Microbial antigenic variation mediated by homologous DNA recombination

Cornelis Vink¹, Gloria Rudenko² & H. Steven Seifert³

¹Department of Pediatrics, Erasmus MC, Rotterdam, The Netherlands; ²Division of Cell and Molecular Biology, Imperial College-South Kensington, London, UK; and ³Department of Microbiology-Immunology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

Abstract

Pathogenic microorganisms employ numerous molecular strategies in order to delay or circumvent recognition by the immune system of their host. One of the most widely used strategies of immune evasion is antigenic variation, in which immunogenic molecules expressed on the surface of a microorganism are continuously modified. As a consequence, the host is forced to constantly adapt its humoral immune response against this pathogen. An antigenic change thus provides the microorganism with an opportunity to persist and/or replicate within the host (population) for an extended period of time or to effectively infect a previously infected host. In most cases, antigenic variation is caused by genetic processes that lead to the modification of the amino acid sequence of a particular antigen or to alterations in the expression of biosynthesis genes that induce changes in the expression of a variant antigen. Here, we will review antigenic variation systems that rely on homologous DNA recombination and that are found in a wide range of cellular, human pathogens, including bacteria (such as Neisseria spp., Borrelia spp., Treponema pallidum, and Mycoplasma spp.), fungi (such as Pneumocystis carinii) and parasites (such as the African trypanosome Trypanosoma brucei). Specifically, the various DNA recombination-based antigenic variation systems will be discussed with a focus on the employed mechanisms of recombination, the DNA substrates, and the enzymatic machinery involved.

Introduction

One of the most intriguing issues in medical microbiology is the interplay between the human host and colonizing microorganisms, which can be of bacterial, viral, fungal, or parasitic origin. While some of these microorganisms have a commensal or symbiotic relationship with their host, others may be envisaged as true pathogens which induce infections that can be harmful and even fatal to the host. On an evolutionary scale, microorganisms evolve relatively quickly and fitness is enhanced by microbial replication and spread, and not by host damage or death. It is reasonable to propose that most pathogens evolve in ways that limit the extent of their pathogenesis in order to increase the duration of the infection and maximize transmissibility.

One of the most effective and widespread strategies employed by pathogens in order to escape the human immune system is antigenic variation. This process can be described as a defined system that allows a pathogen to change its (surface) antigens that are presented to, and targeted by, the host’s adaptive immune system. As a consequence of antigenic variation, subpopulations of antigenically distinct organisms arise within a population that are temporarily not recognized by the primary adaptive immune response. This variability provides the pathogen with an extended window of opportunity to persist within a host or to infect a previously colonized host. It is also possible that antigenic variation provides functional changes to the microorganism that alters interactions with the host that are not dependent on an immune response. For example, antigenic variation of a microbial receptor could lead to changes in its adhesive properties, and it is often difficult to cleanly separate antigenic changes that lead to immune evasion from changes that alter the intrinsic function of the antigen.
While the term ‘antigenic variation’ is typically used to describe the capacity to express alternative forms of a particular antigen, it is also employed to address a process termed ‘phase variation’, in which the expression of an antigen alters between two states of expression. However, in both cases, antigenic variation can be based on either of two mechanisms: genetic or epigenetic mechanisms. Genetic mechanisms involve the direct modification of the DNA sequence of an open reading frame (ORF) or alteration of regulatory elements of an antigen-encoding gene by DNA inversion or changes in polynucleotide repeat number. While changes within the ORF can lead to amino acid changes in the expressed antigen, the modification of the regulatory sequences can result in altered expression levels of the antigen. In contrast to genetic mechanisms, the epigenetic mechanisms that underlie antigenic variation do not involve the modification of primary DNA sequences, but rather rely on processes such as methylation. In particular, the transcription of bacterial genes may be regulated by differential methylation that influences transcription factor binding. These epigenetic processes, however, have previously been reviewed (Deitsch et al., 2009) and will not be discussed here.

Among the genetic processes that govern antigenic variation in pathogens (including viruses), three main mechanisms can be distinguished: (1) a high spontaneous mutation frequency (of DNA or RNA), which is typically seen in viruses such as HIV-1 and influenza virus and is often called ‘antigenic drift’ (Malim & Emerman, 2001; Boni, 2008), (2) reassortment of genome parts or segments, as observed in viruses, which is referred to as ‘anti- genetic shift’, and (3) specific (and often reversible) modification of DNA sequences at defined genomic loci, as observed in bacterial and eukaryotic parasites.

This review will focus on the mechanism by which various human pathogens employ homologous DNA recombination to achieve antigenic variation. By means of homologous DNA recombination events, nonexpressed coding regions can be transferred to one or more distant genomic sites in which the newly introduced DNA sequences can become actively transcribed. The transferred DNA can consist of an entire antigen-encoding ORF, but can also comprise only a segment of such an ORF. Homologous DNA recombination can have two different outcomes: the DNA can be transferred unidirectionally (which is described as gene conversion) and the DNA can be exchanged reciprocally. Examples of both mechanisms are found among a wide range of human pathogens, from bacteria and fungi to parasites. Here, we will discuss the DNA recombination processes that are exploited by important human pathogens in order to vary their antigens and to evade the immune system. Specifically, the antigenic variation systems will be discussed as found in (1) bacteria belonging to the genera Neisseria, Borrelia, Treponema, and Mycoplasma, (2) the fungus Pneumocystis carinii, and (3) the parasite Trypanosoma brucei. The DNA recombination mechanisms that mediate antigenic variation in these organisms will be described, including the DNA substrates that are engaged in the recombination events and the enzymatic machinery involved.

Gene conversion–induced antigenic variation in prokaryotes

The genus Neisseria

The genus Neisseria is part of the family Neisseriaceae that also includes Kingella, Eikenella, Chromobacterium, and Vitreoscilla spp. Most of the Neisseria spp. are obligate human commensal organisms that live in the nasopharynx and rarely cause disease (Berger et al., 1984). The two pathogenic species, Neisseria gonorrhoeae (the gonococcus) and Neisseria meningitidis (the meningococcus), are also only found within humans, but have gained the ability to promote damage and are therefore considered to be true pathogens. It should be noted that the pathogenic Neisseria and commensal Neisseria share some virulence factors that enable the colonization of the host. Their pathogenicity most often results from damage because of the host innate immune system that is induced when the bacteria colonize inopportune anatomical sites, and is not mediated by active processes such as potent cytotoxins.

Neisseria gonorrhoeae is the causative agent of the sexually transmitted infection gonorrhea. There are an estimated 700 000 cases of gonorrhea in the United States each year (Anonymous, 2007) and an estimated 60 000 000 cases of gonorrhea each year worldwide. A majority of infected women are asymptomatic for disease, while a majority of infected men are symptomatic. The most serious outcomes of gonococcal infection are pelvic inflammatory disease, epididymitis, and disseminated gonococcal infections (reviewed in Ehret & Knapp, 1989).

Neisseria meningitidis exists predominantly as a commensal organism in the nasopharynx similar to the commensal Neisseria. In infants, the meningococcus is one of many bacteria that can infect the CNS to cause meningitis, and it is the primary cause of bacterial meningitis in teenagers and young adults. It is presently unknown exactly how the meningococcus transits from the nasopharynx to the meninges, but an increased probability for invasive disease to develop is dependent on a combination of host and bacterial factors (reviewed in Stephens, 2009).

There are three major antigenic variation systems in the pathogenic Neisseria: the pilus, the outer membrane
opacity proteins (Opa), and the lipooligosaccharide (reviewed in Kline et al., 2003). Opa and lipooligosaccharide variation is mediated by random changes in nucleotide repeats that alter gene expression, presumably by slipped strand mispairing during replication. Pilin antigenic variation, however, is mediated by a gene conversion system. There is no evidence that immunity to gonococcal infection leads to protection from reinfection. While there are probably many reasons for this lack of immunity, the high rate of antigenic variability of this organism clearly contributes to this immune evasion. It is less clear why the meningococcus undergoes antigenic variation because this organism behaves more like the commensal Neisseria in most people and because there is no evidence that antigenic variation contributes to invasive disease. It should be noted that the potential number of possible antigenic variants is much fewer in the meningococcus relative to the gonococcus (Stern & Meyer, 1987). There exists also a functional impact of antigenic variation, with the most obvious effects being the change in host cell receptors recognized by bacteria expressing different Opa proteins (Bos et al., 1999) and the creation of under-piliated and nonpiliated variants by way of pilin variation (Segal et al., 1985; Nassif et al., 1993) (see below). Regardless of the reasons, there are common mechanisms promoting antigenic variation in the pathogenic Neisseria. This section of the review will focus on the pilus antigenic variation system that is the only one mediated by a gene conversion mechanism. Almost all of the mechanistic studies on pilin variation were conducted in N. gonorrhoeae and it is assumed that the mechanisms used by N. meningitidis are similar, if not identical.

The pilus phase and antigenic variation systems

Neisseria pili

There is only one pilus expressed by the pathogenic Neisseria, which is a Type IV pilus (Swanson et al., 1971). Type IV pili are expressed by many Gram-negative species of bacteria, and the Neisseria Type IV pilus is a major virulence factor that is required for infection and participates in cell and tissue adhesion, twitching motility, and DNA transformation (reviewed in Forest & Tainer, 1997). In gonococci, but not meningococci, pilus expression leads to a characteristic colony morphology which is readily observable by stereomicroscope (Kellogg et al., 1963; Swanson et al., 1971). When gonococci transit to nonpiliated or under-piliated states (see phase variation below), the colony morphology changes (Fig. 1), and thus pilus-dependent colony morphology is a way to follow and quantify a subset of antigenic variation reactions. While all gonococcal and most meningococcal isolates

![Fig. 1. Results of pilin antigenic variation. A starting piliated variant can produce piliated (P+) antigenic variants, under-piliated (P+/-) antigenic colony morphology variants, or nonpiliated (P-) colony morphology variants. The colony morphologies are shown in the light micrographs from Swanson (1978). P+→P+ and P+→P- variants occur by gene conversion reactions between pilS copies and the pilE locus. P+→P- variation can occur by three distinct mechanisms: (a) gene conversion, (b) PilC variation, or (c) pilE deletion. Only (a) and (b) are reversible. Pili are shown as straight lines where a color change indicates an antigenic variation and/or phase variation event. S-pilin is shown as circles with the same color as the pilin.](image-url)
express antigenically variable Type IV pili designated Class I pili, certain lineages of meningococci and commensal Neisseria express related, nonvariable Type IV pili designated Class II pili. Class II pili are related to the Class I pili but are missing some of the variable portion of the protein (Aho et al., 1987; Cehovin et al., 2010). As both Class I and Class II pilus-expressing meningococci can colonize and cause invasive disease, it is puzzling why antigenic variation of Class I pili is maintained. It is not known whether both Class I and Class II pili have identical biological properties but all pathogenic Neisseria express either a Class I or Class II pilus (see phase variation below).

The pilin antigenic variation system

Pilin antigenic variation is mediated by a gene conversion system that transfers parts of the silent storage pilin gene copies that are located at silent loci into the expressed \( \text{pilE} \) locus (Fig. 2). In gonococci, there are four to five silent loci that in total contain about 18 silent copies (Meyer et al., 1982; Snodgrass et al., 1994). In meningococcal isolates that express variable Class I pili, there is one silent locus that contains five to six silent copies. However, meningococci that express Class II pili carry a truncated silent pilin gene locus, and the expressed pilin gene is found at a different locus on the chromosome. The silent pilin copies contain about 450 base pairs (bp) of potential coding information, and these genes are missing the promoter, ribosome binding site, and the first 35–43 codons and thus do not express a protein product (Figs 2 and 3). In Neisseria, these truncated genes are not designated pseudogenes, because they have a genetic function as sequence donors and are not evolving toward or away from a functional gene. However, in other organisms, silent donor genes are designated pseudogenes, as discussed below. The coding sequence of the expressed pilin gene, \( \text{pilE} \), is about 1.2 kilobase pairs (kb) long and has been separated into five broad regions of primary sequence types (Fig. 2) (Meyer et al., 1982; Howell-Adams et al., 1996).

When pilin antigenic variation occurs, a portion of a silent copy is copied into the \( \text{pilE} \) locus, but the silent copy remains unchanged (Fig. 3). The amount of new sequence transferred (i.e. the recombination tract) is variable and is functionally defined by the first and last nucleotide changes in the (silent) donor sequence that differ from the starting recipient \( \text{pilE} \) sequence. How far outside this tract the recombination occurs cannot always be determined unless there are other nucleotide differences between the donor and recipient copy. These recombination tracts are always bordered by either regions of microhomology or with more extensive regions of homology depending on the relatedness of the recombining genes (Howell-Adams & Seifert, 2000; Criss et al., 2005). The possibility for recombination to terminate anywhere where there are three or more shared nucleotides has discounted the minicassette model (Haas &}

![Fig. 2. Cartoon of the pilin loci of strain FA1090. Depicted in light blue are the five silent loci and the silent copies encoded in each. The \( \text{pilE} \) locus is shown in black (with the variable sequences highlighted in blue) and includes the single silent copy (the upstream silent locus) that is associated with the expressed gene. The conserved, noncoding Sma/Cla repeat is present at the 3′ end of each locus (black oval). In the middle is a detailed cartoon of the \( \text{pilE} \) gene with the constant region (C), semi-variable region (SV), cysteine region 1 (cys1), hypervariable loop (HV1), cysteine region 2 (cys2), and the hypervariable tail (HV2). Conserved DNA segments are shown in black, and variable ones are shown in blue.](image-url)
Meyer, 1986). This model described antigenic variation as using seven minicassettes of variable information to effect pilin antigenic variation (Haas & Meyer, 1986). As little as one nucleotide can change in an antigenic variation event, but the usual recombination tract is on the order of 12-120 bp during antigenic variation events that retain piliation. In contrast, when antigenic variation results in a piliated to nonpiliated phase variation event, the recombination tracts are longer (Criss et al., 2005). Infrequently, the entire silent copy can be transferred into the pilE locus. Inactivation of the mismatch correction system increased the frequency of pilin antigenic and phase variation and produces an increase in the average recombination tract length, suggesting that mismatch correction acts on heteroduplex DNA to limit the extent of DNA tract length during antigenic variation (Criss et al., 2010).

Pilus and colony morphology phase variation

Many bacterial pili are capable of undergoing ‘ON/OFF’ or ‘A/B’ reversible phase variation through the use of invertible DNA segments or differential methylation (reviewed in van der Woude & Baumler, 2004). While there is true ON/OFF pilus phase variation of expression in the pathogenic Neisseria, there is also a reversible state of pilus underexpression that occurs by overlapping mechanisms. As stated above, piliated gonococcal colonies show distinctive colony morphologies when grown on solid media (Kellogg et al., 1963), while under- or nonpiliated cells show different ‘nonpiliated’ colony morphologies (Fig. 1). These shifts in colony morphotype have been used as a surrogate measure for pilin antigenic variation because a majority of colony phase variants show an altered pilE sequence. However, there are several reasons why the transfer of new sequences into the pilE locus may result in a nonpiliated or under-piliated phase variant (Fig. 1).

First, some of the silent copies carry stop codons in their potential coding regions (Koomey et al., 1987; Criss et al., 2005). Transfer of the DNA containing the stop codon will produce a nonpiliated variant. Second, nonpiliated pilus phase variants can also be produced by the switch OFF of the pilus assembly protein PilC (Jonsson et al., 1991). There are two pilC alleles and each carries a string of cytosines in their coding sequence. The number of cytosines can be altered by a slipped strand mispairing mechanism and a phase OFF of either allele occurs when the number of cytosines alters the reading frame to introduce a termination codon. Piliated phase variants are spawned when either pilC allele gains or loses a cytosine and restores the proper reading frame. Interestingly, in the gonococcus, both pilC alleles appear to be functionally equivalent, while in meningococci the two different PilC proteins both allow piliation but differentially influence adhesion to host cells (Nassif et al., 1994). As the frequency of pilC phase variation is at least
10-fold lower than colony morphology changes owing to pilin antigenic variation, pilC variation usually does not alter the analysis of pilin antigenic variation through colony phase variation.

Third, certain combinations of variant pilin sequences can exist in the pilE locus that can interfere with pilus expression for unknown reasons, producing an under-piliated variant (Hagblom et al., 1985). It is likely that these under-piliated phase variants express pilins that cannot interact properly with the portions of the pilus assembly machinery that prevent pilus retraction. Many of these variants also produce a proteolytically processed form of pilin called S-pilin, where the processed subunit can be found as a soluble molecule in culture supernatants (Haas et al., 1987; Koomey et al., 1991; Long et al., 1998).

Fourth, nonrevertible, nonpiliated phase variants can be produced when the pilE gene is deleted, although this is not a true phase variation event, because these variants cannot revert at a high frequency (Segal et al., 1985; Swanson et al., 1985). It is likely that the ability to spawn both nonpiliated and under-piliated variants is important for pathogenesis, because all gonococcal isolates have this ability. Whether pilus phase variation is similarly important for meningococcal pathogenesis has not been directly tested. However, it has been shown that detachment of meningococci from colonies because of the changes in glycosylation enhances transit across epithelial barriers, suggesting that pilus phase variation may provide a similar function (Chamot-Rooke et al., 2011).

Measuring pilin antigenic variation

One of the major challenges in the study of pilin antigenic variation was finding a way to reliably quantify the process and assay the effect of mutations on the variation frequency. Based on the pioneering studies of Kellogg et al. (1963) to associate colony morphology and piliation, the frequency of piliated to nonpiliated colony morphology transitions has been widely used as a surrogate measure of pilin antigenic variation. However, this assay is difficult to perform reproducibly and can be influenced by the different growth rates of piliated versus nonpiliated variants. Some of the variability of this assay was reduced by the development of a kinetic colony morphology phase variation assay (Sechman et al., 2005), which measures the appearance of nonpiliated outgrowths from a piliated colony over time. While being more reproducible, this assay is still sensitive to growth rate and environmental or genetic factors that alter the growth rate.

The most quantitative and informative assay is the sequencing assay, which uses a strain with a lacI regulatable recA gene (recA6) (Seifert, 1997) to allow pilin antigenic variation to be initiated at a specific time (Criss et al., 2005). This assay measures the frequency of pilin antigenic variation, and the variant sequences obtained can also be used to determine whether qualitative changes have occurred such as a shift in the usual spectrum of donor silent copies used (Cahoon & Seifert, 2009; Helm & Seifert, 2010). Most studies have used a version of the colony morphology phase variation assay to measure the effect of mutations on pilin antigenic variation, but the pilE sequencing assay is clearly the most robust, albeit expensive, assay developed to date.

In standard laboratory growth conditions, pilin antigenic variation occurs at rate that produces a frequency of 4-12% progeny colonies with altered pilE sequences after 20 generations of growth (Criss et al., 2005; Rohrer et al., 2005; Helm & Seifert, 2009), and a subset of these variants can be colony phase variants because of the changes at pilE as opposed to pilC phase variants. An analysis of pilin antigenic variation frequency using a PCR-based assay showed that of the in vitro growth conditions tested, only iron limitation altered the frequency of pilin antigenic variation and resulted in an increased frequency (Serkin & Seifert, 1998). No mechanism for this iron limitation–dependent increase in variation frequency has been forthcoming even after a genetic screen was conducted to attempt to identify a gene responsible for the phenotype (Sechman et al., 2005). It is possible that one or more environments are encountered by the bacteria during infection where the available iron is limited, resulting in an increased frequency of variation.

Transformation and pilin antigenic variation

As the Neisseria are very efficient at natural transformation of Neisseria DNA, it was hypothesized that transformation could contribute to pilin antigenic variation. Two studies used DNase I in the growth medium or mutations to inhibit transformation and suggest that transfer of pilS sequences to pilE can occur between cells (Seifert et al., 1988; Gibbs et al., 1989). The conclusion that DNA transformation was the major route of DNA transfer during pilin antigenic variation was challenged by two studies using nonpiliated strains carrying nonsense mutations in pilE. In these studies, the frequency of nonpiliated to piliated phase variation was not altered by inhibiting transformation, suggesting that pilin antigenic variation occurs mainly by an intracellular route (Swanson et al., 1990; Zhang et al., 1992). It is presently accepted that most recombination events that produce pilin variation occur intracellularly and that perhaps a small subset of reactions can utilize transformation as a secondary mechanism.
Polyploidy and pilin antigenic variation

The gene conversion reactions that occur within the context of a bacterial chromosome by an intracellular route were hard to model to allow silent copies that are located upstream as well as downstream of the expression locus to both donate sequences at a high frequency. It was a reasonable assumption that two copies of the pilin sequences might be required to allow these high-frequency gene conversions. When the DNA content of gonococci, both monococci and diplococci, was measured, it was found that both gonococci and meningococci were diploid (Tobiason & Seifert, 2006, 2010), whereas *Neisseria lactamica* carried a single copy of its chromosome (Tobiason & Seifert, 2010). Although a direct link between polyploidy and pilin antigenic variation has not been made, it is intriguing that another antigenically variable organism, that is, *Borrelia hermsii*, has also been reported to be polyploid (Kitten & Barbour, 1992). Despite carrying at least two copies of the chromosome, the pathogenic *Neisseria* are genetically homozygous, as they are unable to stably carry two different alleles at the same chromosomal locus, even under strong selection (Tobiason & Seifert, 2010).

Proteins required for pilin antigenic variation

Over the past 30 years, most of the proteins required for antigenic variation have been identified mainly by genetic means. The roles of these proteins are inferred by the activities of orthologs expressed by other bacterial species, but only a few have been biochemically characterized to confirm these activities.

RecA and modifiers

The conclusion that pilin antigenic variation was mediated by homologous recombination was first made by Koomey and coworkers, who reported that pilin antigenic variation was lost in a gonococcal *recA* mutant (Koomey et al., 1987). This was the first report of an antigenic variation–deficient (AVD) mutant, which is a complete loss-of-function mutant. There are other mutations that lower the frequency of pilin antigenic variation that do not totally prevent gene conversion, and in this review, we will refer to these as producing antigenic variation intermediate (AVI) phenotypes.

The requirement for RecA was restored by complementation with *Escherichia coli* *recA*, which actually provided higher frequencies of pilin antigenic variation because the cloned *rec Ae* gene also carried the cotranscribed *recX Ec* gene (Stohl et al., 2002). This discovery led to the identification of *RecX Ec* as a negative regulator of *RecA Ec* during the SOS response (Stohl et al., 2003). A similar activity was subsequently described for *RecX Ng* which acts on *RecA Ng* to limit the extent of RecA polymerization onto single-stranded DNA (Gruenig et al., 2010). However, while an *E. coli* *recX* mutant has no obvious phenotype, a gonococcal *recX* mutant shows significantly reduced levels of pilin antigenic variation (Stohl & Seifert, 2001). These studies suggested that the length of the *RecA Ng* nucleoprotein filament must be controlled by *RecX Ng* to allow efficient pilin antigenic variation, but why this limitation of *RecA Ng* activity is important has not yet been determined.

The *rdgC* gene was identified as an AVI mutant in a genetic screen (Mehr et al., 2000). The *E. coli* ortholog of this gene was originally identified in a genetic screen that identified mutations that showed a growth defect in a *recA* background (Ryder et al., 1996). The *E. coli* protein was subsequently shown to limit RecA filament growth (Drees et al., 2006), and thus it is possible that RdgC and RecX act together or in parallel to limit RecA activity to promote more efficient pilin antigenic variation.

The role of recombination pathways in pilin variation

The conclusion that pilin antigen variation is mediated by homologous recombination processes was bolstered when a genetic screen for transposon-induced loss-of-function mutants identified the *recO* and *recQ* genes as being important for pilin antigenic variation (Mehr, 1998). These genes are part of the RecF homologous recombination and repair pathway of *E. coli*. As the pathogenic *Neisseria* do not contain a *recF* gene, this pathway was named the RecF-like pathway in these species. Subsequent mutational analyses demonstrated that other RecF-like pathway genes, that is, *recJ* and *recR*, but not *recN*, were involved in pilin antigenic variation (Skaar et al., 2002; Sechman et al., 2005). It should be noted that while the *recA*, *recO*, and *recR* mutants were AVD mutants, the *recJ* and *recQ* mutants were AVI in a kinetic colony phase variation assay (Sechman et al., 2005).

The aforementioned study by Mehr concluded that the RecBCD recombination pathway has no role in pilin antigenic variation (Mehr, 1998). In contrast, other studies have suggested that the RecBCD pathway does play a role in antigenic variation in *Neisseria* (Chaussee et al., 1999; Hill et al., 2007). The contrasting conclusions from these reports prompted a re-examination of the role of the RecBCD pathway in both the MS11 and FA1090 strain backgrounds using the sequencing assay described above (Helm & Seifert, 2009). There was no significant effect of either a *recB* or *recD* loss-of-function mutation in both
strain backgrounds on the frequency or products of pilin antigenic variation. It is likely that these different growth rates, coupled with assays that only detect a subset of potential recombination reactions, generated incomplete data that resulted in the inaccurate conclusion that the RecBCD pathway has a role in antigenic variation.

**Other trans-acting factors involved in pilin antigenic variation**

The most extensive genetic screen for pilin AVD or AVI phenotypes also identified a number of other genes (Sechman et al., 2005). One AVI mutant had a transposon inactivating the ruvB gene, which suggested that the RuvAB helicase may play a role in pilin antigenic variation. The RuvAB helicase participates in many recombination and repair pathways by promoting branch migration of Holliday junctions (West, 1994). After the RuvBC helicase acts, the Holliday junction is cut by endonuclease RuvC. Another helicase that can also catalyze branch migration of Holliday junctions is RecG, although there is no known endonuclease that acts with RecG (Briggs et al., 2004). Mutation of recG, ruvA, ruvB, or ruvC each individually produced an AVI phenotype, suggesting that both Holliday junction processing pathways were required for pilin antigenic variation (Sechman et al., 2006). Interestingly, double mutants made between recG and any of the ruvABC genes were synthetically lethal when RecA, RecO, and RecQ were expressed. Synthetic lethality escape mutants arose that had deleted the pilE gene or when a cis-acting transposon insertion that blocks pilin antigenic variation was present in the same strain and therefore could not undergo pilin antigenic variation. These results suggested that when each Holliday junction processing pathway is inactivated, pilin antigenic variation cannot proceed. However, in the absence of both pathways, a recombination intermediate may be formed, possibly between diploid chromosomes, that cannot be properly resolved. This intermediate may be responsible for the lethality. Nevertheless, there is presently no direct evidence for the existence of this putative intermediate.

Other transposon-generated AVD mutants that came out of that genetic screen included two that affected the expression of threonine biosynthesis genes and the identification of an ABC transporter that influences pilin antigenic variation (Sechman et al., 2005). There is no obvious explanation for why these genes produce a significant drop in pilin antigenic variation, but their isolation suggests that there may be general aspects of *Neisseria* physiology that influence pilin antigenic variation. The final trans-acting factor known to have a role in pilin antigenic variation is the Rep helicase (Kline & Seifert, 2005). The *E. coli* Rep protein is a 3′-5′ helicase that has been implicated in chromosomal replication fork progression (Lane & Denhardt, 1974). Because a gonococcal *rep* mutant shows an AVI phenotype, the Rep helicase may participate in a yet unknown step of pilin antigenic variation (Kline & Seifert, 2005).

**DNA sites and structures required for pilin antigenic variation**

**Conserved sites functioning within the pilin genes**

Several DNA elements have been implicated in having a specific role in pilin antigenic variation. One of these elements is the Sma/Cla repeat, which carries conserved SmaI and Clal restriction endonuclease sites and is found 3′ to all pilin loci (Fig. 2) (Haas et al., 1992). Deletion of the Sma/Cla sequence in strain MS11 produced a reduction in the appearance of pilin antigenic variants (Wainwright et al., 1994). However, this phenotype has not been observed in other strains.

A sequence that clearly has an influence on the gene conversion process is the conserved cys2 region within the pilin sequences (Fig. 2). Two studies indicated that the cys2 region has a yet unknown role in pilin antigenic variation and also suggested that the recipient pilE gene acts differently than donor pilS copies during gene conversion (Howell-Adams et al., 1996; Howell-Adams & Seifert, 1999).

**The guanine quadruplex**

Since the initial description of the gene conversion reactions that allow pilin antigenic variation (Hagblom et al., 1985), it has been considered likely that unique processes existed that can promote high-frequency gene conversion reactions within the context of a bacterial chromosome. The discovery of diploid chromosomes within the pathogenic *Neisseria* provides one aspect of bacterial physiology that helps explain the ability to generate nonreciprocal recombinants within a bacterial chromosome (Tobiason & Seifert, 2006). However, general and directed genetic approaches have hitherto only identified conserved recombination and repair enzymes. If any of the genes required for antigenic variation are essential (e.g. the replicative polymerase), they would not have been identified in a genetic screen. Moreover, many of the transposon-generated mutants carry mutations in genes that do not have an obvious role in pilin antigenic variation (Sechman et al., 2005). However, one of the transposon-generated mutants provided important insight into the process of antigenic variation. This mutant contained a single
transposon that was inserted upstream of the pilE gene. While this insertion did not interfere with pilin expression, it totally prevented recombination from any pilS copy into pilE (Sechman et al., 2005). Further transposon mapping of the pilE upstream region demonstrated that transposon insertions upstream of the AVD insertion had no effect on pilin antigenic variation. Insertions downstream from this site did lower the efficiency of gene conversion, but did not abolish recombination (Kline et al., 2007). Additionally, transposon insertions or point mutations within the pilE promoter that prevented pilin expression had no additional effect on pilin antigenic variation. This demonstrated that the transcriptional activity associated with the pilE locus does not play a direct role in pilin antigenic variation (Kline et al., 2007).

A directed genetic screen identified 12 GC base pairs in the region just upstream of the AVD transposon insertion that were required for pilin antigenic variation (Cahoon & Seifert, 2009). Site-directed mutagenesis conclusively showed that mutation of any of 11 of these 12 GC base pairs produced an AVI phenotype and mutation of the 12th (G3) produced a severe AVI phenotype (Fig. 4). Subsequently, it was shown that mutation of a 13th GC base pair (G0) in combination with the G3 mutation produced an AVD phenotype showing that G0 can partially substitute for the G3 residue only when the G3 base pair is mutated (Cahoon & Seifert, 2011). Mutation of the AT base pairs within the region containing the 12 GC base pairs had no effect on pilin antigenic variation, nor did mutation of bases directly outside this region. This identification of 12 GC base pairs arranged in four sets of three was consistent with a guanine quadruplex (G4)-forming sequence. Biophysical studies confirmed that the G-rich sequence forms a parallel G4 structure in vitro and that single base pair mutations that blocked pilin antigenic variation alter the structure. The NMR structure of the pilE G4 sequence has been solved confirming an all parallel structure with unique properties (V.V. Kuryavyi et al., paper under review). Replacement of the pilE G4 with other G4-forming sequences produces an AVD phenotype as does a switch of the G- and C-rich strand orientation in the chromosome (Cahoon & Seifert, 2009).

Treatment of N. gonorrhoeae with the toxic chemical N-methyl mesoporphyrin IX (NMM), which specifically interacts with G4 structures and not single- or double-stranded DNA, inhibited both pilus phase variation and pilin antigenic variation. Analysis of the variant progeny showed a shift in the spectrum of donor pilS copies used when NMM was present during pilin antigenic variation. NMM treatment also prevented single-stranded nicks from being produced on the G-rich strand (Cahoon & Seifert, 2009). The effect of NMM on pilin variation strongly supports the conclusion that the G-rich sequence forms a G4 structure that is necessary for one or more steps in pilin antigenic variation.

**Summary of Neisseria pilus antigenic variation**

The ability to conduct high-frequency gene conversion reactions from about 18 donor silent copies that are located both upstream and downstream of the recipient pilE locus must require the concerted action of proteins...
Antigenic variation through gene conversion in Borrelia spp.

Antigenic variation also plays a crucial role in the pathogenesis of Lyme borreliosis, which is predominantly caused by four members of the spirochete genus Borrelia, that is, Borrelia burgdorferi, Borrelia garinii, Borrelia afzelii, and Borrelia valaisiana. These pathogens are transmitted to humans through bites of hard-bodied ticks that belong to the genus Ixodes. During the early stages after transmission, a localized inflammation (erythema migrans) can develop at the site of the tick bite. Subsequently, the spirochetes can disseminate throughout the body, leading to dermatological, ocular, neurological, cardiological, and arthritic problems (Colucciello, 2001; Stere, 2001).

Recombination at the vlsE locus of B. burgdorferi

A significant body of information on how antigenic variation is employed by the spirochetes that cause Lyme disease has come from studies on B. burgdorferi. This species expresses a 35-kDa surface-exposed lipoprotein, termed VlsE, which undergoes extensive antigenic variation during infection of mammalian hosts (Zhang & Norris, 1998a; McDowell et al., 2002). The VlsE protein is encoded by the vlsE gene, which forms part of the so-called vls (variable major protein-like sequence) locus, and is located near the telomere of the linear plasmid lp28-1 (Zhang et al., 1997; Zhang & Norris, 1998a,b; Wang et al., 2001; Bankhead & Chaconas, 2007). The vlsE gene represents the single expression site (ES) of the vls locus and is flanked by a contiguous upstream array of 15 silent (unexpressed) vls ‘cassettes’ (vls2 to vls16) (Fig. 5a). These silent cassettes are highly homologous to the central, variable region of vlsE (90–96% identity on the DNA level), which is termed vls1. Both the silent cassettes and the vls1 segment within the vlsE gene are flanked by 17-bp direct repeats. During experimental infection of mice and rabbits with B. burgdorferi, sequence variation was found to occur in the vlsE gene through a series of gene conversion events between segments of any of the 15 silent ‘donor’ cassettes and the vls1 segment within the ES. These events occur within 4 days after inoculation of mice and carry on throughout the course of infection (Zhang & Norris, 1998b; Embers et al., 2007; Coutte et al., 2009). After 1 week of infection, approximately 50% of the spirochetes carry recombinant vlsE sequences. Four weeks after inoculation, the parental vlsE sequence can no longer be detected among the recovered bacteria (Zhang & Norris, 1998b; Coutte et al., 2009). As a result of recombination at the vlsE locus, the sequence of the expression cassette is changed, while the sequences of the silent cassettes remain unaltered (Zhang & Norris, 1998a). The majority of the sequence changes that occur during vlsE recombination events are observed within six variable regions of the vls1 segment (VR1–VR6); these VR sequences are separated or flanked by six invariant regions (IR1–IR6) (Zhang et al., 1997; Coutte et al., 2009).

The role of vlsE recombination in immune evasion

Although antibodies are elicited against invariable, conserved parts of VlsE during infection of mice (Liang & Philipp, 1999, 2000; Liang et al., 1999, 2000), a subset of antibodies was also found to be raised against the surface-exposed, variable regions of the protein. In addition, the latter antibody subset was shown to be VlsE variant-specific (Eicken et al., 2002; McDowell et al., 2002). Sequence variation of the VR regions can therefore alter the antigenicity of the VlsE protein (McDowell et al., 2002). The important role of vlsE recombination in immune evasion by B. burgdorferi was further demonstrated by an increased frequency and complexity of vlsE sequence alterations in bacterial isolates recovered from immunocompetent mice than in isolates obtained from severe combined immunodeficient (SCID) mice (which lack an acquired immune response) (Coutte et al., 2009). Moreover, B. burgdorferi strains that are incapable of VlsE antigenic variation (such as strains lacking plasmid lp28-1) are unable to induce persistent infections in immunocompetent mice (Zhang & Norris, 1998a,b; Labandeira-Rey et al., 2003; Lawrenz et al., 2004; Bankhead & Chaconas, 2007; Coutte et al., 2009). In SCID mice, however, differences in bacterial dissemination and persistence were not observed.
between wild-type bacteria and strains lacking either the entire lp28-1 plasmid (Labandeira-Rey et al., 2003) or the vls locus (Bankhead & Chaconas, 2007).

**The protein machinery involved in vlsE recombination**

While gene conversion at the vlsE site can be observed at high frequency in vivo, during infection of laboratory animals, vlsE switching does not occur in bacterial culture or in the midgut of infected ticks (Indest et al., 2003; Ohnishi et al., 2003; Norris, 2006; Embers et al., 2007). The initiation of vlsE recombination may therefore require some yet unidentified environmental trigger, which is exclusive to the mammalian host of the spirochetes. The identification of such a trigger, however, also requires elucidation of the mechanism of vlsE recombination and identification of the spirochete proteins involved in this process. Although the B. burgdorferi RecA homolog was considered a likely candidate as one of the major protein factors involved in gene conversion events, B. burgdorferi recA null mutants were found to be proficient in vlsE sequence variation (Liveris et al., 2008; Dresser et al., 2009). It was therefore concluded that the RecA protein is dispensable for gene conversion involving vls segments (Liveris et al., 2008). Two other proteins with putative key roles in homologous DNA recombination, however, were found to have an essential role in vlsE recombination and, consequently, in the pathogenesis of infection of B. burgdorferi (Dresser et al., 2009; Lin et al., 2009). These proteins are the orthologs of E. coli RuvA and RuvB, which together form a protein complex responsible for the migration of Holliday junctions in two recombinating (homologous) DNA molecules (Shiba et al., 1991; Iwasaki et al., 1992; Parsons et al., 1992). The importance of both proteins in vlsE recombination was revealed after disruption of genes encoding proteins with a putative function in homologous DNA recombination, repair, or replication (Dresser et al., 2009; Lin et al., 2009). B. burgdorferi mutants carrying mutations in either ruvA or ruvB showed a significant decrease in vlsE recombination. Although these mutants could induce a productive infection in immunocompetent mice during the first week postinoculation, they were cleared from the mice by day

---

**Fig. 5.** Antigenic variation in *Borrelia* species. (a) In *Borrelia burgdorferi*, (parts of) variant, silent vls cassettes (vls2-vls16, in light blue), can be transferred to the vlsE ORF within the vls ES by virtue of gene conversion. This process is dependent on the activity of RuvA and RuvB (Dresser et al., 2009; Lin et al., 2009). The silent vls segments (only vls2 to vls6 are shown) are flanked by 17-bp direct repeats (in red). The promoter within the ES, located upstream of the vlsE ORF, is indicated by a black triangle. (b) In *Borrelia hermsii*, silent vsp and vlp genes can be transferred in their entirety to the ES, thereby replacing the gene that was originally present. The gene replacements occur by gene conversion that involve sequences upstream (the upstream homologous sequence, UHS; in red) and downstream (the downstream homologous sequence, DHS; in blue) of the vsp/vlp ORFs.
In this regard, these mutants displayed a similar phenotype as spirochetes lacking either plasmid lp28-1 or the vls locus, as described above. In SCID mice, however, the ruvA and ruvB mutants displayed wild-type levels of growth and persistence, although almost all recovered clones retained the ‘parental’ vlsE sequence (Dresser et al., 2009; Lin et al., 2009). These results suggest that the reduced infectivity of the ruvA and ruvB mutants in immunocompetent mice is caused by ineffective vlsE recombination, emphasizing the crucial function of vlsE recombination in immune evasion by B. burgdorferi (Dresser et al., 2009; Lin et al., 2009).

The mutant studies by Dresser et al. (2009) and Lin et al. (2009) indicated that the RuvA and RuvB proteins from B. burgdorferi may have a pivotal role in recombinational switching at the vlsE locus. Like their orthologs from E. coli, these proteins may form a bipartite protein complex (RuvAB) that catalyzes the migration of Holliday junctions in recombining DNA. In contrast to RuvA and RuvB, the RecA protein of B. burgdorferi is not involved in vlsE recombination (as discussed above). In E. coli, the RecA protein mediates DNA strand pairing and invasion; these processes precede Holliday junction migration by the RuvAB protein complex (for a review, see Persky & Lovett, 2008). The question therefore remains as to which protein (s) is (are) responsible for homologous DNA strand pairing, strand invasion, and heteroduplex DNA formation during gene conversion events at the vlsE locus. Because recombination between vls segments is a unidirectional process in which sequences are copied from the silent vls donor sites to the vlsE gene, it has been suggested that one or more specialized proteins may be involved in vlsE switching (Dresser et al., 2009; Lin et al., 2009). However, such protein(s) have not yet been identified.

With respect to its RecA independence, vlsE recombination differs drastically from the pilin antigenic variation system of N. gonorrhoeae, which is RecA-dependent (as described extensively above). Nevertheless, at least part of the mechanism of recombination is shared by these systems, as they both require the activities of RuvA and RuvB. Finally, it is likely that antigenic variation systems similar to the B. burgdorferi vlsE system also function in the other spirochetes causing Lyme disease. This was inferred from the presence of sequences homologous to the B. burgdorferi vls locus in the isolates of B. garinii, B. afzelii and B. valaisiana (Wang et al., 2001).

**Antigenic variation in B. hermsii**

While a homologous DNA recombination–based antigenic variation system is also employed by another species from the *Borrelia* genus, B. hermsii, this system differs significantly from that of B. burgdorferi (Norris, 2006). *Borrelia hermsii*, which is one of several spirochetes that cause relapsing fever, successively evades the host’s adaptive immune system by means of antigenic variation (Barbour & Restrepo, 2000). This immune evasion is achieved by continuous exchange of antigenic lipoproteins on the surface of the bacterium. Two different types of antigenic lipoproteins have been identified in *B. hermsii*. These types do not show significant amino acid sequence homology with each other and are designated ‘variable large proteins’ (Vlp, with a molecular weight of ~36 000) and ‘variable small proteins’ (Vsp, with a molecular weight of ~20 000) (Restrepo et al., 1992; Hinnebusch et al., 1998). As the amino acid sequences of the exchanged lipoproteins differ significantly from each other, antibodies generated against one variant (‘serotype’) of lipoproteins will not be effective against a successive, new variant. Consequently, the bacterium causes relapses of disease, in which the newly emerged serotype can proliferate within the host until it is eliminated by yet another wave of serotype-specific antibodies (Barbour et al., 2006). During a single wave of infection, novel serotypes arise in a B. hermsii population at a high frequency (10^-4 to 10^-3 per cell per generation; Stoenner & Dodd, 1982), and these novel serotypes evade the generated immune response against the previous serotype to allow subsequent relapses of infection.

The Vsp and Vlp proteins are encoded by vlp and vsp genes, respectively, which are located on linear plasmids with a size of 28-32 kb (Plasterk et al., 1985; Kitten & Barbour, 1990). Within the B. hermsii genome, there is a single vlp/vsp ES (the vlp/vsp locus) containing a single gene (vlp or vsp). All other ~21 vsp and ~38 vlp genes do not contain a functional promoter and are therefore transcriptionally silent (Barbour et al., 1991; Dai et al., 2006). The single ES is situated near the telomere of a 28-kb linear plasmid and is under the control of a σ^70-type promoter (Kitten & Barbour, 1990; Barbour et al., 1991). Antigenic switching occurs when a particular vsp or vlp gene within the ES is replaced by another gene (either full-length or near full-length) through a nonreciprocal homologous DNA recombination event (Fig. 5b) (Dai et al., 2006). Thus, the originally expressed gene is lost from the ES, while the (archived) copy of the donor gene remains intact (Plasterk et al., 1985; Restrepo et al., 1992). The crossover points for recombination between the vlp/vsp genes are predominantly located at ~60-bp ‘upstream homology sequences’, located around the start codon of the ORFs, and 214-bp ‘downstream homology sequences’, located downstream of the ORFs (Fig. 6b) (Burman et al., 1990; Kitten & Barbour, 1990; Dai et al., 2006). The details regarding this homologous DNA recombination process and the enzymatic machinery involved in Vsp and Vlp variation in B. hermsii have yet to be elucidated.
Antigenic variation through gene conversion in Treponema pallidum

The spirochete *T. pallidum*, the etiological agent of syphilis, utilizes gene conversion in order to vary the amino acid sequence of the TprK protein. This protein is highly heterogeneous among and within *T. pallidum* isolates and is predicted to be localized to the bacterial outer membrane (Centurion-Lara et al., 2000; LaFond et al., 2003; Giacani et al., 2010). TprK is encoded by the *tprK* gene, which contains seven discrete variable regions (termed V1–V7, having lengths of 32–91 bp) within its ORF (Stamm & Bergen, 2000; LaFond et al., 2003; Centurion-Lara et al., 2004). The V regions of TprK were found to induce a humoral immune response in infected rabbits (Morgan et al., 2002). Interestingly, the binding of antibodies to a particular V region was abolished when specific amino acid residues in this region were substituted (Morgan et al., 2003). In addition, the immune response of *T. pallidum*-infected rabbits was shown to select against specific TprK epitopes, resulting in the selection of new TprK variants during the course of infection (Giacani et al., 2010). Thus, the variation in V region sequences within *tprK* can result in antigenic variation.

Modification of the V regions was proposed to depend upon homologous DNA recombination between the V regions and similar sequences outside the (single) *tprK* expression locus (Stamm & Bergen, 2000) (Fig. 6). The *T. pallidum* genome contains ~50 whole or partial V region donor sequences, which can be copied and recombined into the *tprK* expression locus, generating novel, variant V regions (Centurion-Lara et al., 2004). Theoretically, millions of chimeric variants of TprK can be generated by this mechanism. Recombination between V regions and V donor sequences may depend on the presence of 4-bp repeats located internally as well as at the flanks of the donor as well as recipient DNA sequences. As the donor sites appear to be unaltered after recombination, the homologous DNA recombination between the donor sequences and the V regions occurs by a gene conversion mechanism (Centurion-Lara et al., 2004). Interestingly, it was reported that (the sequences of) the V donor sites are stable, indicating that recombination among donor sites does not take place. The gene conversion events involving V regions in *T. pallidum* are therefore directional, such that only the V sequences within the *tprK* gene are used as recipients. Finally, although the protein machinery involved in V region gene conversion has not yet been elucidated, the RecA protein was suggested to play a role in this process (Centurion-Lara et al., 2004).

Gene conversion-mediated antigenic variation in Mycoplasma spp.

Gene conversion events are also considered to play an important role in antigenic variation of *Mycoplasma* spp. These bacterial species belong to the class of *Mollicutes* and are unusual in the sense that they lack a cell wall and represent the smallest known self-replicating organisms. They are generally believed to have evolved from a Gram-positive ancestor by undergoing gradual, but significant,
genome size reductions (Maniloff, 1992). The clinically most important Mycoplasma spp. are Mycoplasma pneumoniae and Mycoplasma genitalium. M. pneumoniae causes a wide range of respiratory infections, including tracheobronchitis, pharyngitis, and atypical pneumonia. This bacterium can cause up to 40% of all community-acquired pneumonias and as many as 18% of the childhood pneumonia cases that require hospitalization (Waites & Talkington, 2004). Mycoplasma genitalium is a pathogen that is able to cause nongonococcal urethritis in men (Taylor-Robinson, 2002) and various inflammatory diseases of the genital tract in women (Moller et al., 1984; Clausen et al., 2001; Cohen et al., 2002; Manhart et al., 2003).

**RepMP elements in the M. pneumoniae genome**

Despite the limited size of the *M. pneumoniae* (strain M129) genome [816 394 base pairs (bp)], which is at least five times smaller than the genome of *E. coli*, it is remarkable that approximately 8% of its sequence consists of repeated DNA elements (Himmelreich et al., 1996; Dandekar et al., 2000). Four different types of these elements were identified by whole-genome sequence analysis and were termed RepMP1, RepMP2/3, RepMP4, and RepMP5, respectively (Su et al., 1988; Wenzel & Herrmann, 1988; Ruland et al., 1990). The multiple copies (variants) of each of these RepMP elements are similar but not identical in sequence. In total, the genome of *M. pneumoniae* strain M129 contains 14 variants of RepMP1 (RepMP1-a to -n), 10 variants of RepMP2/3 (RepMP2/3-a to -j), eight variants of RepMP4 (RepMP4-a to -h), and eight variants of RepMP5 (RepMP5-a to -h). Moreover, these variants can differ significantly in size (Spuesens et al., 2009, 2011). Of the RepMP2/3, RepMP4, and RepMP5 elements, only a single variant is located within an ORF that is known to be expressed into protein (Fig. 7). Variant RepMP2/3-d and variant RepMP4-c are both located within the MPN141 ORF, whereas variant RepMP5-c is present within the MPN142 ORF (Fig. 7). All other variants of these elements, which are found dispersed throughout the genome, are probably not expressed into protein. The MPN141 and MPN142 ORFs are contained within the so-called P1 operon (Walbo & Krause, 2006). ORF MPN141 encodes the ~170-kDa P1 protein, which is the major adherence protein (also termed adhesin) of *M. pneumoniae*. P1 is essential for attachment of the bacteria to host cells, but is also a major immunogen in *M. pneumoniae*-infected individuals (Hu et al., 1983; Leith et al., 1983; Razin & Jacobs, 1992; Seto et al., 2005). The P1 protein forms part of the apical attachment organelle (or ‘terminal tip structure’) of *M. pneumoniae*, which is composed of a complex of adherence proteins and proteins with an accessory role in adhesion. Among these accessory proteins are the P40 and P90 proteins (Layh-Schmitt & Herrmann, 1994). These proteins are proteolytic products of a ~130-kDa precursor protein, which is encoded by the MPN142 ORF (Sperker et al., 1991; Layh-Schmitt & Herrmann, 1992; Catrein et al., 2005) (Fig. 7). Similar to the P1 protein, both P40 and P90 are immunogenic (Leith et al., 1983; Sperker et al., 1991; Layh-Schmitt & Herrmann, 1992; Franzoso et al., 1993; Seto et al., 2005).

![Fig. 7. Structure of the Mycoplasma pneumoniae P1 operon and predicted mechanism of antigenic variation of the P1 protein. The P1 operon contains three ORFs, that is, MPN140, MPN141, and MPN142, the latter two of which encode antigenic surface proteins (P1 and P40/P90). MPN141 contains two variable DNA elements, a RepMP4 element (RepMP4-c; in blue) and a RepMP2/3 element (RepMP2/3-d; in red), which are not unique to the P1 operon: in total, eight RepMP4 variants and 10 RepMP2/3 variants are found dispersed throughout the *M. pneumoniae* genome. The downstream MPN142 gene contains a RepMP5 element (in yellow); this element has eight counterparts in the genome. The RepMP sequences can be transferred (in part) to their homologous sequences within the P1 operon or to other homologous variants (Spuesens et al., 2009, 2011) by means of segmental gene conversion. To illustrate these gene conversion processes, a subset of RepMP2/3 elements is shown in different colors; sequences from these elements can be transferred to other RepMP2/3 elements in the genome, including the element that is located within MPN141. The P1 promoter is indicated by a black triangle.](image-url)
Gene conversion(-like) events involving RepMP elements in *M. pneumoniae*

Based on their sequence similarity, the variants of a given RepMP element were hypothesized to rearrange by means of homologous DNA recombination (Ruland et al., 1990). Indirect support for this hypothesis was obtained from sequence analysis of the variants from various *M. pneumoniae* isolates. In some isolates, sequences derived from one or more RepMP variants (donors) appeared to have been copied, in a unidirectional fashion, to other (recipient) variants (Spuesens et al., 2009, 2011). As a consequence of this sequence transfer, the original (and subsequently replaced) sequences of the recipient seemed to have been lost from the bacterial genome (Kenri et al., 1999; Spuesens et al., 2009, 2010, 2011). A portion of the putative recombination events that were observed in *M. pneumoniae* isolates also involved the RepMP variants located in the MPN141 and MPN142 ORFs. Sequences within these variants appeared to be replaced by sequences originating from other variants in the bacterial genome, leading to alterations in the amino acid sequences encoded by the MPN141 and MPN142 ORFs (Kenri et al., 1999; Dumke et al., 2006; Pereyre et al., 2007; Spuesens et al., 2009, 2010, 2011). This demonstrated that gene conversion(-like) processes involving RepMP elements can indeed lead to changes in the amino acid sequences of important antigenic proteins of *M. pneumoniae*.

While recombination events were also reported to occur among RepMP1 elements (the only type of large repeat elements found outside of the P1 operon), the biological (or immunological) significance of these events, which led to the fusion of certain *M. pneumoniae* ORFs, is as yet unclear (Musatovova et al., 2008). Finally, it is important to note that the (putative) sequence rearrangements among RepMP variants in *M. pneumoniae* have hitherto only been inferred from comparisons between the RepMP sequences obtained from different strains and isolates. The actual recombination of the sequence of a RepMP element within a clonal bacterial line has not yet been demonstrated in *M. pneumoniae*. Such changes, however, have been shown to occur in *M. genitalium*, which is genetically closely related to *M. pneumoniae* (see next section).

Recombination between MgPar elements in *M. genitalium*

Although the ‘minimal’ genome of *M. genitalium* (strain G-37) is even smaller than that of *M. pneumoniae* (580 070 bp; Fraser et al., 1995), it contains a significant number of DNA repeats. These repeats, which are designated as MgPa repeats (or MgPar sequences), constitute approximately 4% of the complete *M. genitalium* genome (Peterson et al., 1993, 1995; Fraser et al., 1995). Like the RepMP elements from *M. pneumoniae*, the *M. genitalium* MgPar elements are believed to provide a source of sequence variation for genes encoding antigenic proteins. These genes include mgpB (or MG191) and mgpC (or MG192), which code for the immunogenic proteins MgPa and P110, respectively (Ma et al., 2007). Both MgPa and P110 play a role in attachment of the bacteria to host cells (Hu et al., 1987; Burgos et al., 2006). The sequences of both genes were found to change at a relatively high frequency within *M. genitalium* strains, both in vitro and in vivo. This sequence variation was found to result from recombination between MgPar sequences (Iversion-Cabral et al., 2006, 2007; Ma et al., 2007). Interestingly, the majority of the observed inter-MgPar recombination events involved reciprocal DNA exchange, although examples of nonreciprocal, gene conversion-like events were also detected (Iversion-Cabral et al., 2007; Ma et al., 2007). In this regard, the mechanism of recombination between repetitive sequence elements in *M. genitalium* appears to differ significantly from that in *M. pneumoniae*, in which exclusively nonreciprocal, duplicative gene conversion events were observed (as described above).

Another major dissimilarity between *M. genitalium* and *M. pneumoniae* is the frequency at which repeated DNA elements seem to recombine. While these recombination events are readily detected in clonal isolates of *M. genitalium* in vitro, this has not yet been reported to occur at all in *M. pneumoniae* (Iversion-Cabral et al., 2006, 2007; Ma et al., 2007; Spuesens et al., 2009, 2011). It was previously suggested that the cause of this difference could reside in the specific activities and/or efficiency of the enzymatic machineries involved in the rearrangement of repeat elements in both species (Sluijter et al., 2010). These machineries, however, have not yet been entirely defined. On the basis of sequence homology, *Mycoplasma* spp. have been predicted to possess a core set of DNA recombination enzymes that includes RecA, single-stranded DNA-binding protein (SSB), RecU, RuvA, and RuvB (Fraser et al., 1995; Himmelreich et al., 1996; Dandekar et al., 2000; Carvalho et al., 2005). The in vitro activities of most of these proteins have recently been analyzed. The SSB protein from *M. pneumoniae* (Sluijter et al., 2008) and the RecA proteins from both *M. pneumoniae* and *M. genitalium* (Sluijter et al., 2009) displayed activities similar to that of their counterparts from other bacteria. In contrast, the RecU, RuvA, and RuvB proteins from *M. pneumoniae* and *M. genitalium* were found to have unique features (Ingleston et al., 2002; Sluijter et al., 2010; Estevão et al., 2011). The most remarkable observations were made for the RecU proteins from both species. The RecU protein from *M. genitalium*, RecU<sub>Mge</sub>...
found to be a potent Holliday junction–resolving enzyme, which cleaves synthetic Holliday junction substrates in a DNA sequence–specific, Mn²⁺-dependent fashion (Sluijter et al., 2010, 2011). In contrast, the RecU protein from M. pneumoniae, RecU<sub>Mpn</sub>, did not exhibit any activity in vitro (neither DNA binding nor DNA cleavage). This inactivity could be attributed to the presence of a glutamic acid residue at position 67 of the protein, which is not conserved in RecU<sub>Mgo</sub> or in any other bacterial RecU sequence (Sluijter et al., 2010). In addition, this (inactive) protein is only expressed by a subset of M. pneumoniae isolates. This subset, which is designated as ‘subtype 2’, represents one of the two major evolutionary lineages of M. pneumoniae. Strains belonging to the other subtype (subtype 1) were found to have a nonsense codon in the RecU<sub>Mpn</sub> gene (or MPN528a) and are therefore only capable of expressing a C-terminally truncated, nonfunctional polypeptide (Sluijter et al., 2010; Spuesens et al., 2010). Thus, all M. pneumoniae strains analyzed to date may be considered as recU mutants. Consequently, this species may be restricted in its DNA recombination capability as opposed to other bacteria, including M. genitalium. Although we cannot rule out that another, yet uncharacterized protein may complement the RecU deficiency of M. pneumoniae, the lack of a functional Holliday junction resolvase could provide an explanation for the relatively low rate of recombination that is observed among repeat elements in M. pneumoniae. In addition, this deficiency may explain why mutagenesis procedures that rely on homologous DNA recombination are highly inefficient in M. pneumoniae. These hypotheses will need to be confirmed by genetic studies in which the growth potential of knock-out and knock-in mutants of an active gene, as well as the proficiency of these mutants in DNA recombination and DNA repair, is measured.

While both M. pneumoniae and M. genitalium can change their antigenic proteins by means of recombination between repeated DNA elements, it is still unknown whether the modification of these proteins actually results in avoidance of antibacterial antibody activity to allow the bacteria to (temporarily) escape from host immune surveillance. Moreover, the protein machinery required for DNA repeat rearrangement in these mycoplasmas still needs to be identified. While it is likely that antigenic variation depends upon many of the protein factors that mediate ‘regular’ homologous DNA recombination and DNA repair processes, this has yet to be established.

### Gene conversion–induced antigenic variation in eukaryotes

In the bacterial pathogens of the genus Borrelia, the variant antigen genes are located near the ends of linear plasmids (see above). This telomeric genomic location for variant antigen gene families also appears to be shared in eukaryotic pathogens, where these gene families are normally predominantly located near chromosome ends (Stringer, 2007; Scherf et al., 2008; Palmer et al., 2009). Examples of such eukaryotic gene families are discussed below and include the msg genes of the fungal pathogen Pneumocystis and the VSG genes of T. brucei, which have an invariably telomeric (or subtelomeric) location.

It is possible that in some cases a telomeric location plays a critical role in the transcriptional control of the variant antigen gene ESs. In addition, telomeres presumably also provide particularly recombinogenic environments, thereby facilitating the creation and maintenance of the genetic diversity in these polymorphic gene families involved in phenotypic variation (including antigenic variation) (Barry et al., 2003; Palmer & Brayton, 2007).

### DNA recombination–based antigenic variation in Pneumocystis spp.

Pneumocystis spp. are haploid fungi that can cause life-threatening pneumonia in immunocompromised hosts. They are found in the lungs of a wide variety of mammalian species, but do not appear to cause disease in hosts that have a fully functional immune system. The best studied member of the Pneumocystis genus is P. carinii, which can colonize and infect rats (Icenhour et al., 2001, 2002). It has been hypothesized that the major surface glycoprotein (Msg) of Pneumocystis spp., which has a yet undefined function, can undergo antigenic variation (Wada & Nakamura, 1996a). The Msg protein is encoded by a family consisting of 80 genes, each with a length of ~3 kb, which are located at the ends of each chromosome (Keely et al., 2005; Keely & Stringer, 2009). It was demonstrated for P. carinii that only one msg gene is transcribed at a time from a single (sub)telomeric locus. This locus is termed ‘the expression site’ (Wada et al., 1995; Sunkin & Stringer, 1996; Wada & Nakamura, 1996b). Consequently, a single Msg protein is expressed on the fungal surface at a time (Angus et al., 1996; Schaffzin & Stringer, 2004). The msg gene in the ES appears to be replaced periodically by another msg gene by means of homologous DNA recombination through a yet unknown mechanism. The substitution of one msg gene by another can lead to the expression of a novel variant of the Msg protein at the fungal surface (Fig. 8). An important and unique feature of the ES is a sequence designated the ‘upstream conserved sequence’ (UCS), located 5’ of the msg ORF (Wada et al., 1995; Edman et al., 1996). Downstream of, and adjacent to, the UCS, a 24-bp conserved sequence is present which is called the ‘conserved recombination junction element’ (CRJE). This CRJE is not only
located at the msg ES in the *P. carinii* genome, but also immediately upstream of all known msg ORFs (Wada *et al.*, 1995; Sunkin & Stringer, 1996; Schaffzin *et al.*, 1999). Owing to the conserved nature of the CRJE as well as its location, this sequence element was hypothesized to have a function in the recombination of msg sequences at the msg ES (Wada *et al.*, 1995). Interestingly, evidence for DNA recombination events between msg sequences was not only found exclusively at the msg ES but also at the silent, donor msg genes (Wada & Nakamura, 1996a; Keely *et al.*, 2005; Kutty *et al.*, 2008). A similar observation was made in the bacterium *M. pneumoniae*, in which the (silent) repetitive RepMP elements are also able to recombine their sequences, as discussed above.

While most of the current knowledge on the msg gene family is obtained from studies with (rat-specific) *P. carinii*, it is to be expected that similar dynamics of msg recombination are found in other *Pneumocystis* spp., including the human pathogen *Pneumocystis jirovecii* (Kutty *et al.*, 2008; Ripamonti *et al.*, 2009). Finally, many questions remain as to the role of Msg variation in host colonization and/or infection by *Pneumocystis* spp. The occurrence of homologous DNA recombination events at the msg ES in *P. carinii* has only been inferred from the detection of a large variety of different msg variants at this site (Sunkin & Stringer, 1996, 1997; Keely *et al.*, 2003). It is yet unknown, however, via which mechanism these recombination events would occur. Moreover, it remains to be determined at which frequency recombination at the ES takes place and whether the introduction of novel msg genes at this site leads to true antigenic variation and, consequently, immune evasion.

**DNA rearrangements and antigenic variation in the African trypanosome T. brucei**

The African trypanosome *T. brucei* is a flagellated unicellular protozoan parasite that causes human African trypanosomiasis. This neglected tropical disease is endemic to sub-Saharan Africa and is transmitted by tsetse flies biting humans as well as a broad range of mammalian reservoir hosts. *Trypanosoma brucei* replicates extracellularly in the mammalian bloodstream where it establishes chronic infections. A characteristic feature of trypanosomiasis is the relapsing parasitemia, where different antigenic variants of *T. brucei* are successively replaced by new switch variants that (temporarily) escape immune recognition. This is a similar process to that seen in the relapsing fever *Borrelia* (see previous section). African trypanosomes are able to colonize the hostile environment of the mammalian bloodstream through an extraordinarily sophisticated strategy of antigenic variation of their surface coat (Taylor & Rudenko, 2006; Horn & McCulloch, 2010; Schwede & Carrington, 2010).
Bloodstream-form trypanosomes express a dense layer of variant surface glycoprotein (VSG) on their surface, coating the entire cell. This VSG coat shields invariant receptor molecules on the trypanosome surface from recognition by host antibodies and also protects the parasite from lysis by the alternative pathway of the host complement system (Schwede & Carrington, 2010). VSG molecules are attached to the trypanosome cell surface via glycosyl-phosphatidylinositol anchors and have a highly variable region at the N-terminus important for immune evasion. A striking feature of VSGs is that proteins that share only 16% amino acid sequence identity can, nonetheless, fold into similar tertiary structures exhibiting a characteristic ‘VSG-fold’ (Blum et al., 1993; Carrington & Boothroyd, 1996; Schwede & Carrington, 2010). This conserved shape of different, antigenically diverse VSGs presumably facilitates tight packing of dissimilar molecules into a protective surface, even while the trypanosome switches between different VSG variants.

An individual trypanosome has a vast repertoire of more than 1500–2000 VSG genes and pseudogenes, which are all kept transcriptionally silent with the exception of one (Fig. 9) (Borst, 2002; Berriman et al., 2005; Marcello & Barry, 2007). The active VSG is located in one of about 15 telomeric VSG ES transcription units (Berriman et al., 2002; Hertz-Fowler et al., 2008). The silent VSGs are mainly located in extensive tandem arrays at the subtelomeres of the large megabase chromosomes (Berriman et al., 2005). In addition, VSGs are located adjacent to the telomere repeats of a large range of T. brucei chromosomes. Telomeres appear to be particularly recombinogenic locations in the cell, allowing efficient movement of silent VSGs into the active ES through DNA rearrangements resulting in a VSG switch.

**The antigenically diverse VSG repertoires**

The T. brucei VSG repertoire is vast, comprising up to 2000 VSG genes and pseudogenes in the genome of strain T. brucei 927 (Marcello & Barry, 2007; Marcello et al., 2007), while the laboratory strain T. brucei 427 has an even larger VSG repertoire (Callejas et al., 2006). Most of these VSGs are localized at subtelomeric positions on megabase chromosomes, where they are arranged head to tail in large tandem arrays (Fig. 9) (Berriman et al., 2005). These arrays of silent VSGs form extensive, haploid (or single-copy) regions, which are attached to conserved diploid chromosomal ‘cores’. This means that although the ‘housekeeping’ genes present in the chromosomal cores are present as two diploid copies, the silent VSGs are normally each present as a single copy within the cell. These arrays of silent VSGs can be very extensive, and an analysis of chromosome 1 in T. brucei 427 showed that more than 75% of the total chromosome length was made up of polymorphic subtelomeric VSG arrays (Callejas et al., 2006). As sister chromosomes do not necessarily both have VSG arrays at a given subtelomere, this extensive variation in the extent as well as number of VSG arrays explains the large difference in karyotypes found between different T. brucei strains (Melville et al., 2000).
The vast majority of silent VSG genes and pseudogenes within the T. brucei genome are located in these extensive subtelomeric VSG arrays. However, single VSGs are also located at the telomeres of a discrete class of small (50-100 kb) T. brucei minichromosomes, of which there are about one hundred within the cell. These minichromosomes consist of palindromic arrays of characteristic 177-bp simple sequence repeats that have a telomere proximal VSG at each end. As no other coding sequences have been found on these small chromosomes, their primary function could be to provide recombinogenic regions for silent donor VSGs (Wickstead et al., 2004). Last, a small subset of VSGs is in the telomeric VSG ES transcription units themselves.

**Mechanisms of switching between different VSGs**

Trypanosomes can switch the active VSG using different mechanisms including DNA rearrangement and transcriptional control (Fig. 10) (Taylor & Rudenko, 2006; Horn & McCulloch, 2010). First, gene conversion can result in a silent VSG being copied from a master copy into the active ES, thereby replacing the previous VSG (Bernards et al., 1981; Pays et al., 1983). Less frequently, a DNA cross-over can occur at two telomere ends whereby the active VSG is exchanged for a silent VSG located at a telomere (Pays et al., 1985; Rudenko et al., 1996). Last, the trypanosome can switch by activating transcription at a different ES while silencing the previously expressed ES (Bernards et al., 1984; Myler et al., 1984). Although transcriptional switches between ESs appear to be relatively frequent (at least early in an infection) (McCulloch et al., 1997; Aitcheson et al., 2005), this switching mechanism only accesses the relatively small pool of VSGs located at ES telomeres.

By far the most important switching mechanism during a chronic infection is mediated by gene conversion reactions, as this allows the trypanosome to insert a copy of any VSG in the genome into the active ES. VSG switching mediated through gene conversion is facilitated by the presence of upstream homology in the form of 70-bp repeat sequences, which are invariably present upstream of the silent VSG (Fig. 11) (Aline et al., 1985). Approximately 92% of the full-length VSGs in T. brucei 927 are flanked upstream by at least one 70-bp repeat (Marcello & Barry, 2007). Downstream homology is present within the 3′ end of the VSG gene (or pseudogene) itself (Marcello & Barry, 2007) (Fig. 11). VSG ESs (which are typically about 40-60 kb long) contain large stretches of
70-bp repeat sequences extending for many tens of kb upstream of the telomeric VSG (Hertz-Fowler et al., 2008). The old VSG can therefore easily be replaced using these two stretches of homology. However, the 70-bp repeat sequences are not obligatory for gene conversions mediating VSG switching. VSG switches can also occur through extensive telomere conversions extending far upstream of the 70-bp repeat sequences, where large segments of a silent ES are gene converted into the active ES (Kooter et al., 1988; Navarro & Cross, 1996). Consistent with this, deletion of the 70-bp repeats from the active ES does not prevent VSG switching occurring through telomeric gene conversions with silent ESs (McCulloch et al., 1997). This indicates that the 70-bp repeat sequences facilitate, but are not obligatory for gene conversion to take place.

An early assumption had been that most VSG switching was a consequence of intact VSG cassettes being copied into the active ES using upstream 70-bp repeat sequences and downstream VSG internal homology. However, this is not always the case. Segmental gene conversion of VSGs during VSG switching has also been shown to occur, particularly in the late stages of chronic infections (Fig. 10b) (Roth et al., 1989; Thon et al., 1989, 1990). The original studies showed that three or four different VSG pseudogenes can be recombined with each other, producing a functional, chimeric VSG (Fig. 12) (Roth et al., 1989). VSG switching mediated through segmental gene conversion was originally thought to be a curiosity rather than a key switching mechanism. However, this view has recently been re-evaluated.

Analysis of the T. brucei genome sequence has shown that the vast majority (> 90%) of the VSG repertoire is composed of pseudogenes containing frame shifts and stop codons, or genes present as truncated copies (Berimagenes et al., 2005; Marcello & Barry, 2007). This has led to a renewed interest in segmental gene conversion as a mechanism of VSG switching. It is now clear that segmental gene conversion is likely to be the most critical VSG switching mechanism during the later stages of a chronic infection, as it allows the trypanosome to recombine different segments of VSGs in a vast number of permutations and combinations to produce new, mosaic VSGs. This ability to create virtually endless numbers of chimeric VSGs during a chronic infection is undoubtedly

---

**Fig. 11.** Duplicative gene conversion can result in a silent VSG being copied into the active VSG ES transcription unit. Above are indicated tandem arrays of silent VSG genes (colored boxes) and pseudogenes (ψ) (gray filled boxes). The VSG ES transcription unit is shown below, with the promoter indicated with a flag and transcription with an arrow. VSGs are flanked upstream by characteristic 70-bp repeats (indicated with vertically hatched box). VSGs also contain highly conserved sequences at the 3′ end (indicated with a yellow box). Gene conversion can occur using these two areas of homology to copy a duplicate copy of a silent VSG into the active VSG ES (indicated below).

**Fig. 12.** Segmental gene conversion of multiple VSG pseudogenes can result in the creation of a new functional chimeric VSG. Three different VSG pseudogenes are indicated above, with disruptions of the ORF indicated with arrow heads and vertical lines. Multiple successive gene conversion reactions can take place, resulting in the creation of a new functional VSG which is a mosaic of segments of the different VSG pseudogenes.
one of the key features allowing the trypanosome to form chronic relapsing parasitemias lasting for many years.

**Hierarchy of VSG activation**

VSG switching is not completely random, and there is a rough hierarchy in the order in which VSG switch variants successively appear during a chronic infection (Caperbn et al., 1977; Robinson et al., 1999; Aitcheson et al., 2005; Marcello & Barry, 2007). Early in an infection, transcriptional activation of new VSG ESs can frequently occur, and the specific ES that is activated is not random. In a study of 127 different VSG switch variants generated as ‘single relapses’ of *T. brucei* 427, over 50% of the switches were one of the two different VSGs (Aitcheson et al., 2005). These preferentially activated VSGs were located in ESs that appeared to be particularly favored for transcriptional activation, possibly as a consequence of their epigenetic state. Consistent with this hypothesis, anecdotal evidence suggests that this preferential hierarchy of ES activation can be reset within a given *T. brucei* strain, again presumably by an epigenetic mechanism (G. Rudenko, unpublished data).

Other early events in an infection are VSG switches mediated through gene conversion of silent VSGs located on telomeres (particularly those located on *T. brucei* minichromosomes) into the active ES (Robinson et al., 1999; Marcello & Barry, 2007). Switching events that are observed a bit later in the infection include gene conversions of chromosome internal VSGs into the active ES. For example, in a study of VSGs that arise during *T. brucei* chronic infections, VSG switches mediated by gene conversions of chromosome internal VSGs were observed at day 14 of the infection (for 14 of 21 sequences analyzed) (Marcello & Barry, 2007). However, the host is continuously immunized against VSGs as they arise during an infection, inducing a selection pressure against the most frequent VSG switch events. As the infection progressed, less frequent VSG switching events were observed. Later in the infection (between days 22 and 28), mosaic VSGs started to predominate, indicating that VSG switching mediated by less frequent segmental gene conversion events became the most productive switching mechanism. In this study, more than 60% of the total VSG switch events had arisen through segmental gene conversions late in the infection (at day 28) (Marcello & Barry, 2007).

**Segmental gene conversion and the creation of VSG diversity**

The extensive VSG gene family is surprisingly diverse, with only limited similarities between different VSGs within a given VSG repertoire. An extensive analysis of 940 VSGs from the *T. brucei* 927 strain showed that about 60% of the VSGs within this repertoire are unique in sequence. The rest of the VSGs form small subfamilies with two to six similar members, with most of the subfamilies only containing two members (Marcello & Barry, 2007). It has been argued that a VSG repertoire structure containing many small VSG subfamilies is key to the formation of expressed mosaic VSG genes. In the study by Marcello & Barry (2007), a reconstruction was presented of how seven putative chimeric VSGs were produced during a chronic infection. In one case, the most plausible reconstruction of one particular VSG switching event entails seven successive gene conversion steps providing a transition between the beginning and the final VSGs. It was suggested that VSG mosaics appear to be constructed through discrete sequential gene conversion steps, with some VSG donors with high-sequence identity participating repeatedly in the process. This would mean that the fact that the VSG repertoire consists of small subfamilies of VSGs with a high degree of sequence identity could facilitate the segmental gene conversion events leading to VSG mosaicism (Marcello & Barry, 2007).

However, exactly where these new chimeric VSGs are made remains a mystery. VSG expression is highly essential for the bloodstream form of *T. brucei*. Despite the fact that the active VSG gene is single copy, it produces the most abundant transcript in the cell (about 10% of the total mRNA), encoding the most abundant protein in the cell (Nilsson et al., 2010). Perturbation of VSG synthesis triggers an extremely rapid cell cycle arrest within hours (Sheader et al., 2005). It is unknown how these new chimeric VSGs are assembled in a way that the bloodstream-form trypanosome is not left (even temporarily) with a compromised VSG coat. It has been proposed that successive gene conversion reactions could occur in ESs located at the silent ES telomeres, resulting in VSG assembly prior to expression. The newly assembled VSGs could subsequently be recombined into the active ES (Marcello & Barry, 2007). However, it is unclear how a process like this could be controlled.

It is possible that high rates of gene conversion naturally occur at the telomeric ESs, facilitating the assembly of mosaic VSGs. Frequent gene conversions between silent and active ESs are known to occur (Kooter et al., 1988; Navarro & Cross, 1996; McCulloch et al., 1997). One could hypothesize multiple successive rounds of gene conversion into ESs, whereby trypanosomes expressing dysfunctional VSGs from the active ES would immediately die. A prediction of this model would be that if rates of gene conversion into the ESs were particularly high, many ESs would contain dysfunctional chimeric VSGs in additional to the functional VSG copy.
repertoire of ESs from laboratory strain T. brucei 427 has been sequenced, and a number of ESs were found to contain VSG pseudogenes (Hertz-Fowler et al., 2008). However, we still do not have the entire repertoire of VSGs sequenced in T. brucei 427 and therefore do not have enough sequence information to determine whether these ES-located VSGs could be partially constructed VSG chimeras.

Although the VSG repertoire of T. brucei strain 927 is currently being investigated (Marcello & Barry, 2007; Marcello et al., 2007), very little is known about the extent of conservation of VSG repertoires between different strains of T. brucei. This issue still needs to be investigated using whole-genome sequencing methods. An initial comparison of 18 expressed VSGs from African T. brucei field strains with the T. brucei 927 strain showed that 14 of these expressed VSGs had homologs within T. brucei 927, with about 71% sequence identity (Hutchinson et al., 2007). Interestingly, it appears as if the VSG repertoires within different strains have diverged to be relatively strain-specific. This rapid generation of diversity in VSG repertoires presumably facilitates the superinfection of an infected mammalian host with different T. brucei strains. It is still unclear how this level of extensive VSG sequence diversity is created and maintained.

The role of double-stranded DNA breaks in VSG switching

It is thought that VSG switching is initiated by the introduction of a double-strand DNA break within the active VSG ES. Evidence supporting this hypothesis was obtained using a T. brucei strain in which a unique I-SceI restriction enzyme site was introduced adjacent to the 70-bp repeats of an actively transcribed VSG gene (Boothroyd et al., 2009). Expression of an inducible I-SceI endonuclease in these cells triggered a 250-fold increase in VSG switching rates. The effect of locating the I-SceI site within different locations of the VSG ES was also investigated, and it was shown that introduction of the double-strand break next to the 70-bp repeats was critical for the high rates of VSG switching observed. It is therefore highly likely that VSG switching often entails the generation of a double-strand break within the 70-bp repeat arrays of an active VSG ES. This double-strand break could possibly be generated by a (yet to be identified) 70-bp repeat-specific T. brucei endonuclease. Alternatively, the 70-bp repeat sequences are highly AT-rich, and high levels of transcription of these sequences could lead to the generation of random DNA breaks. Consistent with this, it has been observed that the 70-bp repeat sequences appear to be particularly prone to breakage in wild-type trypanosomes (Boothroyd et al., 2009).

DNA double-strand breaks normally constitute a danger signal for the cell and trigger a specific DNA damage response including a cell cycle arrest. These double-strand DNA breaks are then subsequently repaired through either homology-dependent repair, nonhomologous end joining (NHEJ), or microhomology-mediated end joining (Hiom, 2010). Double-strand DNA breaks within the cell can be sensed by the MRE11 complex, which regulates double-strand DNA break repair through homology-directed repair or NHEJ. Inhibition of MRE11 expression typically results in cells with increased genome fragility, including an accumulation of chromosomal breaks and sensitivity to ionizing radiation (Stracker & Petrini, 2011). MRE11 is highly conserved, and its role has been investigated in T. brucei. Inactivation of MRE11 in T. brucei leads to an impairment in homologous recombination. Cells exhibit chromosomal instability, as well as sensitivity to double-strand DNA breaks introduced by mutagenizing agents such as phleomycin or ionizing radiation (Robinson et al., 2002; Tan et al., 2002). However, the inactivation of MRE11 in T. brucei did not appear to affect the homologous recombination reactions involved in VSG switching. Possibly, T. brucei possesses an alternative, specialized pathway for sensing the double-strand breaks involved in VSG switching, and linking them to the homologous recombination machinery.

Homology-dependent repair mediated by RAD51

Double-strand breaks can be repaired using NHEJ or homology-dependent repair (Hiom, 2010). There is as yet no evidence for NHEJ in T. brucei, and VSG switching is thought to be mediated through homology-dependent mechanisms that are both RAD51 dependent and independent (Burton et al., 2007; Glover et al., 2008; Horn & McCulloch, 2010). It is possible that there is a dedicated DNA recombination machinery that is used exclusively for VSG switching. However, it can still not be excluded that VSG switching utilizes the normal DNA recombination and repair pathways. The major VSG switching mechanism is gene conversion, which is typically mediated by the RAD51 DNA recombinase. RAD51 is the eukaryotic homolog of the bacterial RecA recombinase and facilitates strand exchange between homologous sequences (San Filippo et al., 2008). Trypanosoma brucei has a RAD51 homolog, as well as four RAD51-related genes, which is an unusually large repertoire for a unicellular eukaryote (Proudfoot & McCulloch, 2005; Dobson et al., 2011). Mutation of RAD51 impairs VSG switching. However, in the absence of RAD51, VSG switching still occurs at a low rate, indicating that more than one DNA recombination pathway can be used to mediate this.
process (McCulloch & Barry, 1999). All four of the *T. brucei* RAD51 paralogs appear to play a role in general DNA recombination and repair. However, only one RAD51 paralog (RAD51-3) is critical for the gene conversion events implicated in VSG switching (Proudfoot & McCulloch, 2005; Dobson et al., 2011). Exactly which interactions or activities allow RAD51-3 to mediate antigenic variation is not known.

In addition to a role of RAD51, the BRCA2 protein is also involved in VSG switching in *T. brucei* (Hartley & McCulloch, 2008). RAD51 function in eukaryotes is typically modulated by a broad range of factors including the BRCA2 DNA repair protein, which, when mutated, confers susceptibility to breast cancer in humans. BRCA2 plays an important regulatory function in homologous recombination mediated by RAD51 and appears to facilitate the assembly of RAD51 onto the single-stranded DNA created around the DNA break, allowing the formation of presynaptic filaments of RAD51 (San Filippo et al., 2008; Holloman, 2011). Interaction of BRCA2 with RAD51 is mediated through a number of BRC repeats (each of approximately 30 amino acids) present within the BRCA2 protein. In most unicellular eukaryotes, the BRCA2 protein contains one to three copies of the BRC repeat. In contrast, the *T. brucei* BRCA2 homolog appears to have undergone an unusual expansion in the number of BRC repeats, with a total of 15 repeats predicted (Hartley & McCulloch, 2008). Deletion of the BRCA2 gene from the *T. brucei* genome resulted in an 8- to 11-fold reduction in VSG switching frequencies, similarly as observed after deletion of either RAD51 or RAD51-3 (Hartley & McCulloch, 2008). There is, however, no direct evidence that the large number of BRC repeats in the *T. brucei* BRCA2 protein is important for antigenic variation, as *T. brucei* strains that carry a BRCA2 protein with only a single BRC repeat are, nonetheless, able to efficiently switch their VSG (Hartley & McCulloch, 2008).

It is perhaps not entirely surprising that deletion of RAD51 or BRCA2 resulted in a reduction in VSG switching mediated through gene conversions. However, the fact that individual deletion of these two genes also led to a reduction in transcriptional activation of new ESs was a surprising and still unexplained result (McCulloch & Barry, 1999; Hartley & McCulloch, 2008). No evidence has yet been found that VSG ES activation is accompanied by DNA rearrangements of any kind. However, VSG ESs are invariably flanked upstream by large arrays of 50-bp simple sequence repeats (Zomerdijk et al., 1990), and DNA rearrangements or gene conversions within these repetitive regions would be extremely difficult to detect.

Homologous recombination in *T. brucei* appears to be similar to that in other eukaryotes in that it is dependent on substrate length. The rate of DNA recombination in *T. brucei* increases as substrate length increases over the range of 25-200 bp (Barnes & McCulloch, 2007). Similar to in other eukaryotes, rates of DNA recombination in *T. brucei* are suppressed by mismatches in the substrate sequences (Barnes & McCulloch, 2007). Mismatch repair in eukaryotes is typically mediated by MutS and MutL homologs, which function as heterodimers. This machinery recognizes mismatches occurring during DNA recombination and can trigger abortion of the recombination reaction (Fukui, 2010). Knockdown of the *T. brucei* mismatch repair genes *MSH2* and *MLH1* (which encode homologs of MutS and MutL, respectively) results in increased homologous recombination between both identical and divergent DNA sequences, indicating that this pathway regulating homologous recombination is present in *T. brucei* (Bell & McCulloch, 2003). However, there is as yet no evidence that this pathway plays a role in VSG switching. It is possible that there is an additional mismatch repair–independent pathway for homologous recombination in *T. brucei*, which is particularly important for DNA recombination on substrates under 200 bp and which was suggested to be independent of RAD51 (Barnes & McCulloch, 2007).

An additional complex that plays a role in DNA rearrangements involved in VSG switching in *T. brucei* is the RTR (RecQ-Top3-Rmi1) complex. This complex (also known as the BLM complex in humans) maintains genome stability by suppressing inappropriate homologous recombination. Blocking synthesis of any component of the RTR complex results in a hyper-recombinogenic phenotype, including increased rates of DNA recombination, higher levels of chromosomal breaks and translocations, and increased rates of loss of heterozygosity (Mankouri & Hickson, 2007). *Trypanosoma brucei* appears to have the equivalent of an RTR complex, and deletion of the gene encoding one of the components of this complex, TOPO3α, results in a sixfold elevation in the frequency of general gene conversion (Kim & Cross, 2010). This *T. brucei* RTR complex also appears to be important for the DNA rearrangements involved in VSG switching, as knock-out of the *T. brucei* TOPO3α protein results in a striking 10- to 40-fold higher rate of VSG switching (Kim & Cross, 2010). This hyper-recombinogenic phenotype leading to increased VSG switching is dependent on RAD51, as knock-out of both proteins abolishes the hyperswitching phenotype (Kim & Cross, 2010). The DNA rearrangements observed after knock-out of *T. brucei* TOPO3α appear to be a consequence of promiscuous recombination throughout the VSG ES, rather than specifically via the 70-bp repeat sequences. This has led the authors of this study (Kim & Cross, 2010) to postulate that *T. brucei* TOPO3α could suppress inappropriate recombinogenic structures continuously arising between the highly similar VSG ESs (Hertz-Fowler et al., 2008).
Non-RAD51-dependent pathways

In addition to homology-dependent repair, double-stranded DNA breaks can also be repaired through NHEJ or microhomology-mediated end joining (Hiom, 2010). Although RAD51-dependent pathways are clearly important for VSG switching, it is also evident that RAD51-independent pathways must also be operating (McCulloch & Barry, 1999; Proudfoot & McCulloch, 2005; Hartley & McCulloch, 2008). In eukaryotes, repair of a double-stranded DNA break can be mediated through homologous recombination or through NHEJ. NHEJ does not appear to occur (at detectable levels) in T. brucei (Burton et al., 2007; Glover et al., 2008; Horn & McCulloch, 2010). Instead, double-stranded DNA breaks that are not repaired through homologous recombination are repaired through microhomology-mediated end joining. This DNA repair pathway has been investigated in T. brucei in cells where a double-stranded DNA break has been introduced using the I-SceI restriction endonuclease (Glover et al., 2008). In this study, double-strand break repair was primarily mediated through homologous recombination mechanisms, whereby allelic sequences on the homologous chromosome were highly preferred as substrates for gene conversion (85% of monitored events) over the same sequence located in an ectopic location (Glover et al., 2008). Of the 57% of the cells that recover from a double-stranded DNA break, about 5% use ectopic homologous recombination or microhomology-mediated end joining, which appears to be a form of micro single-strand DNA annealing. It is possible that this pathway for double-strand DNA break repair is important for VSG switching, but this issue is currently unclear.

In summary, there are still many open questions regarding the mechanism of the DNA rearrangements that mediate VSG switching in T. brucei. Considering the importance of segmental gene conversion events for antigenic variation of VSGs during a chronic infection, it is likely that this process is mediated by a specialized DNA recombination pathway. However, as T. brucei has branched off from other eukaryotes relatively early in evolution (Embley & Martin, 2006), the players that are involved are not necessarily conserved in other eukaryotes and will not be easy to identify in the absence of appropriate genetic screens. As described above, antigenic variation mediated by gene conversion is observed in a number of unrelated pathogens, including parasites other than T. brucei (Palmer & Brayton, 2007). However, antigenic variation in African trypanosomes is unprecedented with regard to the size of the repertoire of antigenically diverse VSG genes and pseudogenes, and the extent of antigenic diversity which can be created through complex, successive gene conversion events. Clearly, the future challenge will be to discover how these events are mediated and controlled.

Conclusions

While a clear overlap exists in many of the genetic mechanisms employed by a wide range of human pathogens in order to vary their outer surface antigens (as summarized in Table 1), it is also obvious that some species have

<table>
<thead>
<tr>
<th>Species</th>
<th>Gene (encoded antigenic protein)</th>
<th>Donor elements (number)</th>
<th>Employed mechanism of DNA recombination</th>
<th>Proteins involved in DNA recombination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prokaryotes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. burgdorferi</td>
<td>vlsE (VlsE)</td>
<td>vls (15)</td>
<td>Gene conversion (segmental)</td>
<td>RuvA, RuvB</td>
</tr>
<tr>
<td>B. hermsii</td>
<td>vsp (Vsp)</td>
<td>vsp (~21)</td>
<td>Gene conversion (segmental)</td>
<td>Unknown</td>
</tr>
<tr>
<td>T. pallidum</td>
<td>tprK (TprK)</td>
<td>V sites (~50)</td>
<td>Gene conversion (segmental)</td>
<td>Unknown</td>
</tr>
<tr>
<td>M. pneumoniae</td>
<td>MPN141 (P1)</td>
<td>RepMP4 (7)</td>
<td>Gene conversion (segmental)</td>
<td>Unknown</td>
</tr>
<tr>
<td>M. genitalium</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P. carinii</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T. brucei</td>
<td>VSG (VSG)</td>
<td>Silent VSG genes (~2000)</td>
<td>Gene conversion</td>
<td>RAD51-3 (a RecA homolog), BRCA2, TOP3α</td>
</tr>
</tbody>
</table>

*See text for details on the various antigenic variation systems. The systems are listed according to the order of description in the text.*


Briggs GS, Mahdi AA, Weller GR, Wen Q & Lloyd RG (2004) Interplay between DNA replication, recombination and...

References


Briggs GS, Mahdi AA, Weller GR, Wen Q & Lloyd RG (2004) Interplay between DNA replication, recombination and...


