Switching trypanosome coats: what’s in the wardrobe?

Jesse E. Taylor\textsuperscript{1,2} and Gloria Rudenko\textsuperscript{1}

\textsuperscript{1}Peter Medawar Building for Pathogen Research, University of Oxford, South Parks Road, Oxford OX1 3SY, UK
\textsuperscript{2}Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK

The African trypanosome \textit{Trypanosoma brucei} is best known for its extraordinarily sophisticated antigenic variation of a protective variant surface glycoprotein (VSG) coat. \textit{T. brucei} has >1000 VSG genes and pseudogenes, of which one is transcribed at a time from one of multiple telomeric VSG expression sites. Switching the active VSG gene can involve DNA rearrangements replacing the old VSG with a new one, or alternatively transcriptional control. The astonishing revelation from the \textit{T. brucei} genome sequence is that <7% of the sequenced VSGs seem to have fully functional coding regions. This preponderance of pseudogenes in the VSG gene repertoire will necessitate a rethink of how antigenic variation in African trypanosomes operates.

Introduction

African trypanosomes including \textit{Trypanosoma brucei} are flagellated unicellular protozoa that infect the mammalian bloodstream, causing African sleeping sickness in humans. Parasites are transmitted between humans and a large range of mammalian reservoir hosts by tsetse flies. Bloodstream-form trypanosomes thrive in an unusually exposed niche for a parasite, remaining extracellular and fully exposed to immune attack during the course of a chronic infection. Cells are coated with a dense sheath of tightly packed variant surface glycoprotein (VSG) homodimers, analogous to a fur coat, that protects against complement-mediated lysis. Synthesis of VSG is essential for the bloodstream-form trypanosome, even \textit{in vitro} [1]. An infected host eventually mounts an effective antibody response against the predominant VSG variants leading to antibody-mediated lysis. However, trypanosomes continually arise that have switched to antigenically different VSGs, prolonging a chronic infection that can last for years. Antigenic variation in African trypanosomes is reviewed in Refs \textsuperscript{[2–5]}.

Switching the protective VSG coat

An individual bloodstream-form trypanosome expresses a single VSG gene, in a mutually exclusive fashion \textsuperscript{[6,7]}, from one of \textasciitilde 20 telomeric bloodstream-form VSG expression site transcription units, which can be present at one or both ends of \textit{T. brucei} chromosomes \textsuperscript{[8]} (Figure 1a). Bloodstream-form VSG expression sites are transcribed as polycistronic transcription units containing members of up to 12 different families of expression site-associated genes (ESAGs) in addition to the telomeric VSG (Figure 2b). By contrast, metacyclic VSG expression sites, which are active in infective metacyclic-stage trypanosomes, contain only the telomeric VSG (Figure 2c). Metacyclic VSG expression sites are active in the host immediately after infection, but are quickly silenced as the trypanosome switches to exclusive activation of one of the bloodstream-form VSG expression sites. Transcription of VSG is maintained in rapidly dividing bloodstream-form trypanosomes, and is only shut down in the quiescent ‘stumpy-form’ trypanosomes that arise when the infection reaches high densities \textsuperscript{[9]}.

Switching the active VSG can involve activation of another VSG expression site (\textit{in situ} switch in Figure 1b). Alternatively, DNA rearrangements including gene conversion or telomere exchange can insert one of many hundreds of silent VSG genes into an active VSG expression site (Figure 1b). VSG genes with highly variable sequences can be moved using mechanisms relying on homologous recombination. The necessary homology can be provided upstream of the donor VSGs by variable numbers of characteristic 70-bp repeats, and downstream by conserved sequences present at the 3’ ends of the VSG genes \textsuperscript{[10,11]} (Figure 3). This event is often a segmental gene conversion, because the crossover between the resident and donor VSG genes can occur at various places within the 3’ ends of both genes \textsuperscript{[11,12]}. It has been shown that successive segmental gene conversions can occur during antigenic variation \textsuperscript{[13]}. Later in an infection VSG switches mediated by more complicated segmental gene conversions can also be detected. This process generates chimeric VSG genes, which can be composed of segments of several (e.g. three or four) different VSG pseudogenes \textsuperscript{[14,15]} (Figure 3b). Both RAD51-dependent and RAD51-independent mechanisms mediate the DNA rearrangements underlying VSG switching, and it has been speculated that the latter set of processes might be responsible for the microhomology-based recombination reactions that are observed in segmental gene conversion events \textsuperscript{[16]}. However, the mechanism mediating VSG gene conversion events of different degrees of complexity still needs to be elucidated. What is clear is that complicated segmental gene conversion events have the potential for generating large numbers of entirely novel mosaic VSGs using different combinations of gene segments from the same set of VSG pseudogenes. The role of VSG pseudogenes in \textit{T. brucei} antigenic variation has been discussed in Refs \textsuperscript{[17,18]}.
The genome sequence has revealed that *T. brucei* has a repertoire of at least 1250 to 1500 VSG genes [19–22] (Trypanosoma brucei GeneDB: www.genedb.org/genedb/tryp). The vast majority of these silent VSGs (at least 1250) are present in several tandem arrays located at subtelomeres [19] (Figure 2a) (Trypanosoma brucei GeneDB: www.genedb.org/genedb/tryp). However, another set of silent VSG genes (~100–200) is present at the telomeres of small mitotically stable minichromosomes of ~50–100 kb [23,24]. Minichromosomes consist of arrays of 177-bp repeats arranged in a large palindrome, with VSG genes at their telomeres [25] (Figure 2d). It is likely that this particularly abundant class of chromosomes (~100 per cell) has expanded in *T. brucei* to increase the number of telomeric VSGs. The VSG genes activated early in an infection using DNA rearrangements seem to be frequently located at telomeres, and are often located on minichromosomes [26,27].

The different VSG switching mechanisms therefore access different sized but overlapping pools of VSGs. Transcriptional switches access the limited set of ~20 VSG genes located in VSG expression sites (Figure 1a). Telomere exchange accesses a repertoire of ~150–250 VSG genes located at the ends of various chromosomes. However, gene conversion is undoubtedly the most important switching mechanism during a chronic infection, because in theory it could enable the movement of any of the silent VSGs into the active VSG expression site [10–13,26,28] (Figure 1a). The classic version of VSG switching presented within the textbooks has been that of silent intact VSG

---

**Figure 1.** VSG genes and VSG switching in African trypanosomes. (a) Genomic location of VSG genes in *T. brucei*. VSG genes are indicated with coloured boxes. VSGs are present in subtelomeric VSG arrays, at telomeres or in one of the many VSG expression site transcription units. VSG expression sites are shown with flags indicating the promoters, and a red arrow indicating transcription at the active VSG expression site. An approximate estimate of the total size of the pool of VSGs in the different genomic locations is indicated above. (b) Different VSG switching mechanisms in African trypanosomes. The coloured rectangular outlines represent trypanosomes expressing a single VSG gene (filled coloured box) from a telomeric VSG expression site. The VSG expression site promoter is indicated with a flag, and transcription with an arrow. Silent VSG genes are located either in tandem arrays at subtelomeric locations or at telomeres, including within VSG expression sites. Switching the active VSG gene can be mediated by different switching mechanisms. Left: gene conversion results in the duplication of a previously silent VSG gene into the active VSG expression site. Centre: telomere exchange involves a DNA crossover within two chromosome ends. Right: in situ switch – transcriptional activation of a new VSG expression site concurrent with silencing of the old one.
cassettes being moved into an active VSG expression site. Although VSG switching using segmental gene conversion of VSG pseudogenes has previously been documented and discussed, the full significance of this switching mechanism was not realized, and many in the field have considered it to be of relatively minor importance [17,18]. However, this view now needs to be re-evaluated in light of new data.

Revelations from the T. brucei genome sequence

The astonishing view revealed by the sequence of the T. brucei genome is that the vast majority of the silent VSG genes within the VSG arrays present on T. brucei mega-base chromosomes are pseudogenes (Figure 2a) [19] (Trypanosoma brucei GeneDB: www.genedb.org/genedb/tryp).

Of the 806 VSG genes analysed in the fully assembled parts of the T. brucei 927 genome sequence, only 7% seem to be intact and containing all recognizable features of known functional VSGs [19]. By contrast, 66% of the VSGs are full-length pseudogenes with frameshifts or in-frame stop codons, 18% of the VSGs are gene fragments, mainly containing VSG 3' ends, and 9% of the VSGs are atypical in some way, in that they lack features that would be expected on functional VSGs [19]. A second surprising feature of the T. brucei genome sequence is that all of these VSG genes are present in arrays ranging in size from three to 250 copies, which are mainly located at subtelomeres [19,22]. The majority of the VSG genes analysed have a strand orientation pointing away from the telomere repeats, unlike the VSGs present within VSG expression site transcription units, which point in the direction of the telomere repeats.

How could all of this VSG gene degeneracy have been missed for so long? Virtually all researchers studying antigenic variation in African trypanosomes have for technical reasons concentrated on VSG switches occurring relatively early in an infection. Typically, single relapses (or switches away from one VSG gene) are analysed. Analysis of VSG switch events occurring during chronic infections is complicated by the fact that it can be impossible to determine whether a switch variant has arisen from a previous antigenic peak, or was already present as a minor variant within the infection. One consequence of this experimental bias for the analysis of early VSG switch events has been to concentrate attention on the most
frequent VSG switch events, and on that portion of the VSG repertoire composed of intact genes most likely to be activated early in infection.

By contrast, because complicated segmental gene conversion reactions leading to the assembly of mosaic VSG genes in an active VSG expression site seem to occur much less frequently, VSG pseudogenes are probably underrepresented in the earliest switch events, and have been studied in relatively few chronic infections in *T. brucei* and the closely related *T. equiperdum* [14,15,29–31]. The fact that segmental gene conversions involving VSG pseudogenes have still been detected in experimental infections, in some cases within weeks of inoculation [29–31], suggests that the contribution of VSG pseudogenes to antigenic variation over the course of a chronic infection could in fact be much greater than previously believed. Different VSG pseudogenes can be sewn together in different ways by segmental gene conversion events, resulting in the creation of a large range of different mosaic VSGs from a limited set of VSG pseudogenes [17,18,32] (Figure 3b). This vastly increases the potential for generation of antigenically diverse VSGs within a chronic infection.

A complication with performing a definitive analysis of antigenic variation in *T. brucei* is that the genome sequence as published is incomplete [19]. Missing are the chromosome ends, which are expected to include all of the estimated 20 *T. brucei* bloodstream-form VSG expression sites and 25 metacyclic VSG expression sites.

**Figure 3.** VSG switching mediated by gene conversion. (a) VSG switching mediated by gene conversion frequently relies on areas of homology upstream and within the 3’ conserved region of VSG genes. VSG genes are indicated with filled coloured boxes, with the particularly conserved 3’ region indicated with black boxes. The telomere repeats located at the end of the active VSG expression site are indicated with triangles. Upstream of both the donor and acceptor VSGs are variable numbers of characteristic 70-bp repeat sequences (vertically hatched boxes). In the active VSG expression sites these 70-bp repeat sequences are typically present as large arrays. Gene conversion results in the duplication of the silent VSG (red box labelled ‘B’) into the active VSG expression site, replacing the previously active VSG (blue box labelled ‘C’).

(b) Mosaic VSGs can be created through the segmental gene conversion of multiple VSG pseudogenes. The figure shows data modified, with permission, from Ref. [32] where different donor VSG pseudogenes are indicated above as coloured boxes, with vertical lines indicating stop codons. Below are schematics of different expressed mosaic VSGs, where segments corresponding to sequences from different VSG pseudogene donors are indicated as coloured regions. This indicates the potential for the assembly of a given set of VSG pseudogenes into a huge range of different chimeric VSGs. (The names down the right-hand side are assigned by Roth et al. [32].)
In addition, the minichromosomes are absent, which can be expected to harbour a total of 100–200 VSG genes depending on whether VSGs are always located at each end. Furthermore, the data as published do not contain complete assemblies of all of the silent VSG gene arrays [19] (Trypanosoma brucei GeneDB: www.genedb.org/genedb/tryp). It is therefore still unclear exactly how many VSG genes T. brucei has, and what percentage of this total is intact.

**How did the VSG pseudogene repertoire accumulate?**

Whatever the exact proportion, the fraction of VSGs that are pseudogenes is clearly very high. Possibly, this is a consequence of genetic drift, whereby relaxed selection on those VSG genes that are activated relatively infrequently enables disruptive mutations to accumulate. Indeed, several features of trypanosome biology might predispose this organism to the accumulation of deleterious mutations in its VSG repertoire. The role of genetic drift in trypanosome evolution is probably enhanced both by the repeated occurrence of selective sweeps of antigenic variants escaping host immune responses, and by population bottlenecks during transmission between the mammalian host and insect vector. Epidemic disease dynamics, as have been reported for human trypanosomiasis [33], will also reduce the effective population size and promote the fixation of deleterious mutations.

In addition, the activation rate of a VSG gene probably influences the selective cost of null mutations. Even if the total rate of VSG switching is relatively high, the activation rate of any single VSG gene might be too small for selection against null mutations to be strong enough to maintain 1000–2000 intact VSG genes. If the distribution of activation rates of VSG genes is skewed so that a large fraction of VSGs are activated infrequently, the strength of selection for intact VSG genes could be comparably skewed. This would increase the proportion of VSG genes containing null mutations.

Another feature that might facilitate the accumulation of pseudogenes in the VSG repertoire is the segmental aneuploidy of the megabase chromosome ends, including the VSG arrays. Although trypanosomes are diploid with respect to their megabase chromosomes, hybridization studies have shown that homology does not extend into the chromosome ends [34]. That this is also true of the VSG arrays is now evident from the T. brucei genome project, which has provided the sequences of corresponding alleles of the right ends of chromosomes V and VIII. Comparison of the VSG arrays on matching alleles, for example all VSGs on one homologue of chromosome V with all VSGs on the other chromosome V homologue, reveals that most VSG genes share little homology in their variable N-terminal domains with any of the VSG genes on the corresponding chromosome allele [19] (Trypanosoma brucei GeneDB: www.genedb.org/genedb/tryp). If the absence of similar VSG alleles on homologous chromosomes results in genes in these locations being repaired less efficiently, for example because there is no template available for mismatch repair, mutation rates could be greater in these regions. Because random mutations are more likely to convert a functional gene into a pseudogene than vice versa, any factor that increases the mutation rate is also likely to increase the rate at which nonessential genes become pseudogenes.

Lastly, it is possible that VSG pseudogenes arise as by-products of processes facilitating diversification of VSG genes within the VSG arrays. Because expressed VSGs bear multiple conformational antibody epitopes [2], generation of an antigenically novel protein might require radical changes to the coding sequence of a VSG that has already been expressed. One way to rapidly generate such changes would be by means of insertion-deletion events, which cause part of the gene to be frameshifted. Although diversification of VSG genes in this manner would carry an appreciable risk of generating pseudogenes containing nonsense mutations, the ability of the trypanosome to assemble functional and antigenically novel VSGs from these pseudogenes might enable null mutations to accumulate in the VSG gene repertoire without reducing its breadth.

**Could a repertoire of VSG pseudogenes be advantageous?**

Rather than being accumulated junk, having such a large repertoire of ‘broken’ VSG genes could have a distinct advantage for an organism that relies on the generation of antigenic diversity to ensure its survival during chronic infection. One consequence of having so many dysfunctional VSG genes is that once the intact VSG genes have been exhausted in the course of an infection, the trypanosome can only survive by using segmental gene conversion of pseudogenes to create an intact VSG. This creation of novel mosaic VSG genes in the active VSG expression site would result in the production of large numbers of potential VSG switch variants from a limited number of VSG pseudogenes, because the same VSG gene segments could be assembled in different ways during the course of the same infection. If the rates of segmental gene conversion are high enough, this ability to create large numbers of novel mosaic VSG variants could extend the duration of individual infections almost indefinitely and facilitate the superinfection of new hosts, even if these have been infected by the same strain of trypanosome [20].

In addition, the presence of VSG pseudogenes could have a role in staggering the preferential hierarchy of VSG activation. VSG genes are activated in a ‘semipredictable’ fashion during a chronic infection (discussed in Ref. [35]). Radically different activation frequencies have the effect of ‘staggering’ which VSG genes are activated during a chronic infection. The conversion of VSG genes into pseudogenes could have been directly selected for to increase the skew in VSG activation frequencies [14,15]. At a population level, a clear hierarchy of VSG gene activation reduces the heterogeneity of VSG variants that an infected host would be exposed to at any given time during the course of a chronic infection and potentially extends the duration of the infection [36].

Where are these new chimeric VSG genes made? A consequence of a gene conversion event moving a pseudogene into an active VSG expression site through duplicative transposition would be a stop in VSG synthesis leading to a cell-cycle arrest [1]. Additional rounds of gene
conversions could repair this VSG pseudogene and rescue the cell. However, this admittedly far-fetched scenario would only work if the cell undergoes a ‘hyperrecombinogenic’ phase before it dies from lack of VSG. Alternatively, and perhaps more likely, VSG genes could be sewn together in a nontranscribed VSG expression site, whereby successive rounds of segmental gene conversion and expression site activation could eventually select for a functional VSG gene and a productive switch [2].

Maintenance of newly generated VSG diversity

Gene conversions sewing together VSG pseudogenes to create novel mosaic VSGs have the potential for generating enormous amounts of VSG gene diversity. How is this diversity maintained within the trypanosome genome if a newly created chimeric VSG gene present within a VSG expression site is destroyed by the incoming VSG during a switch? In theory, gene conversions could move newly created VSG genes back into the VSG arrays. However, this has never been experimentally detected. The relative paucity of intact VSG genes in the VSG arrays suggests that this movement cannot be frequent, at least relative to the rate at which the VSG genes are subsequently degraded. A transcriptional switch to another VSG expression site would also result in the temporary preservation of the newly created mosaic VSG. However, VSG expression sites are relatively unstable repositories for newly created VSGs, because of the selection pressure for gene conversions replacing their telomeric VSG once activated.

Another possibility is that telomere exchange, which occurs regularly during infections [27,37], has an important role in the maintenance of newly generated VSG gene diversity. This mechanism of VSG switching would result in shuttling newly created VSG genes within the active VSG expression site into another telomere, including those of the multitude of nontranscribed minichromosomes. Because there is no selection pressure operating on minichromosome telomeres for VSG switching, movement to this location would enable new chimeric VSG genes to be retained for future switch events. If this exchange is occurring frequently, one would expect that the VSG genes present at the telomeres of minichromosomes would be all or mainly intact, unlike VSG genes present within tandem arrays on megabase chromosomes. The completed sequence of the entire T. brucei genome, including all of the chromosome ends, will tell us if this is the case. If this scenario is correct, an important function of the minichromosomes, in addition to providing a large pool of recombinogenic telomeric VSG genes, would be as a reservoir preserving newly created mosaic VSGs. This would enable trypanosomes at the population level to undergo relatively rapid change of at least part of their VSG gene repertoire.

VSGs and genome architecture

A striking feature of the organization of the VSG gene arrays is that they are subtelomeric. Telomeres seem to be particularly recombinogenic areas of genomes, as has been demonstrated in humans and the malaria parasite Plasmodium [38,39]. This makes chromosome ends a good place to put genes where diversity needs to be generated, as is the case for gene families involved in antigenic variation or host adaptation of the parasitic protozoa [40].

However, for antigenic variation to work, there needs to be mutually exclusive expression of a single VSG gene. There is no evidence for regulation of RNA polymerase II transcription in trypanosomes, and life-cycle-specific regulation of gene expression seems to be almost exclusively post-transcriptional [41]. T. brucei chromosomes seem to be transcribed to a great extent over their entirety as a limited number of large polycistronic transcription units [19,42,43]. How does the trypanosome ensure that its inactive VSG genes are kept silent?

There are presumably multiple mechanisms ensuring that the inactive VSG genes are not constitutively expressed. First, maintenance of the preponderance of silent VSG genes as pseudogenes reduces the risk that functional VSG transcripts are produced from genes residing within the VSG arrays. Second, most of the silent VSG genes present within arrays are either inverted with respect to neighbouring housekeeping gene transcription units or are separated from such units by inverted stretches. However, this is not always the case. For example, there is a tandem array of five VSG genes and pseudogenes that occupies a chromosome internal position on chromosome VII. It is not known if these genes are transcribed as part of the adjacent polycistronic array [19]. Possibly, ‘boundary elements’ are present at the borders between VSG gene arrays and upstream polycistronic transcription units ensuring efficient termination of RNA polymerase II. It will be an experimental challenge to characterize these putative elements.

Future challenges

T. brucei is not alone in making use of pseudogenes to produce new antigenic variants – this mechanism has also been found in prokaryotic pathogens as diverse as Anaplasma marginale, Borrelia burgdorferi, Neisseria gonorrhoeae and Mycoplasma synoviae [44–47]. However, the size of the T. brucei VSG pseudogene repertoire seems unprecedented in the biology of antigenic variation, with potentially far-reaching consequences for the coding potential of the trypanosome genome and its rate of diversification, and for both clinical and epidemiological features of African trypanosomiasis. However, a full evaluation of antigenic variation in African trypanosomes awaits a completed T. brucei genome sequence. A priority will be the cloning and sequencing of all of the different chromosome ends, in addition to the intermediate chromosomes and minichromosomes. How large is this VSG gene repertoire, and what percentage is intact? Is there evidence for pools of VSG genes with different degrees of intactness in different genomic locations? Where are the new mosaic VSG genes pasted together by segmental gene conversions being made? What mechanisms are ensuring that VSG gene diversity is continually being generated and then retained for future use? To what extent does the size, content and genomic location of the VSG gene repertoire differ among African trypanosome lineages, and how does this structural variation affect the epidemiology of different trypanosome subspecies? Hopefully, the answers to these questions will help us to understand how possessing this
extensive wardrobe of VSG coats of varying qualities enables African trypanosomes to be such effective masters of disguise.

Acknowledgements
We would like to thank Bill Wickstead, Keith Gull, Michael Ginger, Kate Harley, Mani Narayanan, Steve Terry, Matt Wand and Rosie Young for stimulating discussions and comments on the manuscript. G.R. is a Wellcome Senior Fellow in the Basic Biomedical Sciences. This research is funded by the Wellcome Trust.

References
6 Chaves, I. et al. (1999) Control of variant surface glycoprotein gene-expression sites in Trypanosoma brucei. EMBO J. 18, 4846–4855
10 Bernards, A. et al. (1981) Activation of trypanosome surface glycoprotein genes involves a duplication-transposition leading to an altered 3′ end. Cell 27, 497–505
13 Pays, E. et al. (1985) Trypanosoma brucei: the extent of conversion in antigen genes may be related to the DNA coding specificity. Cell 42, 821–829
22 Van der Plogt et al. (1982) An analysis of cosmid clones of nuclear DNA from Trypanosoma brucei shows that the genes for variant surface glycoproteins are clustered in the genome. Nucleic Acids Res. 10, 9005–9023
37 Rudenko, G. et al. (1996) Telomere exchange can be an important mechanism of variant surface glycoprotein gene switching in Trypanosoma brucei. Mol. Biochem. Parasitol. 80, 65–75
39 Linardopoulos, E.V. et al. (2005) Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. Nature 437, 94–100