1 ***VSG* mRNA levels are regulated by the production of functional VSG protein**

2

3 Isabella E. Maudlin[a], Steve Kelly[b], Angela Schwede[a*] and Mark Carrington[a*]

4

5 [a] Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge

6 CB2 1QW, United Kingdom.

7

8 [b] Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1

9 3RB, UK

10

11 * emails for correspondence: angela.schwede@gmx.de or mc115@cam.ac.uk

12

13 Highlights:

14 • *VSG* mRNA copy number varies with the identity of the VSG

15 • Early premature termination codons result in degradation of *VSG* mRNA

16 • Late premature termination codons translated to produce incomplete VSG result in

17 an increase in *VSG* mRNA

18 • There is a feedback pathway between non-functional VSG production and *VSG*

19 mRNA levels

20

21 Key words:

22 *Trypanosoma brucei*, VSG mRNA

Abstract

The bloodstream form of *Trypanosoma brucei* persists in mammalian hosts through a population survival strategy depending on antigenic variation of a cell surface coat composed of the variant surface glycoprotein (VSG). The integrity of the VSG coat is essential and blocking its synthesis results in a cell division cycle arrest just prior to cytokinesis. This observation indicates that VSG levels are monitored and that the cell has mechanisms to respond to a disruption of synthesis. Here, the regulation of *VSG* mRNA levels has been investigated by first measuring VSG mRNA copy number, and second using ectopic expression of VSG transgenes containing premature termination codons. The findings are that (i) *VSG* mRNA copy number varies with the identity of the VSG and (ii) a pathway detects synthesis of non-functional VSG protein and results in an increase in *VSG* mRNA levels.

1. Introduction

Trypanosoma brucei can sustain long-term infections in mammalian hosts by antigenic variation of the variant surface glycoprotein (VSG) that forms a coat covering the entire surface of the cell [1]. The active *VSG* gene is located near a telomere at the 3' end of a transcription unit known as a bloodstream expression site (BES). There are 10 to 20 BESs in *T. brucei* genomes but only one BES is transcribed at any one time [2, 3]. Antigenic variation occurs when a low frequency event, either gene conversion in the active BES or a less well characterised epigenetic switch to an alternative BES, results in transcription of a different *VSG* gene [4]. The processes involved in antigenic variation and monoallelic VSG expression have been studied in some detail, the active BES is transcribed by RNA polymerase I (RNAPI) [5] and is located in the nuclear expression site body (ESB), a compartment containing RNAPI lying outside the nucleolus [6]. Expression of protein coding genes by RNAPI evolved in a cell in which the maturation of every cytoplasmic mRNA included *trans*-splicing of a capped 39 nucleotide exon to the 5' end [7], thus the first 39 bases of a *VSG* mRNA are transcribed by RNAPII and the remainder by RNAPI. Presumably, RNAPI transcription allows the exceptionally high levels of *VSG* mRNA in the cell to be transcribed from a single gene. The precise details of VSG monoallelic expression, both the default silencing and the escape from silencing by the active BES, are less well understood but involve a VEX/CAF complex identified in a whole genome RNAi screen for loss of monoallelic VSG expression [3, 8].

There is strong evidence that mechanisms have evolved to ensure that the supply of

59 newly synthesised VSG is adequate to maintain the densely packed coat as new
60 plasma membrane is added during cellular growth. Blocking VSG synthesis by RNAi
61 knockdown of the mRNA or by blocking translation using a morpholino oligonucleotide
62 results in cell division cycle arrest at the onset of cytokinesis [9, 10]. It is assumed that
63 cytokinesis is the point in the cell cycle where there is the highest demand for new VSG
64 coated membrane. Trypanosomes with a common genetic background but expressing
65 different VSGs proliferate at different rates in identical culture conditions, providing
66 evidence that the identity of the expressed VSG can affect cellular growth rates with
67 some VSGs effectively constraining proliferation [11]. As VSGs constitute 10 to 20% of
68 total protein and VSGs have very different amino acid sequences, the rate of successful
69 folding and export from the endoplasmic reticulum may be limiting, thus affecting growth
70 rates. Studies have reached mixed conclusions, either that VSG protein is synthesised
71 in amounts greater than is needed and the excess degraded [12, 13] whereas another
72 report found that just sufficient is synthesised [14].
73
74 As might be expected for an mRNA encoding a highly abundant protein, *VSG* mRNA
75 has a long half-life in proliferating cells with measurements ranging from 4.5 hours [15]
76 down to 1 hour [16]. Its levels can be modulated by the cell as expression of a second
77 VSG from a transgene reduces steady state mRNA levels from the endogenous *VSG*
78 gene. Inducible expression of a T7 RNA polymerase-driven *VSG6* (also known as
79 VSG121) transgene located in the non-transcribed spacer of a ribosomal RNA gene
80 cluster was shown to cause a rapid decrease in endogenous *VSG2* (also known as
81 VSG221) mRNA levels and a slower transcriptional attenuation of the active BES

82 dependent on the activity of DOT1B (disruptor of telomeric silencing-1B) [17]. It is not
83 clear whether attenuation of the active VSG expression site is an entirely intranuclear
84 event or if there is a cytoplasmic contribution. In a second set of experiments, insertion
85 of a *VSG117* gene in various locations in the genome of cells expressing VSG2 resulted
86 in a reduction of *VSG2* mRNA proportional to the amount of *VSG117* mRNA expressed
87 [16]. These observations imply that the total VSG mRNA level is regulated and that the
88 mechanism can recognise different *VSG* mRNAs despite their diverse sequences. The
89 only obvious conserved motif in *VSG* mRNAs is a 16 nucleotide sequence (16-mer) in
90 the 3'UTR close to the poly(A) tail. Although the 16-mer is necessary for the very high
91 levels of cytoplasmic *VSG* mRNA [16], no mechanism has been characterised for how it
92 is involved in positive regulation of high levels of *VSG* mRNA and it is unknown whether
93 it is also involved in down regulation when a second *VSG* mRNA is expressed.
94
95 Here, expression of a second *VSG* from transgenes containing a premature termination
96 codon (PTC) at various locations in the open reading frame was used to produce
97 cytoplasmic degradation of *VSG* mRNA [18, 19] and the consequent effect on *VSG*
98 mRNA measured. In yeast and metazoa, recognition of a PTC during translation
99 triggers a reduction in the half-life of an mRNA via the nonsense-mediated decay (NMD)
100 pathway mediated by a complex containing the RNA helicase UPF1 (reviewed in [20,
101 21]). In yeast, NMD is not a binary process and only rarely is a PTC-containing mRNA
102 completely absent from a cell (for example [22]). The degree of reduction in the PTC-
103 containing mRNAs is affected by the location of the PTC within the open reading frame,
104 the nearer to the initiation codon the more probable that NMD will be triggered on any

105 round of translation [23]. In trypanosomes, investigation of PTC-containing mRNAs
106 showed a similar relationship between PTC location in an ORF and reduction in mRNA
107 levels but any role for the UPF1 homologue remained ambiguous [24].
108
109 Expression of PTC-containing *VSG* transgenes and measurement of transgene and
110 endogenous *VSG* mRNA levels confirmed the presence of a pathway that reduced
111 PTC-containing *VSG* mRNAs, with the decrease being proportional to the proximity of
112 the PTC to the initiation codon. Measurements of the *VSG* mRNA in wild type cells
113 showed that VSG6 expressing cells have a higher copy number than VSG2 expressers
114 and that double expressers had intermediate levels. An unexpected finding was that
115 expression of *VSG* transgenes with PTCs close to the C-terminus resulted in increases
116 in both transgenic and endogenous *VSG* mRNAs. Together, these experiments provide
117 further evidence for a pathway that decreases PTC-containing mRNAs, and a second
118 pathway that regulates *VSG* mRNA levels that is linked to production of functional VSG
119 protein.
120
121
122
123

124 **2. Materials and Methods**

125 **2.1. Plasmids**

126 The sequence of the VSG6 BES was taken from Genbank entry FM162569 and VSG2
127 BES from FM162566. Genomic DNAs from yeast containing these cloned telomeres (a
128 kind gift of Gloria Rudenko) were used as templates for PCR reactions to recover the
129 fragments described below. Constructs to introduce a *VSG2* transgene were based on
130 the plasmid p3952. Digestion of p3952 with restriction enzymes Acc65I and SacI
131 released a 4.6 kbp fragment (Supp. Fig. 1A) that contained, in order: bases -800 to -300
132 upstream of the VSG6 initiation codon (5' targeting); XhoI site; bases -182 to -1
133 upstream of the VSG2 initiation codon (VSG2 splice acceptor and 5'UTR); an SphI site;
134 VSG2 ORF; PacI site; bases 1 to 615 downstream of the VSG2 termination codon
135 (VSG2 3'UTR and polyadenylation); XmaI site; a tubulin/blasticidin resistance cassette
136 (alpha to beta tubulin inter-ORF, blasticidin-S-deaminase ORF, beta to alpha tubulin
137 inter-ORF); SpeI site; bases -300 upstream to +210 downstream of the VSG6 initiation
138 codon (3' targeting) (see Supp. Fig. 1B for sequence). The plasmid was constructed by
139 joining fragments using the restriction enzyme sites listed above. Subsequent changes
140 to the VSG2 ORF used SphI and PacI. NEB Phusion DNA polymerase was used for
141 site-directed mutagenesis according to the manufacturer's instructions, oligonucleotides
142 are listed in Supp. Table 1.

143

144 **2.2. Cell culture and sample preparation**

145 *Trypanosoma brucei* Lister 427 bloodstream form trypanosomes were used as parental
146 cell lines throughout this study. HMI-11 [25] was used to culture bloodstream form

147 trypanosomes with 5 μg/ml blasticidin to maintain genetically modified cell lines. All
148 experiments were performed using logarithmically growing cells at a density of less than
149 1 x 10^6 cells/ml. Preparation of whole cell lysates, western blotting, preparation of total
150 RNA and northern blotting all used standard techniques [26] with the following details.
151 The primary antibodies used for Western blotting were: (i) Rabbit anti-VSG2 N-terminal
152 peptide [27] (a kind gift of Peter Overath), (ii) mouse monoclonal anti-VSG6 (a kind gift
153 of Miguel Navarro), and (iii) mouse monoclonal L13D6 anti-paraflagellar rod proteins
154 (PFR) 1 and 2 (a kind gift from Keith Gull). Secondary antibodies were: (i) goat anti-
155 Rabbit AlexaFluor 680, and (ii) goat anti-Mouse IRDye800. The LI-COR Odyssey
156 infrared imaging system was used to visualise western blots. For northern blots, the
157 relevant whole VSG ORF was used as a probe. Phosphorimaging was used to
158 quantitate RNA from northern blots, background from the blot was subtracted from the
159 band-of-interest, and then normalised against the rRNA loading control signal.
160

161 **2.3. RNAseq for transcript abundances**

162 Total RNA was prepared as above for single sample RNAseq to estimate *VSG* mRNA
163 abundance. The cDNA libraries were prepared and sequenced at the Beijing Genomics
164 Institute (Shenzhen, China) [28] In brief, polyadenylated RNA was purified from total
165 RNA, converted to cDNA using random hexamer primers sheared and size selected for
166 fragments ~200 bp in length using the Illumina TruSeq RNA Sample Preparation Kit v2.
167 RNAseq of the resulting libraries was used for the determination of transcript
168 abundances. Sequencing was performed on an Illumina Hiseq 2000 (Illumina, CA)
169 platform. Paired end reads were subject to quality trimming and adaptor filtering using

170 Trimmomatic [29] using the settings “LEADING:10 TRAILING:10
171 SLIDINGWINDOW:5:15 MINLEN:50”. The quality filtered paired-end reads were then
172 mapped to the complete set of CDS from version 6 of *T. brucei* genome annotation
173 using bowtie2 [30] and transcript abundances were estimated using eXpress [31]. The
174 sequence reads are in EBI ArrayExpress accession E-MTAB-9122.
175
176 **3. Results**
177 **3.1. VSG2 transgenes were expressed and PTCs were effective in terminating**
178 **translation**
179 Expression of a VSG protein is essential for the proliferation of bloodstream form
180 trypanosomes [9]. To investigate any cytoplasmic control of *VSG* mRNA levels
181 independently of cell viability, trypanosomes were made to express two *VSG* mRNAs
182 from a single bloodstream form expression site (BES), generating double expressers,
183 using a similar construct design to others [32] (Figure 1A). This approach allowed the
184 manipulation of one *VSG* mRNA whilst maintaining viability through expression of the
185 endogenous VSG. In the experiments here, the parental cell line was *Trypanosoma*
186 *brucei* Lister 427 expressing VSG6 and the transgene encoded VSG2 including the
187 native 5’ and 3’UTRs (Supplementary Figure 1).
188
189 In addition to a construct designed to express wildtype VSG2 after insertion into the
190 active VSG6 expression site, a series of VSG2 mutants were made by introducing
191 premature termination codons (PTCs) (Figure 1B). VSG2 is encoded by a 476 codon
192 open reading frame (ORF) and the mature VSG contains 433 amino acids after

193 processing which removes a 26 residue N-terminal signal sequence and a 17 residue
194 C-terminal GPI-anchor addition sequence [33]. Mature VSG2 has two folded domains,
195 the N-terminal domain (residues 27 to 377, all numbering from the nascent polypeptide),
196 an unstructured linker (378 to 400), the C-terminal domain (401 to 442) and a second
197 linker (443 to 459) with the GPI-anchor attached. The PTCs in the VSG2 open reading
198 frame were located at codons 64, 167 and 262 in the N-terminal domain, 392 in the
199 interdomain linker, 426 in the C-terminal domain and 460, the first residue of the GPI-
200 addition sequence designed to produce a mature VSG without a GPI-anchor (Figure
201 1B).
202
203 Three independent clones of each transgenic cell line were selected and any effect on
204 proliferation was measured in continuous culture with daily passage over 5 days (Figure
205 2A and Supp. Table 2). On averaging the three clones for any one cell line we observed
206 that: (i) expression of a wild type *VSG2* transgene had little effect on proliferation when
207 compared to the VSG6 expressing parental cell line, (ii) cell lines that contained the
208 transgene with a PTC at codon 262 grew more slowly with a ~10-fold reduction in
209 population growth after 5 days as measured by cumulative cell number, (iii) the three
210 cell lines with PTCs closest to the wild type stop codon (codons 392, 426 and 460) all
211 had slightly reduced proliferation over 5 days, with 2- to 3-fold reduction in population
212 growth as measured by cumulative cell number, (iv) PTCs at other locations had no
213 effect.
214
215 Expression of VSGs in the cell lines was tested by western blotting for both the

216 endogenous VSG6 and any product from the *VSG2* transgene (Figure 2B). VSG2 was
217 detected using an antiserum raised against a peptide corresponding to the N-terminal
218 15 residues of mature VSG2 [27], these residues were present in all VSG2s encoded by
219 the PTC-containing transgenes. VSG2 polypeptides of the expected molecular weight
220 were detected in cell lines containing PTCs at codons 262, 392, 426 and 460 as well as
221 more abundant expression of the wild type *VSG2* transgene. No polypeptides were
222 detected from *VSG2* transgenes with PTCs at 67 and 167 (Figure 2B), these were
223 assumed to be below the limit of detection in this experiment. These findings indicated
224 that the VSG2 transgenes were expressed as expected and that the PTCs were
225 effective in terminating translation.
226
227 **3.2. PTC-containing *VSG2* transgenes are subject to NMD and an additional**
228 **pathway that detects non-functional VSG resulting in increased VSG mRNA**
229 The expression of *VSG* mRNAs in cell lines containing each transgene was measured
230 using quantitative northern blotting (Figure 3A). mRNA levels were estimated as the
231 average of three independent clones and expressed relative to the *VSG* mRNA from
232 wild type cell lines expressing either VSG2 or VSG6 (Supp. Table 3), for example a
233 value of 0.60 represents 60% of the *VSG* mRNA present in the wild type cell line
234 expressing that specific VSG. The values were then plotted against the location of the
235 stop codon in the VSG2 transgene (Figure 3B).
236
237 The first observation from these measurements was that the insertion of the wild type
238 *VSG2* transgene resulted in the expression of two VSG mRNAs as expected, but the

239 levels of the *VSG2* and *VSG6* mRNAs were not the same, *VSG2* mRNA expression
240 was 0.39 and *VSG6* mRNA was 0.81.
241
242 This observation was investigated further by using RNAseq to estimate *VSG* mRNA
243 abundance as transcripts per million transcripts (TPM). RNA from wild type cell lines
244 expressing either VSG2 or VSG6 was used along with a cell line expressing wild type
245 VSG2 in a VSG6 background at two time points after electroporation of the transgene
246 construct. RNAseq data was processed to produce estimates of mRNA abundance as
247 transcripts per million transcripts (Table 1). In the wild type cell lines, *VSG* mRNA levels
248 varied: *VSG2* mRNA represented 15% and *VSG6* mRNA 22% of total mRNA. In the cell
249 line expressing transgenic VSG2, total *VSG* mRNA was intermediate with 18% and 20%
250 of total mRNA respectively, and the ratio of *VSG* mRNAs was 0.26:0.74 (VSG2:VSG6)
251 (Table 2).
252 The RNAseq estimates of the abundance of *VSG2* and *VSG6* mRNAs were in
253 agreement with the northern blot data. When the northern blot measurements were
254 converted to abundance estimates, the ratio of mRNAs was 0.25:0.75, similar to that
255 estimated by RNAseq (calculation in Supp. Figure 2). Thus, the relative abundance of
256 *VSG* mRNA can vary with the identity of the VSG and the decrease in *VSG6* mRNA on
257 introduction of a *VSG2* transgene indicated that there is a regulatory system able to
258 adjust the total *VSG* mRNA concentration. The adjustment could be occurring at the
259 transcriptional and/or post-transcriptional level.
260

261 Next, the expression of *VSG2* mRNAs in cell lines with PTC-containing transgenes was
262 measured by northern blotting (Figure 3 and Supp. Table 3). The expression showed a
263 position-dependent effect, with PTCs at codons 64, 167, 262, 392, 426 and 460
264 resulting in relative levels of 0.05, 0.09, 0.14, 0.31, 0.45 and 0.65 (Figure 3B). The
265 measurements of *VSG2* mRNA for the first four PTCs was as expected for an NMD
266 pathway and provides further evidence that such a pathway is present in trypanosomes.
267 However, an unexpected observation was that in the cell lines with PTCs at codons 426
268 and 460, the level of *VSG2* mRNA, 0.45 and 0.65 respectively, was greater than the
269 wild type *VSG2* transgene, 0.39.
270
271 When expression of the endogenous *VSG6* mRNA was measured, two observations
272 were made: first, expression of all the PTC-containing *VSG2* transgenes resulted in an
273 increase in VSG6 mRNA from 0.81 for the wild type to at least 1.15, more than the
274 amount present in the parental cell line (Figure 3B). Second, for VSG2 transgenes with
275 PTCs close to the C-terminus, there was a larger increase to 1.39 and 1.57 for PTCs at
276 426 and 460 respectively (Figure 3B). Thus, synthesis of a faulty VSG protein triggers a
277 general increase in *VSG* mRNA.
278
279 **4. Discussion**
280 Trypanosomes tolerate the expression of two VSGs with little effect on growth in culture
281 [32] and, as VSG expression is essential, investigations have used ectopic expression
282 of a second VSG to investigate regulation. Here, transgenes encoding *VSG* mRNAs
283 containing PTCs were used to trigger *VSG* mRNA decay in the cytoplasm and the

284 effects on both transgene and endogenous *VSG* mRNA were measured. The insertion
285 of a *VSG2* transgene upstream of the active *VSG6* gene (Figure 1A) followed the
286 original approach to successfully force the simultaneous expression of two VSGs [32],
287 and was used here to ensure that both transgenic and endogenous VSGs were co-
288 transcribed. In previous work [32], three different VSG transgenes were inserted
289 upstream of the endogenous *VSG2* gene and, in each case, high levels of expression of
290 VSG2 and the transgenic VSG were detected by immunofluorescence and western
291 blotting [32]. Here, a *VSG2* transgene was inserted upstream of the endogenous *VSG6*
292 gene and measurements of wild type transgenic and endogenous VSG expression by
293 western blotting were similar to earlier work (Figure 2B).
294
295 When the *VSG* mRNA levels were measured, the findings were: first, wild type cells
296 expressing either VSG2 or VSG6 contained different amounts of *VSG* mRNA as a
297 fraction of total mRNA: *VSG2* mRNA was 15%, and *VSG6* mRNA was 22%. Cells
298 expressing a wild type *VSG2* transgene contained intermediate levels of *VSG* mRNA
299 (Figure 3 and Table 1). These measurements could arise from a genuine difference in
300 VSG mRNA levels and/or result from differences in all other mRNAs, that is the amount
301 of VSG mRNA remains constant but all other mRNAs change. It is not possible to rule
302 out the latter but the former is more likely given probable folding efficiency variation in
303 different VSG proteins discussed below.
304
305 Second, cell lines expressing a wild type *VSG2* transgene in a VSG6 background
306 contained a decreased amount of *VSG6* mRNA. However, the cells did not contain

307 equal numbers of the two *VSG* mRNAs, the ratio was approximately 1:3 (*VSG2*:*VSG6*)
308 measured by both RNAseq and northern blotting (Figure 3 and Table 1).
309
310 Third, there was the expected position dependent response of *VSG2* mRNA levels to
311 the inclusion of a PTC, the nearer the PTC to the N-terminus the lower the steady state
312 levels of *VSG2* mRNA (Figure 3). However, reduction in *VSG2* mRNA only occurred
313 with PTCs located in the N-terminal domain or the inter-domain linker. This decrease in
314 *VSG2* mRNA resulted in an increase in *VSG6* mRNA.
315
316 Fourth, *VSG2* transgenes containing a PTC in the C-terminal domain or at codon 460,
317 which directed synthesis VSG2 without a GPI-anchor, resulted in both *VSG2* and *VSG6*
318 mRNAs increasing 1.6 to 2-fold compared to the levels in cells expressing a wild type
319 VSG2 transgene (Figure 3). These variations in copy number indicate that it is unlikely
320 that there is a direct counting mechanism for *VSG* mRNA, but rather a feedback from
321 the production of functional VSG protein.
322
323 The first two findings concern the difference in *VSG* mRNA copy number in cells
324 expressing different VSGs. Since the variation is present in the cell line expressing two
325 VSGs from the same BES, it does not result from differential transcription and must be
326 explained by post-transcriptional process(es). The major determinant of stability for
327 most mRNAs in trypanosomes is codon use, expressed numerically as a gene
328 expression codon adaptation index (geCAI) [26]. In this case, the geCAI value for *VSG2*
329 is 0.33 and *VSG6* is 0.35, and this difference might contribute towards the higher levels

330 of *VSG6* mRNA but is unlikely to be the sole determinant. The second observation that
331 expression of a second VSG results in a reduction of the endogenous *VSG* mRNA has
332 been reported before [16, 17].
333
334 The expression of VSG2 with a PTC in the N-terminal domain or inter-domain linker
335 resulted in the reduction in *VSG2* mRNA levels, inversely proportional to the length of
336 the residual open reading frame (Figure 3). This is a typical characteristic of an NMD
337 pathway [23] and has been reported before in trypanosomes [24]. The decrease in
338 *VSG2* mRNA resulted in an increase in *VSG6* mRNA to ~1.5 fold more than cells
339 expressing a full length *VSG2* transgene and ~1.2 fold more than parental cells
340 expressing VSG6. This increase to levels above that present in wild type cells indicated
341 that the trypanosome has spare capacity for the production of *VSG* mRNA and is
342 emphasised in cells expressing *VSG2* transgenes with PTCs in the C-terminal domain
343 or immediately after the mature C-terminus, so that a complete but not GPI-anchored
344 VSG was synthesised. In the latter case, both *VSG2* and *VSG6* mRNAs increased 1.6
345 to 2-fold.
346
347 One model arising from the measurements above is that *VSG* mRNA was subject to two
348 pathways. First, an NMD pathway that decreased the amount of PTC-containing mRNA
349 but became less effective when the PTC was close to the C-terminus. Second, a
350 pathway that detected unsuccessful production of membrane-anchored VSG and
351 triggered an increase in total *VSG* mRNA. The second pathway is apparent as an
352 increase in *VSG6* mRNA in cells expressing VSG2 with an early PTC and an increase

353 in both *VSG* mRNAs in cell expressing VSG2 with a late PTC, when the NMD pathway
354 is less potent.
355
356 The kinetics and route of VSG synthesis are well characterised (reviewed in [34, 35]).
357 As VSGs are translated by ER-associated ribosomes, the PTC-containing *VSG* mRNAs
358 would have directed translocation of a nascent VSG into the ER. It is likely that several
359 could not fold correctly and were potential substrates for an unfolded protein response
360 (UPR) pathway. However, there is evidence that the UPR is absent in bloodstream form
361 trypanosomes [36] and that GPI-anchorless VSG is selectively retained and degraded in
362 the lysosome [37, 38]. Our favoured model for the function of the pathway that
363 increases *VSG* mRNA in response to incorrect VSG protein is that it evolved to
364 compensate for variation in the rates of successful folding of different VSGs. If only a
365 small fraction of VSG protein folds successfully, the pathway increases *VSG* mRNA and
366 if a larger fraction of the VSG folds correctly, then less *VSG* mRNA is produced. This
367 model predicts that VSG2 folds more efficiently than VSG6. The model is also
368 consistent with the observation that bloodstream form trypanosomes seem to have a
369 capacity for degrading a large amount of non-functional VSG without any great effect on
370 proliferation.
371
372 As *VSG* mRNAs are highly variable in sequence, how is the increased mRNA that
373 occurred in response to the anchorless VSG specific to *VSG* mRNA? Different *VSG*
374 mRNAs have one specific conserved sequence feature, the '16-mer' element in the 3'
375 UTR [16] and this is an obvious candidate for mediating a specific response but the

376 experiments here do not distinguish whether this happens in the cytoplasm via mRNA
377 half-life and/or in the nucleus via transcription rate or more efficient mRNA maturation.
378
379 The strict selection pressure imposed by the mammalian immune response on
380 bloodstream trypanosomes makes it likely that VSG synthesis is tightly regulated. Here,
381 evidence is provided for sensors that detect the production of functional VSG coated
382 membranes and a response that alters *VSG* mRNA levels. The regulation of VSG
383 synthesis is central to the survival of the African trypanosome in a mammalian host and
384 as such its regulation will contain further complexities.
385
386 **5. Author contributions**

387 I.E.M. generated cell lines, performed proliferation experiments, performed sample
388 preparation and subsequent western and northern blotting and analyses. S.K.
389 performed RNA sequencing data analysis. A.S generated cell lines and prepared RNA
390 for RNAseq. M.C. and A.S. devised and supervised the study. M.C. analysed data and
391 wrote the manuscript. The final version of the manuscript was approved by all authors.
392
393 **Acknowledgements**

394 This work was funded by the Wellcome Trust award 093008/Z/10/Z. MC is a Wellcome
395 Trust Investigator 217138/Z/19/Z. We would like to thank Sarah Marsden, Markus
396 Engstler, Mark Field, Jay Bangs and Nancy Standart for reading and commenting on
397 the manuscript.
398
399

400 **References**

401 [1] M.R. Mugnier, C.E. Stebbins, F.N. Papavasiliou, Masters of Disguise: Antigenic
402 Variation and the VSG Coat in Trypanosoma brucei, PLoS Pathog 12(9) (2016)
403 e1005784.
404 [2] I. Cestari, K. Stuart, Transcriptional Regulation of Telomeric Expression Sites and
405 Antigenic Variation in Trypanosomes, Curr Genomics 19(2) (2018) 119-132.
406 [3] J. Faria, L. Glover, S. Hutchinson, C. Boehm, M.C. Field, D. Horn, Monoallelic
407 expression and epigenetic inheritance sustained by a Trypanosoma brucei variant
408 surface glycoprotein exclusion complex, Nat Commun 10(1) (2019) 3023.
409 [4] N. Sima, E.J. McLaughlin, S. Hutchinson, L. Glover, Escaping the immune system by
410 DNA repair and recombination in African trypanosomes, Open Biol 9(11) (2019)
411 190182.
412 [5] A. Gunzl, T. Bruderer, G. Laufer, B. Schimanski, L.C. Tu, H.M. Chung, P.T. Lee,
413 M.G. Lee, RNA polymerase I transcribes procyclin genes and variant surface
414 glycoprotein gene expression sites in Trypanosoma brucei, Eukaryot Cell 2(3) (2003)
415 542-51.
416 [6] M. Navarro, K. Gull, A pol I transcriptional body associated with VSG mono-allelic
417 expression in Trypanosoma brucei, Nature 414(6865) (2001) 759-63.
418 [7] C.E. Clayton, Gene expression in Kinetoplastids, Curr Opin Microbiol 32 (2016) 46-
419 51.
420 [8] L. Glover, S. Hutchinson, S. Alsford, D. Horn, VEX1 controls the allelic exclusion
421 required for antigenic variation in trypanosomes, Proc Natl Acad Sci U S A 113(26)
422 (2016) 7225-30.

423 [9] K. Sheader, S. Vaughan, J. Minchin, K. Hughes, K. Gull, G. Rudenko, Variant
424 surface glycoprotein RNA interference triggers a precytokinesis cell cycle arrest in
425 African trypanosomes, Proc Natl Acad Sci U S A 102(24) (2005) 8716-21.
426 [10] C.P. Ooi, T.K. Smith, E. Gluenz, N.V. Wand, S. Vaughan, G. Rudenko, Blocking
427 variant surface glycoprotein synthesis alters endoplasmic reticulum exit sites/Golgi
428 homeostasis in Trypanosoma brucei, Traffic 19(6) (2018) 391-405.
429 [11] D. Liu, L. Albergante, T.J. Newman, D. Horn, Faster growth with shorter antigens
430 can explain a VSG hierarchy during African trypanosome infections: a feint attack by
431 parasites, Sci Rep 8(1) (2018) 10922.
432 [12] M.C. Field, T. Sergeenko, Y.N. Wang, S. Bohm, M. Carrington, Chaperone
433 requirements for biosynthesis of the trypanosome variant surface glycoprotein, PLoS
434 One 5(1) (2010) e8468.
435 [13] Y.N. Wang, M. Wang, M.C. Field, Trypanosoma brucei: trypanosome-specific
436 endoplasmic reticulum proteins involved in variant surface glycoprotein expression, Exp
437 Parasitol 125(3) (2010) 208-21.
438 [14] C. Tiengwe, K.A. Muratore, J.D. Bangs, Surface proteins, ERAD and antigenic
439 variation in Trypanosoma brucei, Cell Microbiol 18(11) (2016) 1673-1688.
440 [15] B. Ehlers, J. Czichos, P. Overath, RNA turnover in Trypanosoma brucei, Mol Cell
441 Biol 7(3) (1987) 1242-9.
442 [16] S. Ridewood, C.P. Ooi, B. Hall, A. Trenaman, N.V. Wand, G. Sioutas, I. Scherwitzl,
443 G. Rudenko, The role of genomic location and flanking 3'UTR in the generation of
444 functional levels of variant surface glycoprotein in Trypanosoma brucei, Mol Microbiol
445 106(4) (2017) 614-634.

446 [17] C. Batram, N.G. Jones, C.J. Janzen, S.M. Markert, M. Engstler, Expression site
447 attenuation mechanistically links antigenic variation and development in Trypanosoma
448 brucei, Elife 3 (2014) e02324.
449 [18] S. Durand, J. Lykke-Andersen, Nonsense-mediated mRNA decay occurs during
450 eIF4F-dependent translation in human cells, Nat Struct Mol Biol 20(6) (2013) 702-9.
451 [19] S.C. Rufener, O. Muhlemann, eIF4E-bound mRNPs are substrates for nonsense-
452 mediated mRNA decay in mammalian cells, Nat Struct Mol Biol 20(6) (2013) 710-7.
453 [20] J. Wen, S. Brogna, Nonsense-mediated mRNA decay, Biochem Soc Trans 36(Pt 3)
454 (2008) 514-6.
455 [21] F. He, A. Jacobson, Nonsense-Mediated mRNA Decay: Degradation of Defective
456 Transcripts Is Only Part of the Story, Annu Rev Genet 49 (2015) 339-66.
457 [22] J. Wen, S. Brogna, Splicing-dependent NMD does not require the EJC in
458 Schizosaccharomyces pombe, EMBO J 29(9) (2010) 1537-51.
459 [23] R. Losson, F. Lacroute, Interference of nonsense mutations with eukaryotic
460 messenger RNA stability, Proc Natl Acad Sci U S A 76(10) (1979) 5134-7.
461 [24] P. Delhi, R. Queiroz, D. Inchaustegui, M. Carrington, C. Clayton, Is there a classical
462 nonsense-mediated decay pathway in trypanosomes?, PLoS One 6(9) (2011) e25112.
463 [25] H. Hirumi, K. Hirumi, Axenic culture of African trypanosome bloodstream forms,
464 Parasitol Today 10(2) (1994) 80-4.
465 [26] J. de Freitas Nascimento, S. Kelly, J. Sunter, M. Carrington, Codon choice directs
466 constitutive mRNA levels in trypanosomes, Elife 7 (2018).

467 [27] C.G. Grunfelder, M. Engstler, F. Weise, H. Schwarz, Y.D. Stierhof, M. Boshart, P.
468 Overath, Accumulation of a GPI-anchored protein at the cell surface requires sorting at
469 multiple intracellular levels, Traffic 3(8) (2002) 547-59.
470 [28] M. Fiebig, S. Kelly, E. Gluenz, Comparative Life Cycle Transcriptomics Revises
471 Leishmania mexicana Genome Annotation and Links a Chromosome Duplication with
472 Parasitism of Vertebrates, PLoS Pathog 11(10) (2015) e1005186.
473 [29] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina
474 sequence data, Bioinformatics 30(15) (2014) 2114-20.
475 [30] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat
476 Methods 9(4) (2012) 357-9.
477 [31] A. Roberts, L. Pachter, Streaming fragment assignment for real-time analysis of
478 sequencing experiments, Nat Methods 10(1) (2013) 71-3.
479 [32] J.L. Munoz-Jordan, K.P. Davies, G.A. Cross, Stable expression of mosaic coats of
480 variant surface glycoproteins in Trypanosoma brucei, Science 272(5269) (1996) 1795-7.
481 [33] M. Carrington, N. Miller, M. Blum, I. Roditi, D. Wiley, M. Turner, Variant specific
482 glycoprotein of Trypanosoma brucei consists of two domains each having an
483 independently conserved pattern of cysteine residues, J Mol Biol 221(3) (1991) 823-35.
484 [34] J.S. Silverman, J.D. Bangs, Form and function in the trypanosomal secretory
485 pathway, Curr Opin Microbiol 15(4) (2012) 463-8.
486 [35] P.T. Manna, C. Boehm, K.F. Leung, S.K. Natesan, M.C. Field, Life and times:
487 synthesis, trafficking, and evolution of VSG, Trends Parasitol 30(5) (2014) 251-8.
488 [36] C. Tiengwe, A.E. Brown, J.D. Bangs, Unfolded Protein Response Pathways in
489 Bloodstream-Form Trypanosoma brucei?, Eukaryot Cell 14(11) (2015) 1094-101.

490 [37] V.P. Triggs, J.D. Bangs, Glycosylphosphatidylinositol-dependent protein trafficking
491 in bloodstream stage Trypanosoma brucei, Eukaryot Cell 2(1) (2003) 76-83.
492 [38] U. Bohme, G.A. Cross, Mutational analysis of the variant surface glycoprotein GPI-
493 anchor signal sequence in Trypanosoma brucei, J Cell Sci 115(Pt 4) (2002) 805-16.
494
495
496

497 **Figure legends**

498 **Figure 1. Integration of the *VSG2* transgenes into the *VSG6* bloodstream
499 expression site (BES) and consequent effect on growth of transgene containing
500 cell lines**

501 **A.** Site of integration of transgenes. Electroporation of Acc65I and SacI digested
502 targeting construct (p3952; Supp. Figure 1) containing either the wild-type or a mutant
503 *VSG2* ORF into bloodstream form trypanosomes expressing VSG6 and selection with
504 blasticidin resulted in stable cell lines expressing either wild-type or mutant VSG2 in
505 addition to wild-type VSG6. Sequence of the BES3 containing VSG6 is from Genbank
506 FM162569; intact ORFs are indicated and the vertical ticks are spaced at 10 kbp.

507 **B.** Map of VSG2 showing the location of PTCs, at codons 64, 167, 262, 392, 426 and
508 460. The signal peptide, the N-terminal domain, linkers (1 and 2), the C-terminal
509 domain, and the GPI-anchor signal sequence, and the two N-linked oligosaccharides
510 (NLO and residue number) are shown. The approximate molecular weights (kDa) of
511 wild-type or each of the PTC-containing VSG monomers are shown.

512

513

514

515 **Figure 2. VSG2 transgenes were expressed and PTCs were effective in
516 terminating translation**

517 **A.** Measurement of cell number over a five day time course for three independent
518 clones of each transgenic cell line. Proliferation is expressed as cumulative cell number

519 /10000 and the standard error is shown (Supp. Table 2). Two panels are used for
520 clarity.
521 **B.** Western blot of whole cell lysates for the series of cell lines expressing a *VSG2*
522 transgene in a VSG6 background. One clone of each cell line is shown: VSG6, the
523 parental cell line expressing VSG6. VSG2, a distinct cell line expressing VSG2 from its
524 endogenous locus. The remaining cell lines express a *VSG2* transgene in a VSG6
525 background: VSG6 VSG2 WT, or WT VSG6 with PTCs at one of the following codons
526 within VSG2: 64, 167, 262, 392, 426 or 460 were loaded on the gel. Whole cell lysates
527 of 2 x 10^6 cell equivalents were loaded per lane and the white asterix (*) indicates
528 truncated VSG2 polypeptides of the expected predicted molecular weights.
529
530 **Figure 3. Expression of PTC-containing *VSG2* transgenes and measurement of**
531 **transgene and endogenous *VSG* mRNA levels**
532 **A.** Northern blot analysis of *VSG2* and *VSG6* mRNA levels in one set of cell lines
533 expressing a *VSG2* transgene in a VSG6 background. To assist with quantitation,
534 mRNA from unmanipulated VSG6 and VSG2 expressing cell lines was also analysed as
535 a titration of the amount of total RNA loaded. A relative loading of 1 indicated 1 µg total
536 RNA. The blot was probed for rRNA to adjust quantitation for precise loading.
537
538 **B.** *VSG* mRNA levels determined from quantitative northern blots as above, expressed
539 relative to the *VSG* mRNA level measured in unmodified cell lines VSG2 (blue) and
540 VSG6 (red). First, an average value for cell lines expressing a wild type *VSG2*
541 transgene in a VSG6 background (VSG6+VSG2 transgene) is shown followed by a plot

542 of *VSG* mRNA expression against the location of the stop codon in the *VSG2*
543 transgene. The error bars are standard error of the mean from measurements of three
544 independent clones for each transgenic cell line. The measurements of *VSG6* mRNA in
545 all cell lines with a PTC-containing *VSG2* transgene were significantly different ($p<0.05$
546 by Student's unpaired t-test) to *VSG6* mRNA in cell lines with a wild type *VSG2*
547 transgene (Supp. Table 3).

Table 1. Measurement of VSG mRNA abundance by RNAseq. Two moderately abundant mRNAs (DHH1 and NOT1) an an abundant mRNA (RPL4) are shown for comparison

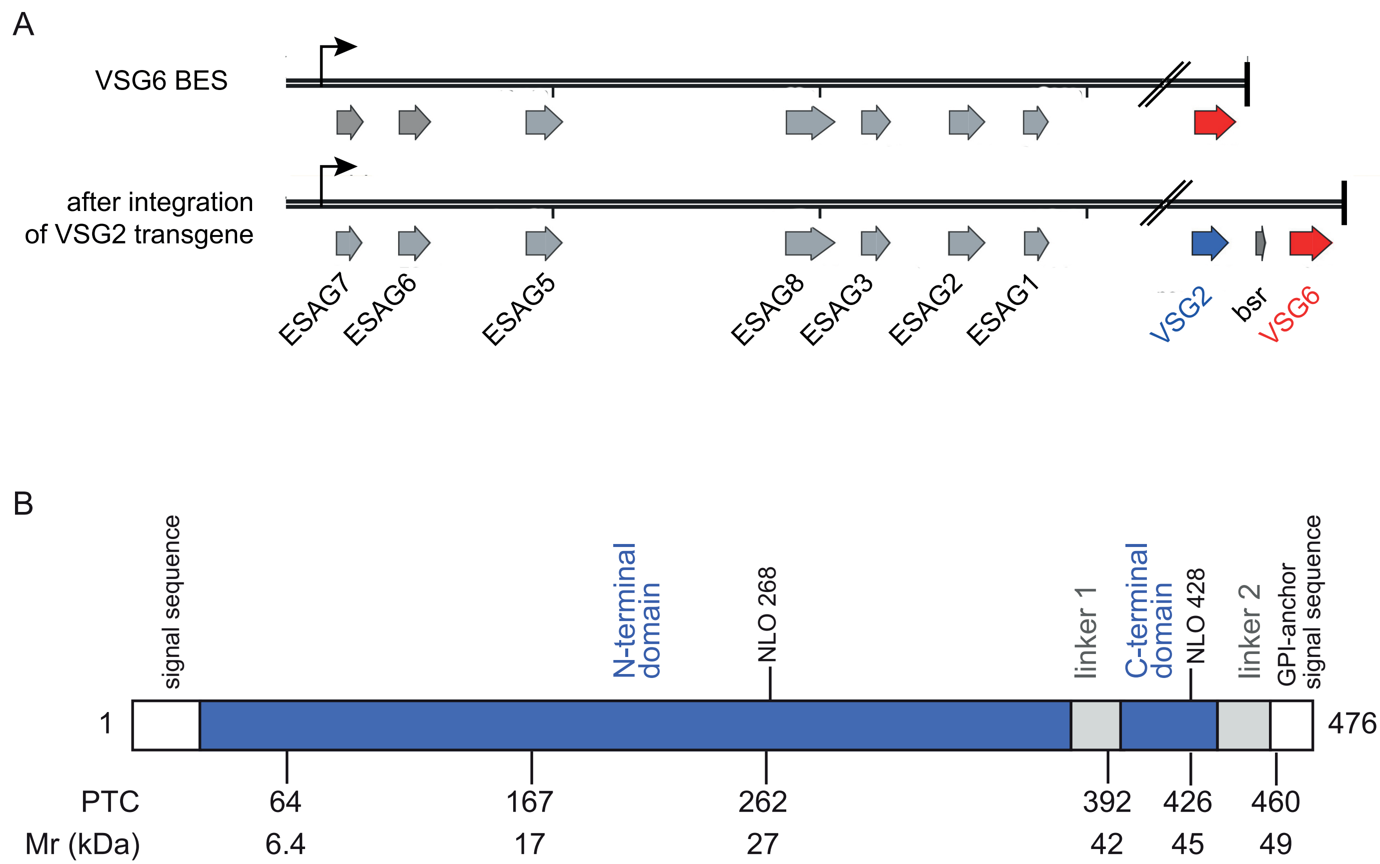
mRNA expression in transcripts per million transcripts (TPM)

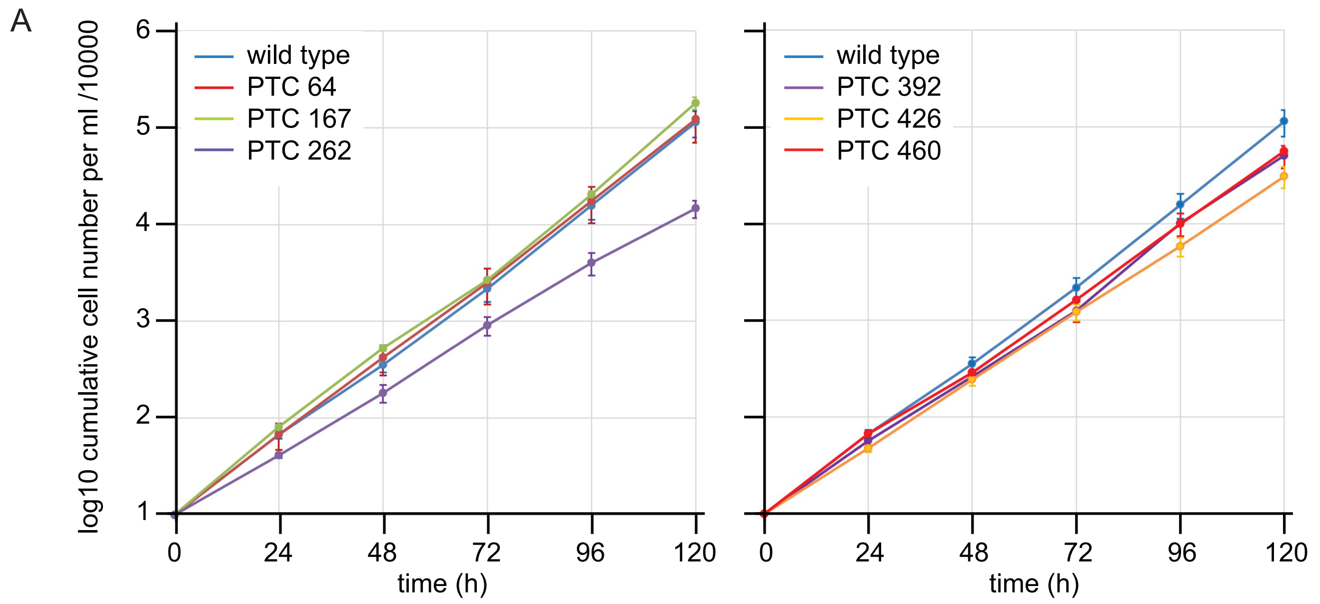
Gene	Accession number	*T. brucei* L427 VSG2	*T. brucei* L427 VSG6 parental	*T. brucei* L427 VSG6:VSG2 d28	*T. brucei* L427 *VSG6:VSG2* d44
VSG6	Tb427.BES15.12	25	219777	136573	147716
VSG2	Tb427.BES40.22	151633	5	47619	51168
total VSG		151658	219782	184192	198884
DHH1	Tb427.10.3990	294	259	292	306
NOT1	Tb427.10.1510	197	161	135	155
RPL4	Tb427.03.5050	2214	1891	2202	2060

Table 2. The relative abundance of VSG mRNA in a transgenic cell line expressing wild type VSG2 compared to wild type cell lines

Fractional VSG mRNA expression

Gene	Accession number	*T. brucei* Lister 427 VSG6 parental	*T. brucei* Lister 427 VSG2	*T. brucei* Lister 427 VSG6 p3952 VSG6:VSG2 double expresser d28	*T. brucei* Lister 427 VSG6 p3952 VSG6:VSG2 double expresser d44
VSG6	Tb427.BES15.12	1	0	0.74	0.74
VSG2	Tb427.BES40.22	0	1	0.26	0.26


A
VSG6 BES
after integration
of VSG2 transgene
ESAG7
ESAG6
ESAG5
ESAG8
ESAG3
ESAG2
ESAG1
VSG2
bsr
VSG6
B
signal sequence
N-terminal domain
NLO 268
linker 1
C-terminal domain
NLO 428
linker 2
GPI-anchor signal sequence
1
476
PTC
64
167
262
392
426
460
Mr (kDa)
6.4
17
27
42
45
49



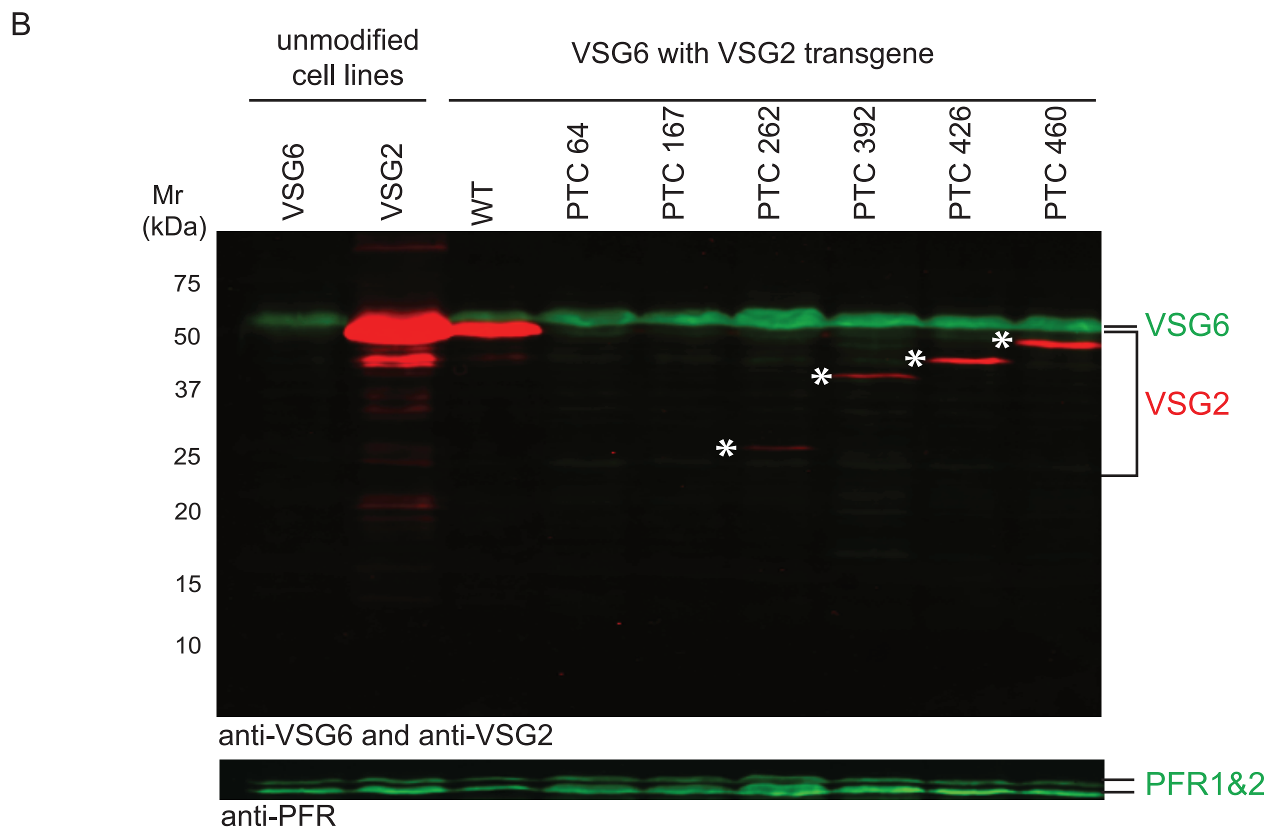
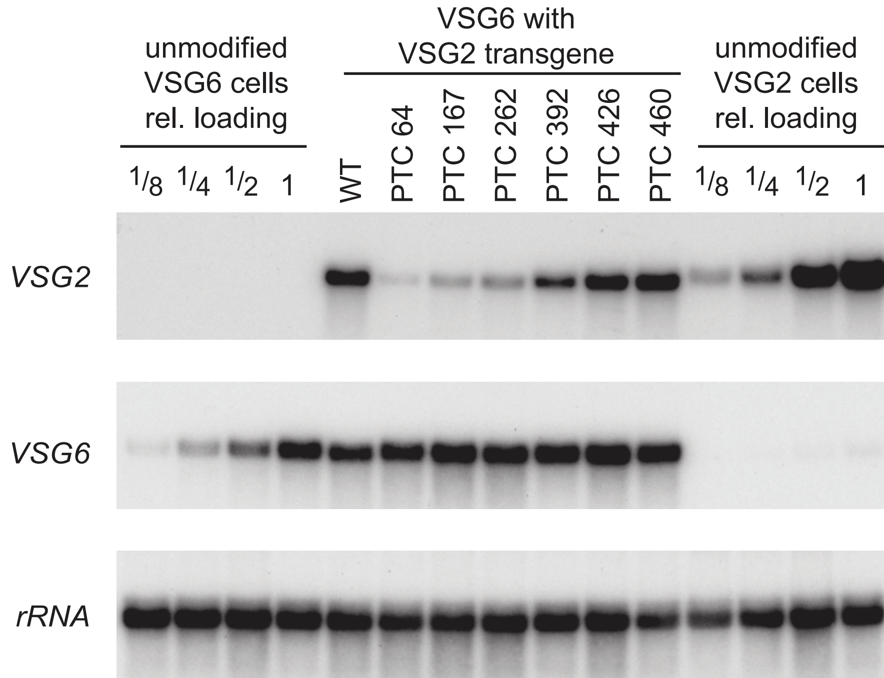


Figure 2

A



B

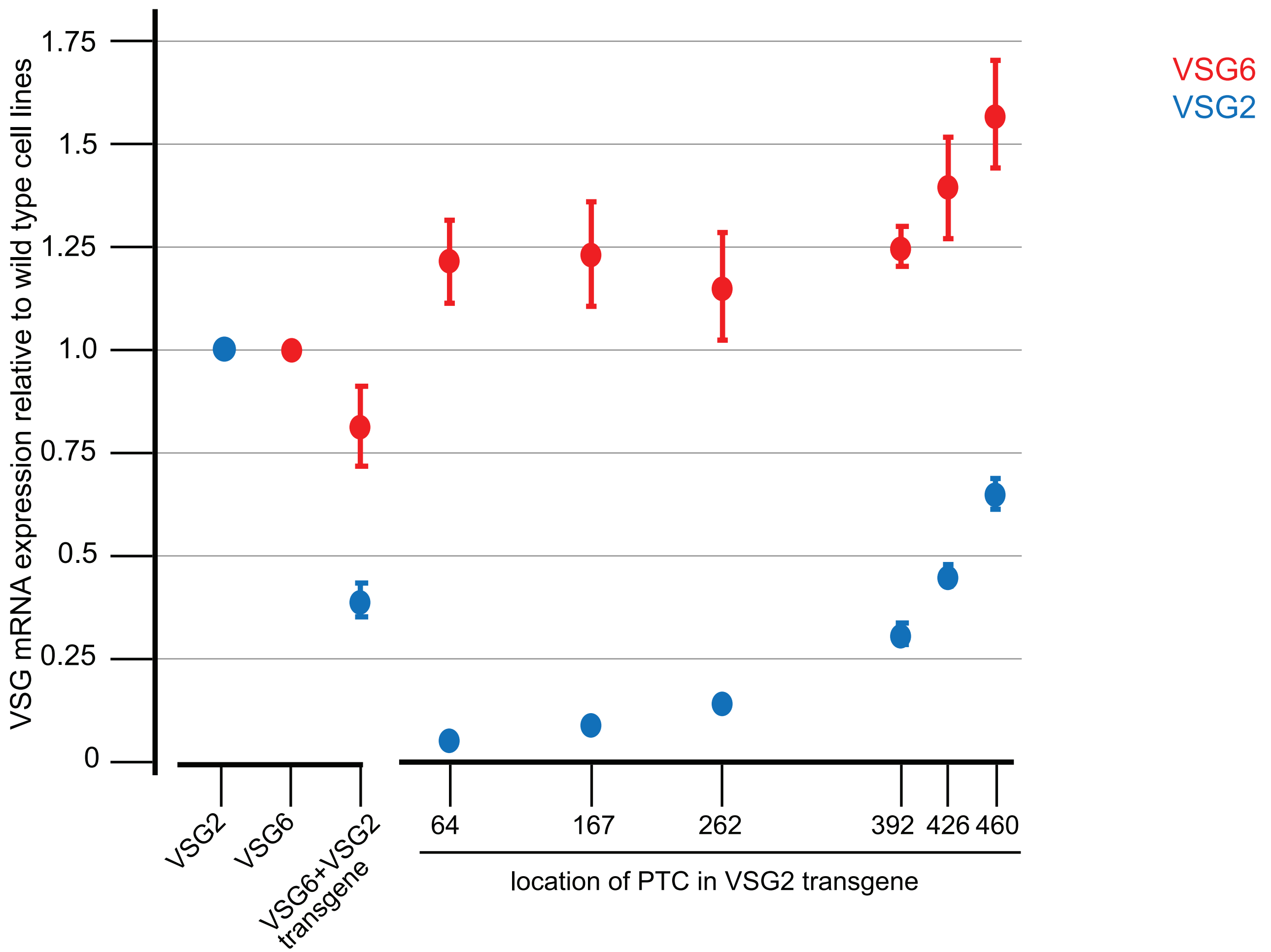


Figure 3