-binding features, a transmembrane sequence, and a terminal segment that serves as a submembrane anchor. The extracellular portion of these proteins contains two to four regions that resemble cysteine-rich domains of certain \textit{Plasmodium} molecules involved in erythrocyte invasion, including the Duffy antigen binding proteins (4) and EBA175 (154); these regions have therefore been termed Duffy binding-like domains (129). The extracellular regions near the amino terminus of the protein also contain a cysteine-rich interdomain region, which, with the first Duffy binding-like domain, may form a head structure important to binding specificity. Additional studies have shown that purified \textit{PIEMP1} proteins bind thrombospondin and the endothelial receptors CD-36 and intercellular cell adhesion molecule type 1 and have presented evidence for binding to distinct domains of the molecule (17, 58). Sequence analysis and gene hybridization studies have shown dramatically different complements of \textit{var} genes among parasites (170). This overall diversity is indicative of vast numbers of genetically diverse strains over the geographical range of \textit{P. falciparum} malaria. Generation of diverse subpopulations from erythrocyte-stage recombination events and frequent sexual recombination in the course of mosquito transmission (10, 71, 185) appear to be potent sources of strain diversity and suggest one reason why continual exposure to a variety of parasite strains is evidently required for disease-controlling immunity among individuals living in malarial regions. These sources of diversity may also explain why sterilizing immunity does not develop in adults native to regions of high \textit{P. falciparum} transmission; these individuals generally achieve at best a balanced state of persistent infection wherein parasitemia is controlled and severe disease does not develop.

Pregnancy has long been associated with a notable decline in malaria resistance, returning otherwise semi-immune women to a susceptible state during which they can develop severe malaria with devastating consequences (102, 103, 107, 117). This susceptibility is thought to involve the vulnerability of the placenta to parasite sequestration, perhaps by exposure of surfaces that carry a high density of such receptors as chondroitin sulfate A (CSA) (57). Particular forms of \textit{PIEMP1} presumably mediate the binding to CSA by erythrocytes infected with particular \textit{P. falciparum} strains (137, 139). The placenta may therefore make the fetus and mother especially suscepti-
Piliated meningococci and gonococci have been observed to adhere to human cells but not to many animal cells (humans are the only natural reservoir of these bacteria). We note that other molecules, for example PilC and the outer membrane proteins Opa and Opc, play an important role in adherence interactions. The Opa (opacity) proteins, found in both \( \text{N. meningitidis} \) and \( \text{N. gonorrhoeae} \) (105, 147), and Opc, expressed by many strains of \( \text{N. meningitidis} \) (3, 147), are known to play an important role in virulence (22, 56, 98, 178, 179), since their expression mediates host cell invasion (182). Individual proteins of the Opa and Opc groups have approximate molecular masses of 27 kDa and are characterized by relatively basic isoelectric points (3, 121). The high variability of their surface expression (3, 182) involves both variation in promoter activity affecting individual protein levels and frame shifts, which, at the translational level, result in reversible on/off switching. There are as many as 11 related \( \text{N. gonorrhoeae} \) \( \text{opa} \) genes and 3 or 4 \( \text{N. meningitidis} \) \( \text{opa} \) genes at different chromosomal locations (5, 23). The encoded Opa sequences show close similarity except for small semivariable regions near the \( \text{N} \) termini and two centrally located hypervariable regions. Opc in \( \text{N. meningitidis} \) is encoded by a single gene and shows only limited homology at the amino acid level to the Opa proteins (121), but it is variably expressed through a mechanism which modulates transcription.

In \( \text{N. meningitidis} \), Opa and Opc proteins are located on the surface of the bacteria under the capsule, so that the major binding properties of these proteins may be masked in capsule-positive organisms or bacteria which have sialylated lipopolysaccharides (LPS) (181, 182). For capsule-negative bacteria, however, Opa proteins may mediate bacterial attachment to epithelial human cells in the early stages of disease (22, 98) and Opc may mediate attachment to both endothelial and epithelial cells (182, 183). Frequent isolation of capsule-deficient strains of \( \text{N. meningitidis} \) from the nasopharynx suggests that these proteins play an important role in interactions with mucosal epithelial cells (37).

LPS and CP Are Major Determinants of \( \text{H. influenzae} \) Virulence

The surfaces of pathogenic bacteria display sugar and lipid molecules that are accessible to the immune system and have important roles in the virulence and course of infection. Both capsular polysaccharide (CP) and LPS, for example, are major components of the cell surface of gram-negative bacteria and are known to be critical virulence determinants of \( \text{Haemophilus influenzae} \), a major pathogen that causes meningitis and respiratory tract infections in children. The absolute production of CP has also been correlated with virulent infections (74, 86). Bacteria that produce no CP are less virulent but have the advantage that they are not recognized by host antibodies to capsule antigen (73). The absence of CP may also give the bacteria a greater ability to attach to or invade host epithelial cells (165).

In conjunction with the polysaccharides of encapsulated strains, LPS impedes host clearance mediated by complement and phagocytes (113, 116). Antigenic variation in LPS is generated by changes in the expression of enzymes involved in its synthetic pathway. Populations of bacteria that display different surface molecules (2, 42, 100) have been shown to result from the loss or gain of LPS core saccharide structures (100, 186).
Major Outer Membrane Proteins Play a Principal Role in the Antigenic Variation of Borrelia spp.

The Borrelia hermsii spirochete is responsible for a relapsing fever that is characterized by recurrent febrile attacks separated by asymptomatic intervals. The symptomatic episodes are marked by antigenically distinct waves of Borrelia in the bloodstream, each of which requires a specific antibody response for clearance. An abundant outer membrane protein, termed the variable major protein (VMP), is the primary variant antigen recognized in infection (15). The serotype specificity of each bacterium is thought to result from a single form of this protein that is expressed from one of a family of VMP genes in the organism (12, 15, 136). Switches among the expressed members of this family of genes allow small subpopulations to continually arise during infection, providing a reservoir of new serotypes that allow the infection to stay ahead of the immune response. A single B. hermsii bacterium has been shown to produce up to 40 different serotypes (136), and switching rates among different VMP genes have been estimated to range from \(10^{-6}\) to \(10^{-5}\) (168).

VMPs are generally classified by molecular mass into large (37- to 40-kDa) and small (19- to 22-kDa) categories (14, 136) and are encoded by genes located on multicopy linear plasmids (13, 81, 130). The small and large VMPs share \(\approx 40\%\) identity; members of the small class of B. hermsii VMPs have an overall identity of 70 to 80% (135, 136). Divergence among different VMPs occurs predominantly in the internal regions of the primary structure, whereas sequences at the amino and carboxy termini are largely conserved.

The VMPs are thought to play an important role in determining the tissue and organ localization of Borrelia spp., probably through a direct effect on cell-cell interactions (33). In the rodent pathogen Borrelia turicatae, subpopulations expressing one VMP were able to colonize the brains of infected mice whereas subpopulations expressing different VMPs failed to do so. VMP types therefore altered both the tissue tropism and immunogenic profile of these organisms.

Proteins similar to VMPs are found on the surface of Borrelia burgdorferi, the causative agent of Lyme disease (36). These B. burgdorferi surface proteins, collectively referred to as OspC, share sequence homology and antigenic cross-reactivity with the small B. hermsii VMPs, suggesting that families of variable surface proteins may be a common feature of different Borrelia species.

Differential Expression of Developmentally Regulated Genes Appears To Be Involved in Phase Variation by C. albicans

Candida albicans is a fungal species that often occurs in the oral cavity, vulvovaginal region, and anorectal regions of healthy persons without causing pathogenic effects (120, 159). It can, however, produce overt local or systemic disease. A common distressing problem is symptomatic vulvovaginal inflammation in women, sometimes with no obvious predisposing condition. More seriously, in immunocompromised individuals and patients with predisposing conditions such as diabetes and intravenous drug abuse, C. albicans can enter and spread within tissues (120), causing life-threatening disseminated or invasive infection. The pathogenic potential of this organism seems to be mediated by its adherence properties and phenotypes that affect such virulence traits as the bud-to-hypha transition (7), sensitivity to antifungal drugs (158), cell surface antigenicity (6, 8), epithelial tropism (79), and in vitro sensitivity to neutrophils and in vitro oxidants (82).

Some of these traits may be controlled by switch events that produce coordinated changes in multiple virulence characteristics. Switching events can alter the cell surface antigenicity and adherence properties of C. albicans (6, 8, 79), consistent with a role for variable surface molecules in tissue invasion and localization. Entire sets of phenotypically important genes may be coordinately switched on or off, thereby driving phase variation of complex phenotypes and virulence parameters (126, 161). There is also evidence for differential expression of developmentally regulated genes in the phase variation of C. albicans phenotypes (161). In the switch from white to opaque colony morphology, a number of opaque-specific antigens as well as opaque-specific mRNAs have been identified (6–8, 161). Both activation and deactivation of phase-specific genes have been found in the white-to-opaque transition (160). These phase-specific genes of C. albicans are dispersed throughout the genome, but no genomic rearrangements, transpositions, or mutations have yet been correlated with switches in expression (110, 111). The genes may instead be regulated by one or more trans-acting factors encoded by a master regulatory site. One hypothesis is that this regulatory site undergoes a reorganization or alteration that is responsible for developmental regulation of the phase-specific genes downstream in the metabolic pathway (161). This proposal is supported by gel retardation assays of white- and opaque-specific gene control regions after incubation with cell extracts containing DNA binding proteins (157, 162). Results from these assays are consistent with the presence of regulatory proteins that bind to the control regions in a phase-specific manner. A master regulatory site, however, remains to be identified.

Changes in Host Receptors for Pathogens Can Influence Their Adherence and Tissue Invasion

On the host side of the host-parasite interaction, changes in cell surface or tissue properties can produce effects on the virulence of an infection that are as dramatic as changes in the pathogens themselves. One example is found in severe disease due to Streptococcus pneumoniae. This gram-positive bacterium colonizes approximately 1 in 2.5 humans, but only 1 in 200 go on to develop overt pneumonia, otitis media, or septicemia. In the human lung, both the opaque and the disease-causing transparent pneumococci adhere to the surface of lung cells, and there appears to be no advantage of one bacterial type over the other in colonization of healthy lung tissue (48). However, when a viral infection is present in the respiratory tract, induction of host cytokines activates the receptor for platelet-activating factor on the surface of lung epithelial cells (47, 65). The transparent pneumococci can attach to this receptor and invade the systemic circulation via pulmonary capillaries or lymphatic vessels (131). This probably underlies the predisposition of virally infected patients to the complications of pneumococcal pneumonia, septicemia, and meningitis.

A second example of host tissue changes leading to severe disease is found in the malaria of pregnancy. Placental sequestration sites in women with previous disease-controlling immunity provide a new hold for Plasmodium falciparum infection that can rapidly lead to severe disease, threatening the lives of both mother and fetus. Chondroitin sulfate A is reported to be among the molecules at these sequestration sites (57).

The Capacity for Rapid Variation in Surface Molecules Is a Common Evolutionary Feature of a Wide Spectrum of Pathogens

Table 1 lists several general characteristics of variable surface molecules critical to the survival and propagation of pathogens discussed in this review. These surface molecules are generally characterized by regions of dramatic variability in
primary structure, and they tend to be encoded by exclusively expressed members of large families of genes which have multiple variant forms. Regions of the molecules that are conserved often are found where the proteins are anchored within the surface membrane or otherwise not directly exposed to the external environment and the immune system of the host. High levels of mutability serve to produce the great numbers of different antigenic forms. Posttranslational modifications can also be a potent source of immune system evasion, because surface antigenicity is often dramatically affected by additions such as polysaccharide groups.

In addition to evasion of attack by the immune system, antigenically variable surface molecules often play other roles in tissue invasion or in countering of other host defenses. Accumulating evidence indicates that the major variable surface antigen of *P. falciparum*-infected erythrocytes is responsible for sequestering the parasites away from the clearing action of the spleen (90, 107). The variable surface antigens of African trypanosomes (173) and polysaccharide groups covering the surface of a number of pathogenic bacterial species are thought to interfere with phagocyte binding (113, 116). Other variable surface antigens of bacteria play roles in tropism and tissue-specific invasion. These include the proteins of bacterial fimbriae and pili, Opc and Opa, and the variable major proteins of *Borrelia*. Major proteins involved in localization and tissue invasion by *Candida* spp. remain to be identified.

### GENETIC MECHANISMS OF VARIATION IN BACTERIAL AND PROTOZOAL PATHOGENS

#### High Rates of Mutation Can Lead to Greater Genetic Variability and Increased Ability To Adjust to Unstable Environments

The rates at which different bacterial strains undergo mutations or acquire DNA from separate organisms can vary substantially. A number of strains of both *Escherichia coli* and *Salmonella* pathogens with unusually high mutation rates have been isolated (92). In most cases, these strains have arisen through mutations in methyl-directed mismatch repair, which leads to both an increase in the rate of mutation (hypermutable phenotype) and an increase in the occurrence of recombination between divergent sequences (108). Increased recombination can lead to a greater propensity to acquire DNA from separate organisms and even separate species (101, 133). Such “mutator” phenotypes are thought to provide the pathogen with the ability to generate rapid variation in an unstable environment so as to escape the host immune system or to develop resistance to drug therapy (92). While bacterial strains that exhibit a general hypermutable phenotype have the advantage of increased variation throughout their genome, several pathogens have evolved the specific ability to generate variation at particular loci involved in host pathogen interactions. This more refined form of antigenic variation can provide the pathogens with distinct selective advantages in their effort to evade their hosts and to evade the immune response.

### Mechanisms of Phase Variation Govern the Expression of Many Genes Encoding Surface Molecules

Phase variation refers to reversible on-off switching of surface proteins or carbohydrates (or epitopes of these molecules) at much higher frequencies than the average underlying genetic mutation rate. For bacteria, these phase variation rates are typically 1/100 to 1/1,000 per generation. We note, however, that the phase variation paradigm of on-off switching is often a simplified representation of complex processes falling under the broader description of antigenic variation. In this context, the combinatorial nature of molecular switching can include phase variations which affect different epitopes and act interdependently with other mutational or recombinatorial events at several levels.

In its classical form, on-off switching in phase variation results in the reversible loss or gain of specific structures from the cell surface. These switches are usually controlled at the level of transcription of the encoding genes. For example, phase variation of *E. coli* type 1 fimbiae has been shown to involve variable DNA methylation and a transcription regulator (153,
Salmonella spp. can switch flagellum type through site-specific inversion of a genetic element (1). Other examples include modification of the promoter regions that appear to play key roles in the variable expression of lipoproteins on the surface of Mycoplasma hyorhinis (190, 191) and the expression of the surface molecules C5a and M protein in Streptococcus pyogenes (91).

Control of fimbrial gene expression by H. influenzae and Bordetella pertussis and of variable lipoprotein expression in M. hyorhinis involves the addition or subtraction of base pairs or repeat units in promoter elements (reviewed in reference 115). These changes are thought to alter the interactions of RNA polymerase with other binding components of the transcription machinery (175, 188, 191). Such alterations lead to changes in gene expression and determine whether bacteria possess surface fimbriae.

Changes in the translational frame of the lic genes, encoding a protein necessary for LPS synthesis, are responsible for the variation of LPS structures at the surface of Haemophilus spp. (187). These translational alterations involve additions or deletions of CAAT repeat sequences in the lic2 open reading frame (69). Regulation of pilC expression also has been shown to involve the loss or gain of nucleotides in a homopolymeric cytosine tract of the open reading frame (78). In these examples, changes in the DNA repeats are thought to result from a mechanism of “slipped-strand mispairing” (95, 169). A similar mechanism involving the addition or deletion of CTCTT elements produces disruption or reestablishment of the opa open reading frame in N. meningitidis and N. gonorrhoeae, resulting in Opa-negative and Opa-positive bacteria (3) (Fig. 1).

Another mechanism of phase variation in H. influenzae is found for the CP. Production of these polysaccharides appears...
to be increased by amplification of genes involved in capsule biosynthesis (86). These genes are contained within a locus referred to as cap (74). Amplification appears to involve homologous recombination facilitated by the insertion element IS1016 flanking cap copies (87). The entire locus is a compound transposon that can expand or shrink with changes in the number of individual copies (up to five copies have been reported).

**Gene Conversion and Intra- and Intergenomic Recombination Generate Extensive Repertoires of Antigenically Variant Molecules**

Loss of expression of the *N. gonorrhoeae* pilin subunit gene (*pilE*) results frequently from deletions, especially in the promoter region affecting transcription (70, 151). The pilin-negative variants, however, can revert to pilus expression through intragenomic, *recA*-dependent recombination events in which homologous but not identical sequences from any of several other loci (about 20 are known) are transferred by nonreciprocal gene conversion (64, 83, 84). Another important mechanism involves the conservative, reciprocal reinstatement of a functional pilin gene through intergenic null or telomeric expression of DNA (60). *Neisseria* spp. naturally transform at high rates by taking up DNA released from the lysis of neighboring *Neisseria* cells; the donor DNA usually is derived from members of the same species (53). However, in the case of *N. meningitidis*, interspecies transfer may occur, and it has been proposed that class II pilin sequences originated from the related *Neisseria lactamica*, which, like *N. meningitidis*, can exist as a commensal organism of the nasopharynx (148). Through these mechanisms of variation, the potential repertoire of different amino acid sequences in the pilins of *Neisseria* strains has been calculated to comprise more than a million variants (146).

Intragenomic recombination events involving gene conversion also give rise to frequent variant forms in other organisms (72, 123, 152). In trypanosomes, one form of VSG switching involves duplicative transpositions by which silent VSG gene cassettes are replicated from chromosome-internal regions into telomeric expression sites (72, 123) (Fig. 1). These transposition events depend upon sequence homology between the duplicated cassette and the gene in the expression site (123, 140). Gene cassettes with extensive similarity to active copies in telomeric sites appear to be readily duplicated in their entirety, so that the complete open reading frames of the cassettes are expressed. Partial conversion events that produce hybrid genes also occur, usually late in trypanosome infections, among VSG sequences that diverge in parts of their open reading frames. Sequences from different silent VSG genes have been recognized in these chimeric forms (141, 142, 171). They may thus act as an important source of novel VSGs that increase the antigenic repertoire and enable the parasite to evade advanced stages of immune system recognition (140).

Two mechanisms have been observed to produce switches in VMP gene expression by *Borrelia hermsii*. Different VMPs in this organism are encoded by genes located on 28- to 32-kb linear plasmids that have hairpin telomeres (13, 130). One mechanism of switching involves the unidirectional, duplicative transposition of a silent VMP gene cassette into an expression site by gene conversion, thus replacing the previously expressed VMP gene (81, 104, 130) (Fig. 1). The second mechanism of VMP gene switching involves an intraplasmid deletion event involving an active VMP gene and an inactive (or pseudogene) sequence at an adjacent location (135). The result is the expression of a new, fused form. Both mechanisms are active in serotype switches during *B. hermsii* infection (134).

**Point Mutations Are Frequently Involved in Immune Evasion by Pathogens**

Simple point mutations within the reading frame of genes encoding surface proteins can have profound effects on the antigenicity of pathogenic organisms (Fig. 1). Examples of the effects of such point mutations are numerous and well recognized in such viruses as influenza A virus and HIV. Among other examples in bacteria, antigenically variant forms of *Streptococcus pyogenes* result from point mutations in the gene encoding the M protein (66). More than 80 different serotypes resulting from such mutations have been defined. Activation of *Borrelia* VMP genes through intramolecular rearrangement is often followed by the introduction of multiple point mutations in the coding region of the activated gene (134). Other VMP pseudogenes are thought to serve as templates for these mutations.

Extensive point mutations have also been identified within the coding region of trypanosome VSGs after gene conversion events (97). In one study, the transposed gene copy from a telomere was found to differ from the original template copy in up to 35 positions (96). These mutations have been associated with duplication of silent genes from other telomeric locations, and they display a distinct strand bias. The point mutations may therefore result from the modified bases that are found within silent telomeric VSG copies but not within silent VSG cassettes located in internal chromosomal sites (96). One hypothesis to explain this bias is that the presence of methylated or otherwise modified DNA in silent telomeric VSG genes may affect the fidelity of the gene conversion mechanism, producing an elevated frequency of point mutations and generating heterogeneity in the expressed VSG proteins. Alternatively, observations that the mutations are found in the coding regions of the VSG gene, not in the surrounding transposed DNA, and that they tend to be clustered within a region of the VSG that is exposed on the outer surface of the VSG coat may indicate that these mutations arose not during but after transposition, in the months required for isolation of the population (45). The availability of new techniques with selective markers which allow immediate selection of VSG transposition events in transformed parasites should help resolve this question.

**Mechanisms of Antigenic Variation Include Other Active (and Uncharacterized) Mechanisms of Switching among Gene Expression Sites**

African trypanosomes may also generate antigenic variants by changing transcription from one telomeric expression site to another. Each trypanosome may contain as many as 25 to 50 telomeric expression sites (46), only 1 of which is active at a time, and switches in their activation therefore can lead to expression of different VSG and ESAG genes. The silencing of expression sites is developmentally regulated, because ribosomal promoters placed within silent expression sites are repressed in blood stage trypanosomes but active in insect forms (76). How VSG switching occurs is not understood. One hypothesis is that DNA modification plays an important role. Comparative restriction analysis of an expression site in its active and inactive states has provided evidence for an association of nucleotide methylation with inactivity (19, 124). In addition, an unusual DNA base, β-D-glucosyl-hydroxymethyluracil (referred to as J) (62), is found in trypanosome DNA and is enriched in telomeres (61). The presence of this base is reduced in activated VSG expression sites but is restored when
The expression site is silenced (61, 62). These forms of DNA modification, methylation or incorporation of J, may directly silence an active VSG site or may tighten the silencing of the site in a way that is reminiscent of the method by which methylated cytosine is thought to act in animal cells (25, 38, 54).

The mechanisms of expression site switching remain unknown for antigen variation in *Candida* spp. and *P. falciparum*. Transcription of the *P. falciparum* var genes can occur from many different expression sites throughout the genome (68, 170). These expression sites are found in internal regions of the chromosomes as well as in subtelomeric regions (68, 144, 170), unlike the expression sites of trypanosome VSG genes, which are thought to be exclusively telomeric (46, 174). Evidence indicates that multiple mechanisms of var control may exist: in one study (155), no restriction differences suggestive of rearrangement within a var cluster did correlate with a switch in the transcription of an expressed gene, implicating recombination as one source of diversity in the var repertoire.

**PERSPECTIVE: VARIATION IN PATHOGENIC MICROBES IS MEDIATED BY HYPERMUTABLE LOCI**

Evolutionary biologists have for some time taken a keen interest in the interactions of microbe and host and the ensuing conflicts which determine their coevolutionary trajectories. This mutuality has attracted such colorful metaphors as gene-for-gene arms races (49) or the Red Queen hypothesis (176). Microbes can produce overwhelming infections because of their relatively short generation time and because they have evolved powerful mechanisms for generating phenotypic diversity as an efficient strategy for adapting to rapidly responding immune system defenses and the broad range of polymorphisms characteristic of different host tissues. These mechanisms of variation may be of particular relevance in the successful spread of an infection through a host population. A particular microbial species may be subjected to evolutionary bottlenecks in the course of transmission from individual to individual or during translocation from one anatomical site to another within an individual host. Indeed, for some pathogenic bacteria, it has been shown that an entire population of infecting organisms can be derived from a single clone (114). In this context, efficient mechanisms for generating phenotypic diversity become crucial as the microbes are confronted with sudden and variable challenges to their survival. Our review provides some examples of the rich panoply of molecular mechanisms through which pathogenic microbes generate phenotypic diversity.

We wish to emphasize the distinction between the genetic mechanisms of phenotypic variation with which this review has been concerned and those of classical gene regulation, exemplified by such phenomena as catabolite repression or two-component signaling mechanisms. Although classical gene regulation can produce phenotypic change, the stochastic mechanisms of antigenic variation attributable to hypermutable loci offer an enormous range of diversity, enabling the pathogen populations to cope with the challenges of maintaining infection within the host. These hypermutable genetic sequences have been termed contingency loci to emphasize their role in facilitating rapid diversification while maintaining requisite constraints on the trade-offs implicit in rapid genetic change (115). Although many of these genetic mechanisms of phenotypic variation appear to be involved in evasion of host immunity, they also are found in organisms on which the impact of immune defenses is unknown or absent, such as the intestinal parasite *Giardia* (112, 119) and the free-living organism *Paramecium* (94).

The hypermutability of contingency loci and its potential to generate extensive variation is confined to a minority of nucleotide sequences within a genome. This localization has apparently evolved to produce a repertoire of variant molecules that modulate such properties as antigenicity, motility, chemotaxis, attachment to host cells, acquisition of nutrients, and sensitivity to antibiotics while avoiding the deleterious effects that high mutation rates would impose on other conserved housekeeping functions, e.g. cell cycle control, metabolism, signaling, and synthesis.

The molecular switching and variability provided by contingency genes in protozoal, bacterial, and fungal pathogens are important in determining whether a particular pathogen is cleared from the host, persists to cause a relatively benign infection, or produces severe or even fatal disease. This heterogeneity, driven by the contest between the host defenses and the pathogen, may be affected by changes in the susceptibility of host tissues through polymorphisms or alterations in the distribution and density of host cell surface molecules. In some diseases caused by pathogens discussed in this review, infected individuals acquire only partial immunity and often are subject to recrudescences and recurrent infections. Our potential to control such infections will be improved by a better understanding of the factors that determine the molecular basis of the mechanisms which underlie the production of the extensive diversity of molecules involved at the host-pathogen interface.

**REFERENCES**

VOL. 61, 1997 ANTIGENIC VARIATION AND VIRULENCE 291


