Cell, Vol. 109, 5–8, April 5, 2002, Copyright ©2002 by Cell Press

Antigenic Variation and Allelic Exclusion

Minireview

Piet Borst[1]
The Netherlands Cancer Institute
Division of Molecular Biology and Center for Biomedical Genetics
Plesmanlaan 121
1066 CX Amsterdam
The Netherlands

Cells often express only one gene from a set of two or more. African trypanosomes appear to accomplish this monoallelic expression by segregating the selected gene into a specific nuclear body. The possibility that such a structure might explain monoallelic expression in other multigene systems is discussed here.

Variation is the spice of life and a prerequisite for survival in a competitive world. Diversity can be generated by errors in nucleic acid replication or repair but can also be programmed, as exemplified by African trypanosomes. These unicellular eukaryotic parasites multiply freely in the bloodstream of their mammalian host and can cause deadly disease in humans (sleeping sickness) and in other mammals. Trypanosomes escape total destruction by the host immune system by regularly changing their coat (reviewed in Borst and Ulbert, 2001; Pays et al., 2001; Cross et al., 1998; Barry and McCulloch, 2001; Borst, 2002). The gene for the main coat protein, the variant surface glycoprotein (VSG), is transcribed from a long transcription unit located at the end of a chromosome, a telomeric expression site (ES). As illustrated in Figure 1A, the trypanosome can change its coat by replacing the transcribed *VSG* gene in the ES by a different *VSG* gene. Most potential donor *VSG* genes are clustered in nontelomeric chromosome regions, and these "chromosome-internal" genes can be transposed to an active ES by a gene conversion event that displaces the resident gene. The large repertoire of about 10^3 chromosome-internal *VSG* genes is further expanded by some 100 minichromosomes of 50–100 kb carrying a *VSG* gene at their termini. These genes can enter the active ES either by gene conversion or by reciprocal recombination (Figure 1A). All these DNA rearrangements appear to be directed by short blocks of sequence homology adjacent to donor and target genes.

It should be clear from Figure 1A that a single telomeric ES would suffice to express the entire repertoire of *VSG* genes. It therefore came as a surprise when *Trypanosoma brucei* was found to have about 20 different ESs, which look alike in overall topography and in the sequence of their promoter region. Trypanosomes can switch at low frequency from one ES to another (Figure 1B).

ESs do not serve primarily to provide an alternative way to switch coat but are used to express different sets of expression site-associated genes, which are cotranscribed with the *VSG* gene (Figure 1). Two of these genes encode the subunits of a heterodimeric transferrin receptor (Tf-R). Trypanosomes can, therefore, make as many as 20 slightly different Tf-Rs, and it has been shown that these differences can strongly affect the affinity for the Tfs of different mammals. An attractive idea is that this diversity of Tf-Rs allows the trypanosome to deal with the diversity of Tfs in its mammalian hosts and has therefore helped the trypanosome to extend its host range (Bitter et al., 1998).

The multiplicity of ESs adds an additional layer of complexity to antigenic variation and raises several questions: how does the trypanosome choose one out of 20 expression sites to be active? How does it keep the other 19 inactive (allelic exclusion)? How does it switch from one expression site to the next? These questions are of concern to more than just trypanosomologists, because, as illustrated in Table 1, several other pathogenic protozoa use a transcriptional mechanism for generating antigenic variation. Even though gene replacement is the dominant mechanism for variation of surface antigens in trypanosomes and in *Pneumocystis* (Borst, 2002; Stringer and Keely, 2001), *Plasmodium* and *Giardia* use a transcriptional (epigenetic) control mechanism. In mammals, we are confronted with the same questions, for instance, when a single olfactory receptor gene is selected for activation out of a family of 1000 related genes (Kratz et al., 2002).

How allelic exclusion works in such complex systems is not known, and it may be naive to expect the same mechanism for this process in trypanosomes and in humans. Several different mechanisms are used in other instances where genes need to be repressed, e.g., in genomic imprinting (Sleutels and Barlow, 2002). Nevertheless, it may be useful to examine whether the recent findings of Navarro and Gull (2001) on selection of a single transcribed *VSG* gene in trypanosomes might also apply to monoallelic gene expression elsewhere in nature.

Early work on the *VSG* ESs of African trypanosomes had established that the *VSG* genes are not transcribed by RNA polymerase II, the polymerase transcribing most other trypanosomal protein coding genes, but by a polymerase highly resistant to α-amanitin, presumably RNA polymerase I (Pol I). In other organisms, Pol I resides in the nucleolus and only transcribes the rRNA genes. *VSG* genes are not transcribed in the nucleolus, however, but somewhere in the nucleoplasm. This "somewhere" has now been defined more precisely by Navarro and Gull (2001). In a series of elegant experiments, they show that the active *VSG* ES is present in a nuclear body, the expression site body (ESB), which stains with antibodies against Pol I and labels with BrUTP in the presence of high concentrations of α-amanitin, sufficient to inhibit RNA polymerases II and III (Figure 2). The ESB can be detected in postmitotic nuclei before cytokinesis is complete, suggesting that each daughter cell inherits an ESB with an attached active ES. The ESB can also explain why trypanosomes trapped in the act of switching their ES contain two active ESs close together (Chaves et al., 1999). The ESB is not an accessory nucleolus, as it does

[1]Correspondence: p.borst@nki.nl

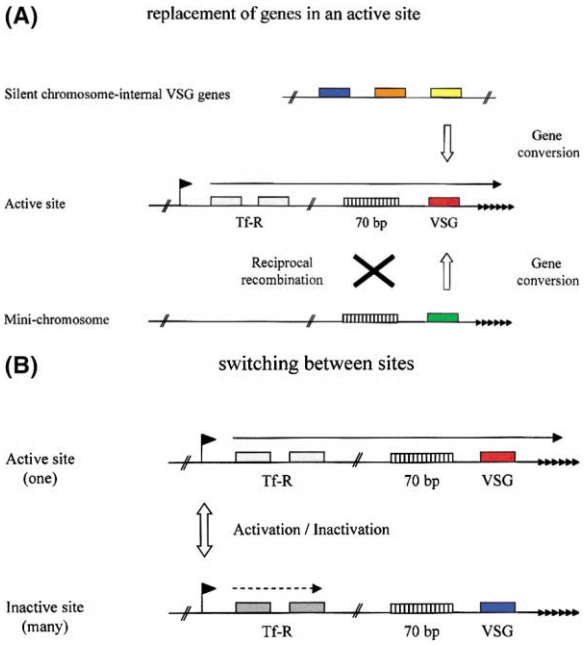

Figure 1. Scheme of a Variant Surface Glycoprotein (VSG) Gene Expression Site of *T. brucei*

VSG gene expression site of *Trypanosoma brucei*, indicating the various ways in which VSG genes can be replaced in an active site (A) and the switching between sites (B). The active site is transcribed from the promoter (flag) down to the VSG gene (genes represented by colored blocks, transcript represented by continuous line). The inactive expression site is only partially transcribed (broken line). Abbreviations: Tf-R, genes encoding the two subunits of the heterodimeric transferrin receptor; arrowheads, telomeric repeats; 70 bp, imperfect 70 bp repeats. For background information, see Borst and Ulbert, 2001; Pays et al., 2001; Cross et al., 1998; Barry and McCulloch, 2001; Borst, 2002.

not stain with antibodies against a nucleolar protein, fibrillarin. Navarro and Gull emphasize that the ESB is not just an ad hoc assembly of components required to transcribe the active ES either, because the ESB survives removal of DNA with DNase I. It will be necessary, however, to demonstrate that the ESB contains unique structural components in order to prove that the ESB is a specific subnuclear structure existing in the absence of the active ES and its transcripts.

The model proposed by Navarro and Gull (2001) for ES control is a simple one: monoallelic expression is guaranteed, because there is only space for one ES in each ESB. This would explain the observations that

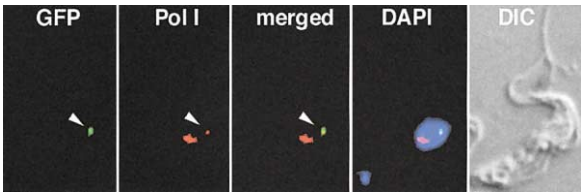

Figure 2. Expression Site Body (ESB) of *T. brucei*

The ESB (arrowhead) is visualized by location of the active telomeric expression site (by tagging with GFP; green) and a monoclonal antibody against RNA polymerase I (Pol I) (red). The anti-Pol I antibody also detects the larger nucleolus. DNA is stained with DAPI (4′,6-diamino-2-phenylindole dihydrochloride), revealing the nucleus and the smaller kinetoplast (mitochondrial DNA). DIC, differential interference contrast image of the cell (courtesy of Navarro and Gull).

trypanosomes cannot stably maintain two fully active ESs and that the inactivation of an active ES is coupled to the activation of a silent one (Chaves et al., 1999). Allelic exclusion is stringent, indeed. An interesting complication is that the "silent" ESs are not completely silent, but that the upstream part of the long transcription unit is transcribed to some extent (Figure 1B). Therefore, silent ESs are transcription competent and although transcription is initiated, it is not maintained. Only within the ESB can an ES avoid transcriptional attenuation and efficiently complete transcription of the entire ES. Silent ESs appear to be randomly distributed in the nucleus. They do not seem to be relegated to a specific silencing compartment of the nucleus, e.g., the area close to the nuclear membrane, as observed for silenced genes in yeast.

The "privileged location" model for the active ES has been proposed before, and some of the old questions remain: what is the nature of the privileged site? How do ESs get to this site and stay there? How is an ES replaced at low frequency in the site? Does this happen during DNA replication/mitosis/cytokinesis? How does the trypanosome prevent stable occupation of the site by two ESs? Clearly the ES body still needs hands and feet.

Nevertheless, one may ask whether an ESB could also explain other examples of monoallelic expression of multigene families, such as the *var* genes of *Plasmodium* (see Table 1). Recent work on the *var* gene family shows that the 50 *var* genes are scattered over several chromosomes and are transcriptionally controlled by an epigenetic mechanism (Deitsch et al., 2001; Scherf et al.,

Table 1. Antigenic Variation of Eukaryotic Parasites: Mechanisms Used to Control Expression of Surface Antigen Genes

Organism	Surface Antigen	Number of Genes (approx.)	Mechanism of Expression Control
Trypanosoma brucei	VSG	1000	Gene replacement (major) Transcriptional (minor)
Plasmodium falciparum[a]	PfEMP1 (var)	50	Transcriptional
Pneumocystis carinii	MSGI	100	Gene replacement
	MSGII	>20	Transcriptional?
Giardia lamblia	VSP	150	Transcriptional

Abbreviations: VSG, variant surface glycoprotein; PfEMP1, *Plasmodium falciparum* erythrocyte membrane protein 1; MSG, major surface glycoprotein; VSP, variant specific surface protein. See text for details. This table was modified from Borst, 2002.
[a]The PfEMP1 is not on the surface of the parasite, but on the surface of the erythrocyte in which it resides.

1998). If *var* genes are placed on a plasmid, they are only silenced when the conserved intron normally present in *var* genes is intact (Deitsch et al., 2001). Interestingly, full repression requires passage of the parasite through S phase (Deitsch et al., 2001), suggesting that the assembly of a transcriptionally silent chromatin structure occurs on freshly replicated DNA, as also observed in other systems. How this epigenetic silence is broken when a *var* gene is activated is not known. DNA rearrangements do not appear to be involved (Deitsch et al., 2001; Scherf et al., 1998), and entry into a privileged site, e.g., an ESB, could explain activation.

Another multigene family studied intensively is the olfactory receptor (OR) gene family, which is as complex as the trypanosome VSG gene family, occupying about 1% of the mammalian genome. The 1296 murine OR genes are distributed over nearly all mouse chromosomes (Zhang and Firestein, 2002), and each olfactory neuron only expresses a single OR gene (see Kratz et al., 2002). Since the OR genes are diploid, the neuron is not only able to select 1 out of 1000 different genes, but also (at random) one of the two alleles. An extra allele introduced in the germline can be exclusively activated in some neurons without activation of the corresponding endogenous genes (Serizawa et al., 2000). Three models have been considered to explain this "most enigmatic" selection mechanism for OR genes. (1) The selected OR gene may be transposed into an ES. There is no evidence for this mechanism. (2) Each OR gene may have a different set of regulatory motifs recognized by several different combinations of activators. The transgene experiments mentioned above do not support this complex model, as an extra OR gene allele can be activated without activation of the identical endogenous allele. (3) Finally, Lane et al. have invoked a mechanism wherein "a single OR transcription complex resides within olfactory neurons that can stably accommodate only a single OR gene" (Lane et al., 2001). This is very similar to the trypanosome ESB discussed here.

In summary, gene families may either be controlled by a cassette mechanism, in which genes are slotted by DNA recombination into an ES, or at the transcriptional level. Transcriptional control could occur in any of four ways. (1) By the activation/inactivation of a promoter through DNA rearrangements. There is no evidence for this mechanism in any of the multigene families discussed here. (2) By a mechanism involving a locus control region, as deduced for the β-globin locus in mammals (Grosveld, 1999). Although invoked for more complex gene families than the β-globin gene family (Kratz et al., 2002), this mechanism cannot easily explain the control of gene families widely scattered over multiple chromosomes. (3) By negative epigenetic control, such as telomeric silencing. This mechanism is suitable for stochastic activation/inactivation of genes, but its hallmark is that each gene is controlled independently. Hence, it should be possible to select for simultaneous activation of multiple genes. This mechanism is unsuitable for large gene families in which expression needs to be limited to one gene at a time. It has been shown not to apply in the case of trypanosome ESs (see Borst and Ulbert, 2001). (4) By positive epigenetic control, involving a limiting factor required for activity. This factor could either be a (set of) limiting diffusible proteins or a privileged site, such as the ESB. By introducing an element of cooperativity into such a protein assembly, it could be used to restrict gene activity to one at a time. To avoid accidental activation of genes, this mechanism could involve active repression of silent copies, only relieved by the positive regulatory element. Passage through S-phase would require reestablishment of the privileged position of the chosen gene, with a small chance that another one would take over in the case of parasites that can switch the active gene.

For a unicellular organism, not at risk for cancer, gene replacement is a fine mechanism for antigenic variation. The two eukaryotes that use this mechanism extensively, *Trypanosoma* and *Pneumocystis*, both use telomeric expression sites. The telomeric position has the advantage that new genes can be brought in by reciprocal translocation or gene conversion if the donor genes are also telomeric (Borst, 2002; Stringer and Keely, 2001). In a subtelomeric location, (silent) variant antigen genes may also be exposed to elevated mutation rates, leading to increased rates of diversification (Freitas-Junior et al., 2000). In contrast, DNA rearrangement is a risky mechanism for multicellular organisms with large genomes, as each DNA rearrangement carries the risk of a misrearrangement, resulting in cancer. This is amply demonstrated by the high frequency of B and T cell lymphomas produced as a side effect of immunoglobulin and T cell receptor gene formation. Hence, the preference of mammals for epigenetic (transcriptional) controls (Chess, 1998; Ohlsson et al., 1998). The African trypanosome may provide a clue how such controls work.

Selected Reading

Barry, J.D., and McCulloch, R. (2001). Adv. Parasitol. *49*, 1–70.

Bitter, W., Gerrits, H., Kieft, R., and Borst, P. (1998). Nature *391*, 499–502.

Borst, P. (2002). In Antigenic Variation to Mobile DNA II, L.C. Craig, R. Craigie, M. Gellert, and A. Lambowitz, eds. (Washington, DC: ASM Press), pp. 953–971.

Borst, P., and Ulbert, S. (2001). Mol. Biochem. Parasitol. *114*, 17–27.

Chaves, I., Rudenko, G., Dirks-Mulder, A., Cross, M., and Borst, P. (1999). EMBO J. *18*, 4846–4855.

Chess, A. (1998). Science *279*, 2067–2068.

Cross, G.A.M., Wirtz, L.E., and Navarro, M. (1998). Mol. Biochem. Parasitol. *91*, 77–91.

Deitsch, K.W., Calderwood, M.S., and Wellems, T.E. (2001). Nature *412*, 875–876.

Freitas-Junior, L.H., Bottius, E., Pirrit, L.A., Deitsch, K., Scheidig, C., Guinet, F., Nehrbass, U., Wellems, T., and Scherf, A. (2000). Nature *407*, 1018–1022.

Grosveld, F. (1999). Curr. Opin. Genet. Dev. *9*, 152–157.

Kratz, E., Dugas, J.C., and Ngai, J. (2002). Trends Genet. *18*, 29–34.

Lane, R.P., Cutforth, T., Young, J., Athanasiou, M., Friedman, C., Rowen, L., Evans, G., Axel, R., Hood, L., and Trask, B.J. (2001). Proc. Natl. Acad. Sci. USA *98*, 7390–7395.

Navarro, M., and Gull, K. (2001). Nature *414*, 759–763.

Ohlsson, R., Tycko, B., and Sapienza, C. (1998). Trends Genet. *14*, 435–438.

Pays, E., Lips, S., Nolan, D., Vanhamme, L., and Pérez-Morga, D. (2001). Mol. Biochem. Parasitol. *114*, 1–16.

Scherf, A., Hernandez-Rivas, R., Buffet, P., Bottius, E., Benatar, C.,

Pouvelle, B., Gysin, J., and Lanzer, M. (1998). EMBO J. *17*, 5418–5426.

Serizawa, S., Ishii, T., Nakatani, H., Tsuboi, A., Nagawa, F., Asano, M., Sudo, K., Sakagami, J., Sakano, H., Ijiri, T., et al. (2000). Nat. Neurosci. *3*, 687–693.

Sleutels, F., and Barlow, D.P. (2002). Adv. Genet. *46*, 119–163.

Stringer, J.R., and Keely, S.P. (2001). Infect. Immun. *69*, 627–639.

Zhang, X., and Firestein, S. (2002). Nat. Neurosci. *5*, 124–133.