Nucleic Acids Research, 2003, Vol. 31, No. 14 ***3993–4000***
DOI: 10.1093/nar/gkg445

The frequency of gene targeting in *Trypanosoma brucei* is independent of target site copy number

Bill Wickstead, Klaus Ersfeld and Keith Gull*

School of Biological Sciences, University of Manchester, 2.205 Stopford Building, Oxford Road, Manchester M13 9PT, UK

Received April 2, 2003; Revised and Accepted May 7, 2003

ABSTRACT

We have investigated the effect of target copy number on the efficiency of stable transformation of the protozoan parasite *Trypanosoma brucei*. Using a single strain of the organism, we targeted integrative vectors to several different genomic sequences, occurring at copy numbers ranging from 1 to ~30 000 per diploid genome, and undertook a systematic assessment of both transformation and integration efficiencies. Even over this vast copy number range, frequency of gene targeting was the same for all sites. An independence of targeting frequency and target copy number is characteristic of mammalian homologous recombination and is unlike the situation in budding yeast. It is also not seen in the related parasite *Leishmania*, a distinction that may be the consequence of the different usage of recombination within the mechanisms of pathogenicity in the two species.

INTRODUCTION

The ability to make specific transgenic manipulations to an organism's genome is a hugely powerful tool in studies of gene structure and function. Gene targeting, by means of integration of exogenous DNA into host DNA, enables the investigation of *cis*-effects at the site of integration that could never be probed by *trans* approaches such as the use of autonomously replicating vectors. It also provides a method of stable transformation which can be consistent in terms of copy number and extremely stable over time, even in the absence of selection. For these reasons, much work has been done to make gene targeting a standard, though not necessarily straightforward, biotechnological tool in many systems, as it has been in yeast for many years.

The protozoan parasite *Trypanosoma brucei* is the causative agent of the fatal disease sleeping sickness. *Trypanosoma brucei* and the related parasites *Trypanosoma cruzi* and *Leishmania* spp. are significant insect-borne pathogens of tropical and subtropical regions. Aside from their impact on human health, these species are also important model systems for the study of basic cell biology, a position substantially aided by the advent of reproducible methods for the introduction of exogenous DNA into their genomes (1–5). Gene targeting has proved invaluable in these organisms as a method of analysing gene function and investigating chromosome structure. As well as being used for gene disruption, in *T.brucei* gene targeting is the predominant approach for stable transformation when *cis*-effects are not desired. This is despite the existence of relatively stable non-integrative vectors (half-lives in the absence of selection ranging from ~1.5 to ~17 cell generations) (6,7). Such a situation exists because the autonomously replicating vectors described to date do not provide sufficient improvements in transformation efficiencies to offset the greater stability and consistency afforded by integrative approaches.

At present, the key to gene targeting in all organisms is the soliciting of the cell's own homologous recombination machinery to deliver exogenous DNA to a specified chromosomal locus. Generally for vertebrate cells, integration of DNA at the target site by homologous recombination is obscured by a much higher frequency of random (i.e. non-homologous) integration events (reviewed in 8), necessitating the use of approaches such as positive–negative selection, promoter and polyadenylation traps and marker–target gene fusions to isolate true targeted recombinants (see 9–11). In the Kinetoplastidae *Trypanosoma* and *Leishmania* (2,3,12), as for *Saccharomyces cerevisiae* (13), the reverse is true, with nearly all integration events proceeding by homologous recombination.

Given the importance of gene targeting to the study of parasite pathogenesis and to the wider study of biology more generally, knowledge of the parameters that influence the frequency of gene targeting is of considerable interest. Factors which affect gene targeting in mammalian and/or yeast systems include: the cell line used, the length of homology between donor and target, the extent of homology to target, the dose of vector and the target copy number. Here we have investigated in particular the relationship between the target copy number and the frequency of gene targeting. Using a single *T.brucei* strain, we determined the efficiency of integration at genomic sequences occurring at low, intermediate and very high copy numbers and found no difference in the frequency of gene targeting. These data are in contrast to

*To whom correspondence should be addressed at present address: Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK. Tel: +44 1865 285 455; Fax: +44 1865 285 691; Email: keith.gull@pathology.oxford.ac.uk
Present address:
Bill Wickstead, Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

Targeting sequence	Promoter		
	-	Ti-P_{EP1}	Ti-P_{T7}
β-tubulin	•		
rDNA spacer		•	•
177-T1		•	
rev 177-T1		•	
177-T2		•	•
3x177-T2		•	
VSG-G4		•	•
VSG-S8		•	

Figure 1. Anatomy of vectors used in this study for stable integrative transformation of *T.brucei*. Vectors essentially differ only in the targeting sequence and promoter used. Targeting sequences were: β-tubulin gene; rDNA spacer; the minichromosomal *VSG*s G4 and S8; four different versions of the 177 bp repeat, a single Type 1 repeat (177-T1), a Type 1 repeat in the reverse orientation (rev 177-T1), a single Type 2 repeat (177-T2) and three direct Type 2 repeats (3 × 177-T2). Constructs for the transformation of PCF cells contained either no promoter or a tetracycline-inducible procyclin EP1 promoter (Ti-P_{EP1}). Constructs for the transformation of BSF cells contained a tetracycline-inducible T7 polymerase promoter (Ti-P_{T7}). *GFP* and *Hygro*^{r} open-reading frames abut splice acceptor sequences (SAS) and 3′-processing sequences (3′-PS) derived from endogenous trypanosomal genes. Targeting sequences and promoters were combined as shown.

the situation in *S.cerevisiae* and *Leishmania*, where gene targeting frequency is dependent on target copy number (14,15), and suggest that, in common with mammalian cells, the search for homology is not rate limiting to homologous recombination in *T.brucei*.

MATERIALS AND METHODS

Targeting constructs

All plasmids for the stable transformation of procyclic form (PCF) cells were derived from a construct we have named pGad7. This root vector contains a polycistron encoding green fluorescent protein (GFP) and a hygromycin resistance marker under the control of a tetracycline-inducible procyclin EP1 gene promoter (16). Processing sequences are: an EP1 splice acceptor sequence with aldolase 3′-processing sequence for the GFP gene; an actin splice acceptor sequence and 3′-processing sequence for the hygromycin resistance marker gene. Upstream of the promoter is a small multicloning site to accommodate targeting sequences. Downstream of the polycistron are two T7 terminator sequences and a recognition site for the meganuclease I-SceI. A map and full sequence of pGad7 are provided as Supplementary Material and a detailed lineage can be obtained from the authors (see also Fig. 1).

pGad8-VSG-G4 and pGad8-VSG-S8 contain sequences for targeted integration of the plasmids into the minichromosomal variable surface glycoprotein (VSG) genes *VSG* G4 and *VSG* S8, respectively (accession nos AF294807 and AF294806). Sequence for targeting *VSG* G4 (including an engineered BamHI recognition site) was produced in two halves by PCR from *T.brucei* strain 427 gDNA. Primers used were: CTGGAGCTCTTCTCGCATTAAAGCCAC and CGGGATCCCTTGCAACCTGTTTCATC; CGGGATCCCAACTGAAGTCAGGGCAA and GACCTCGAGACAGTTCGTCGATGCTTG. These amplicons were digested with the restriction nucleases SacI/BamHI and BamHI/XhoI, respectively, and used in a three-way ligation with SacI/XhoI-digested pGad7 to produce the plasmid pGad8-VSG-G4. Sequence for targeting *VSG* S8 in pGad8-VSG-S8 was produced from a single PCR using the primers CTGGAGCTCTCCAGCAAACGAGCGGAT and GACCTCGAGGCCTCCAGCTTGAGTTTG and gDNA template. A unique PvuII site in the *VSG* S8 gene obviated the need for extra restriction sites to be engineered into this sequence.

Targeting sequences in pGad8-177t1 and pGad8-rev177 are single 'Type 1' 177 bp repeats in the 'forward' and 'reverse' orientations, respectively (see Results). Both contain engineered BamHI recognition sites. Targeting sequence in pGad8-177t1 was produced in two halves by PCR from a *T.brucei* strain 427 gDNA template: ~90 bp sized DNA from a PCR using the primers CTGGAGCTCTAAATGGTTCTTATACGAATG and CGGGATCCTATTGCACACATTAAAAGTT; ~90 bp sized DNA from a PCR using the primers CGGGATCCTTAATTACAAGTGTGCAACA and GACCTCGAGTTAACACTAAAGAACAGCGTT. These amplicons were cloned as for pGad8-VSG-G4. The 'reverse' 177 bp repeat in pGad8-rev177 was produced from pGad8-177t1 by PCR with the primers GACCTCGAGTAAATGGTTCTTATACGAATG and CTGGAGCTCTTAACACTAAAGAACAGCGTT.

Targeting sequences in pGad8-177t2 and pGad8-3×177t2 are 'Type 2' 177 bp repeats, with engineered BamHI recognition sites. Targeting sequence in pGad8-177t2 was produced as for pGad8-177t1 but with the primers: CTGGAGCTCTTTAATGGTCCTTATACG and CGCGGATCCGTAATTAATATGGCACAC; CGCGGATCCGTGTGCAACATTAAATAC and GACCTCGAGACCCATTAAACACTAAAG. The longer targeting sequence in pGad8-3×177t2 was produced as for pGad8-177t2, but from ~270 bp sized DNAs.

The rDNA spacer targeting sequence (including an engineered NotI site) was produced by PCR from the plasmid pLew100 (17) using the primers CTGGAGCTCATATAGTTG and GACCTCGAGCATTTGTTCTTCTAC. Ligation of this amplicon into pGad7 gave the vector pGad8-rDNA. The pGad8-tubulin construct contains no promoter, relying on transcriptional read-through following integration for marker expression. It was obtained by replacing the promoter and EP1 splice acceptor sequence of pGad7 with a partial β-tubulin gene and aldolase splice acceptor sequence.

Constructs for stable transformation of bloodstream form (BSF) cells, pGad9-rDNA, pGad9-177t2 and pGad9-VSG-G4, were produced by replacement of the tetracycline-inducible EP1 promoters in pGad8-rDNA, pGad8-177t2 and pGad8-VSG-G4, respectively, with tetracycline-inducible T7 promoters (see Fig. 1). This was achieved by PCR amplification of the Ti-T7 promoter from pLew82 (17) with the primers CCGCTCGAGCCTGATTAATACGAC and CCCTTGCTCACCTAG, followed by XhoI/HindIII digestion and ligation into XhoI/HindIII-cut pGad8-rDNA, pGad8-177t2 and pGad8-VSG-G4.

Cell lines

PCF and BSF cell lines were derived from *T.brucei* strain 427 culture-adapted cells. PCF 'PTP' cells were a kind gift of P. Bastin (Laboratoire de Biophysique, Muséum National d'Histoire Naturelle, Paris, France). This cell line expresses the tetracycline repressor and a phleomycin resistance marker (*Bleo*^{r}) from the tubulin locus and was generated by the transformation of PCF strain 427 cells with linearised pHD449 vector (18), followed by selection with 2 μg ml^{-1} phleomycin. BSF cells were the 'single-marker cell line' described by Wirtz *et al.* (17) and express the tetracycline repressor and T7 RNA polymerase. PCF cells were cultured at 28°C in SDM-79 (19) supplemented with 10% (v/v) foetal calf serum. BSF cells were cultured at 37°C, 5% CO_2 in HMI-9 (20).

Trypanosome transformation

Transformations were carried out essentially as in Wirtz *et al.* (21). For PCF cells, 2.5×10^7 actively dividing cells were harvested by centrifugation (5 min at 500 *g*), washed once in ice-cold ZPFM buffer (22), centrifuged again and then resuspended in ZPFM at 5×10^7 cells ml^{-1}. An aliquot of 20 μg of linearised vector was added to these cells prior to electroporation twice at 1.7 kV, 110 μF with 3×100 μs pulses (Electro Square Porator; BTX). Since the constructs used for transfection are of approximately equal size (6.1–7.0 kb), this is a roughly equimolar plasmid dosage. Post-electroporation, cells were allowed to recover in normal growth medium for 16 h with induction (1 μg ml^{-1} tetracycline), after which time a cell count was performed and the proportion of cells positive for GFP was assessed by flow cytometry. Stable PCF transformants were selected for by aliquoting cells into 96-well plates at 2×10^5, 2×10^4 and 2×10^3 cells well^{-1} in the presence of 20 μg ml^{-1} hygromycin B.

BSF cell transformations were as for PCF cells excepting the following: ZPFM supplemented with 1% (w/v) glucose at 37°C was used in place of cold ZPFM; 50 μg of linearised vector was used; cells were electroporated once only; induction was with 1 μg ml^{-1} doxycycline. Much lower numbers of stable transformants are expected for BSF than for PCF, due to the lower transformation rate and low number of cells surviving electroporation. For this reason, BSF transformants were selected in a single 24-well plate containing 10 μg ml^{-1} hygromycin B.

Both PCF and BSF cells were passaged with selection for at least 16 days to ensure true positives and all wells tested contained GFP-expressing cells in which the hygromycin resistance gene was integrated into an endogenous chromosome ($n = 46$; data not shown).

Assessment of transformation efficiencies

An assessment of the number of positive transformants arising was made from the number of positive wells by assuming a Poisson distribution, $\mathbf{P}(\lambda)$, where λ is the mean number of positive transformants per well. Hence, the probability of any well being negative, p_0, is given by $p_0 = e^{-\lambda}$ and a measurement of λ can be made as follows,

$$\lambda = \ln[1 - (n_+/N)]$$

where n_+ is the number of positive wells and N is the total number of wells. This was considered a good measure of λ for any plate so long as it had no more than 80% of wells positive.

For each transformation in PCF cells, up to three estimates of λ could be gained (one from each plate) and up to three independent transformations were performed for each targeting sequence. In BSF cells, where the transformation efficiencies are much lower, only one estimate of λ per transformation was possible and four independent transformations were performed for each targeting sequence analysed. Integration efficiencies (PCF only) were assessed from the transformation efficiencies by dividing by the proportion of cells positive for GFP at 16 h post-electroporation.

Database mining

Sequence data were produced by the Pathogen Sequencing Group at the Sanger Institute (Wellcome Trust Genome Campus, Hinxton, UK) and can be obtained from ftp://ftp.sanger.ac.uk/pub/pathogens. From the large number of GSS reads containing 177 bp repeat sequence, we picked 20 reads at random. These contained a total of 42 complete (i.e. AluI–AluI recognition site) copies of the 177 bp repeat.

RESULTS

Transformation efficiencies for procyclic form cells

In *T.brucei*, since integration by homologous recombination predominates (2), the efficiency of stable integrative transformation can be taken as a direct measure of the frequency of gene targeting. This predominance of homologous recombination was confirmed by southern hybridisation for a sample of the positive transformants generated here; each of 46 clonal transformants examined showed evidence of a single integration event at the targeted loci (data not shown; Wickstead *et al.*, submitted for publication). We compared the efficiency of stable transformation of *T.brucei* cells when five sites present at very different copy numbers were targeted (see Table 1). Initially, we transformed a single strain laboratory culture of PCF cells. These cells are derived from the insect mid-gut life cycle stage of the parasite. The vector target sequences were: the β-tubulin gene (23); the region upstream of the promoter for the 18S rRNA gene (referred to here as the rDNA spacer) (24); the 177 bp repeat (25); *VSG* G4; *VSG* S8 (26). Figure 1 shows the anatomy of the plasmids used.

Two of the integration target sites, the β-tubulin gene and the rDNA spacer, are found on the diploid megabase sized chromosomes (MBCs) of *T.brucei* which carry all of the constitutively active portion of the trypanosomal genome. Both sequences are multicopy, but differ in their distributiona: α- and β-tubulin genes alternate in an array of up to 19 tandem

Table 1. The copy numbers in the *T.brucei* genome of the five sequences targeted in this study

Target sequence	Copy number[a]	Number of arrays[a]	Copies per array	References
β-Tubulin gene	34–38	2	17–19	(27,29)
rDNA spacer	15–20	6–7	~3	(24,30,31)
177 bp repeat	~30 000[b]	~100	100–500	(25,33)
VSG G4	4	4	1	(26)
VSG S8	1	1	1	(26)

[a]Per diploid genome.
[b]Our data from physical mapping of several MCs (Wickstead *et al.*, submitted for publication) suggests this lower copy number for the 177 bp repeats than estimates made elsewhere (33).

repeats at the diploid *TUB* locus (27–29), while the rDNA spacer can be found dispersed on several chromosomes at a total copy number of 15–20 (30,31).

The remaining three targeting sequences were derived from sequences on *T.brucei* minichromosomes (MCs). These small linear chromosomes (30–150 kb in size) appear to be aneuploid (26) and a heterogeneous population of roughly 100 MCs is maintained by the organism as a reservoir for *VSG* genes (reviewed in 32). Each of the minichromosomes consists predominantly of a repeat of ~177 bp (33), which is also found on the intermediate sized chromosomes (200–700 kb). The 177 bp repeat thus exists at a very high copy number (30 000–50 000 copies per diploid genome, depending on estimate). Conversely, the two targeted minichromosomal *VSG*s, *VSG* G4 and *VSG* S8, are low copy number sequences, found as single (i.e. non-tandemly repeated) subtelomeric genes on four or one MCs, respectively (26).

Figure 2A and B shows the transformation efficiencies for vectors targeted to different genomic sites in PCF cells. The efficiency of stable transformation is the product of the efficiency of transfection (the proportion of cells into which exogenous DNA is introduced) and the efficiency of integration (the proportion of cells containing exogenous DNA that subsequently integrate the DNA into the genome). We have noticed that the efficiency of transfection is exquisitely sensitive to experimental conditions and subject to a large experiment to experiment variation. In order to look at factors affecting the homologous recombination of DNA, rather than the introduction of DNA into the cell, we used a measure of the transfection efficiency, the proportion of cells positive for GFP at 16 h post-electroporation, to find the efficiency of integration in the transformation experiments. An assessment of transfection efficiency (and hence integration efficiency) could not be made for the tubulin-targeting construct since, unlike the other four vectors, it contains no promoter; transcription of the encoded marker genes relies on read-through by endogenous RNA polymerase which can only be achieved following integration into a host chromosome. Figure 2C and D shows the integration efficiencies for the transformations in Figure 2A and B.

The data clearly show that there is no relationship between target copy number and the efficiency either of transformation or of integration of exogenous DNAs into the *T.brucei* genome. Genomic targets present at from 1 to ⩾30 000 copies are targeted at the same frequency. Also, there is no significant barrier to the integration of constructs containing active promoters into the normally transcriptionally silent MCs; two low copy *VSG*s can be targeted with the same efficiency as

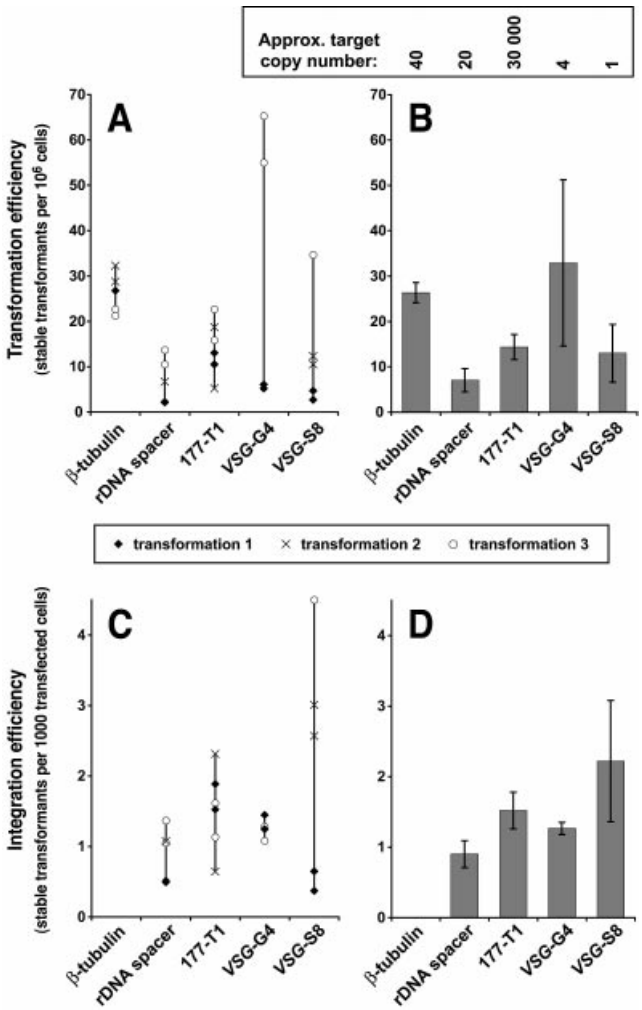

Figure 2. Efficiency of transformation (**A** and **B**) and efficiency of integration (**C** and **D**) in PCF cells when targeting genomic sequences occurring at different copy numbers. Cells were transformed with the plasmids shown in Figure 1. Targeting sequences were: β-tubulin gene; rDNA spacer; Type 1 177 bp repeat (177-T1); *VSG* G4; *VSG* S8. Data are from up to three assessments of efficiency from each of up to three independent transformations performed for each targeting sequence, as shown in (A) and (C). The same data sets are shown in (B) and (D) as average efficiencies with standard errors. Transformation efficiencies are expressed as number of stable transformants per 10^6 cells at 16 h post-electroporation. Integration efficiencies are transformation efficiencies divided by the proportion of cells positive for GFP at 16 h post-electroporation and are expressed as number of stable transformants per 1000 transfected (GFP-positive) cells.

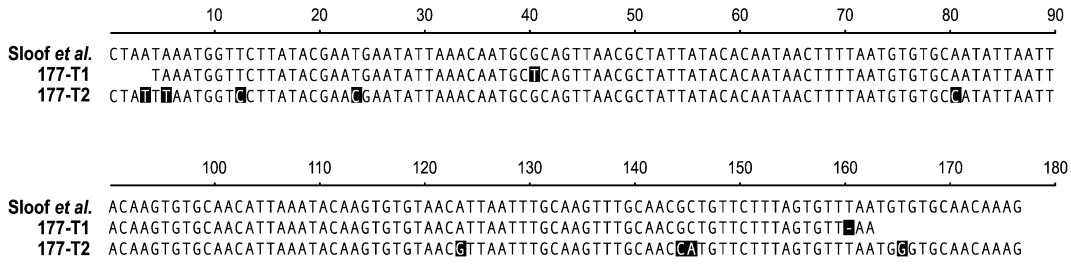

Figure 3. An alignment to show the heterology existing between 177 bp repeats. Sequences are: the original sequence described by Sloof *et al.* (25); the 'Type 1' 177 bp repeat (177-T1); the 'Type 2' 177 bp repeat (177-T2). The Type 2 repeat sequence is identical to the consensus 177 bp repeat sequences mined from the *T.brucei* genome sequencing project. Nucleotides differing from the sequence of Sloof *et al.* (25) are shaded.

multicopy loci on MBCs and the MC 177 bp repeats. Integration into MCs is not unprecedented, since Zomerdijk *et al.* (31) have shown that transcription is possible from an orphaned rDNA promoter on an unusual MC.

177 bp repeat heterogeneity

Efficient homologous recombination requires a high degree of sequence identity between substrates (34–36). This requirement is attributed to the activity of long patch mismatch repair systems which identify and dissect the heteroduplexed DNA of recombination intermediates formed between non-identical sequences (37,38). In *Leishmania*, gene targeting has been shown to be ablated by 13% divergence between target and vector (14). It is likely that a similar situation exists in *T.brucei*, since if several non-identical genomic sites (heteroalleles) exist, those with the highest homology to the input plasmid are targeted preferentially (39).

Given the high homology requirement for gene targeting, it is possible that the efficiency of transformation when targeting the 177 bp repeat might reflect heterology between repeats rather than true target copy number independence. For example, a requirement for absolute identity in a population of very heterogeneous repeats would seriously reduce the effective copy number. To investigate this possibility further, we mined the database at The Sanger Institute (http://www.sanger.ac.uk/Projects/T_brucei/) for information on the 177 bp repeats resulting from the ongoing *T.brucei* genome sequencing project. Forty-two randomly selected examples of the repeat showed a very low degree of heterology, with an average of 99% identity between repeats (data not shown). However, the consensus of the mined sequences was only 93% identical to the 177 bp repeat targeting sequence in the plasmid used in Figure 2 (see Fig. 3). This original 177 bp repeat targeting sequence will be referred to below as a 'Type 1' 177 bp repeat. It shares 99% identity with the sequence described by Sloof *et al.* (25) in the original description of the 177 bp repeat.

From the strain of *T.brucei* used in transformation experiments, we isolated a 177 bp repeat identical to the consensus of the mined sequences. This is referred to here as a 'Type 2' 177 bp repeat. From the degree of homology found between 177 bp repeat sequences on the database, it is expected that ~40% of all 177 bp repeats will be 100% identical to this sequence (equivalent to a copy number of at least 8000). When this Type 2 repeat was used as the targeting sequence, the frequency of gene targeting and efficiency of integration were the same as that for the 93% identical Type 1 sequence (Fig. 4). This demonstrates that repeat heterology is not the cause of the apparent target copy number independence of targeting the 177 bp repeat. Nor is it the case that the direction of the repeat influences integration. A Type 1 repeat in the reverse orientation targets exogenous DNA to MCs just as effectively (Fig. 4).

Lengths of targeting sequences

In mammalian systems, the frequency of gene targeting is exponentially dependent on the length of homology shared between the donor plasmid and target locus. This holds for lengths of homology up to ~10 kb, after which the recombination frequency levels off and then decreases slightly (34). In contrast, *S.cerevisiae* homologous recombination shows a linear relationship between substrate length and amount of recombination (40). *Leishmania* requires smaller target sequence lengths than mammalian systems. The highest gene targeting frequencies are observed when the length of target homology is $\geqslant$2 kb, with the critical factor seeming to be the length of homology on the shorter arm of the vector ($\geqslant$1 kb) rather than the total length of targeting sequence. Vectors with <200 bp homologous sequence at one or both ends yielded no positive transformants in *Leishmania* (14).

Table 2 lists the lengths of the targeting sequences in the vectors used to transform PCF *T.brucei* cells in this study. Targeting sequences consisting of only one copy of the 177 bp repeat (either Type 1 or Type 2) have smaller regions of homology at the vector ends (~90 bp) than those used to target other genomic sites. There was concern that this situation might result in aberrantly low rates of recombination for vectors targeting this 177 bp repeat. Were this the case, increasing the length of target homology would produce a concomitant increase in gene targeting for this site. We increased the 177 bp repeat targeting sequence length 3-fold (three direct Type 2 repeats), making the vector ends of comparable length to those of the rDNA spacer targeting plasmid. However, this resulted in no increase in the frequency of gene targeting (Fig. 4), indicating a true target copy number independence for gene targeting, rather than an artefact caused by target sequence length.

Furthermore, in this study we found no relationship between length of target homology, either smallest arm length or total length, and efficiency of transformation or integration for the sequences tested. Also, the lengths of homology necessary for efficient homologous recombination are much lower than for *Leishmania* or mammalian cell cultures. These data are consistent with the work of Shen *et al.* (41), who demonstrated

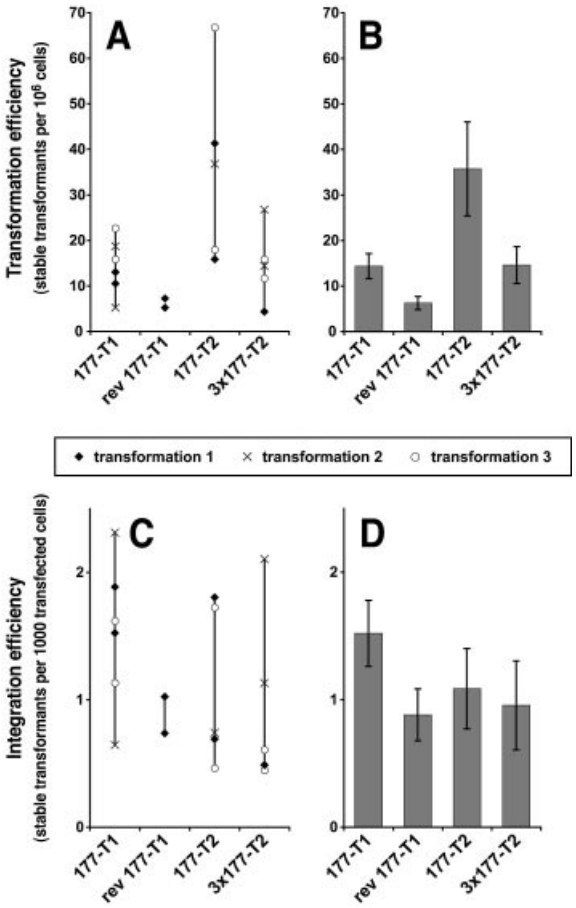

Figure 4. Efficiency of transformation (**A** and **B**) and efficiency of integration (**C** and **D**) in PCF cells when targeting different versions of the 177 bp repeat. Cells were transformed with the plasmids shown in Figure 1. Targeting sequences were: a single Type 1 repeat (177-T1); a Type 1 repeat in the reverse orientation (rev 177-T1); a single Type 2 repeat (177-T2); three direct Type 2 repeats (3 × 177-T2). Data shown are up to three assessments of efficiency from each of up to three independent transformations performed for each targeting sequence, as shown in (A) and (C). The same data sets are shown in (B) and (D) as average efficiencies with standard errors. Transformation efficiencies are expressed as number of stable transformants per 10^6 cells at 16 h post-electroporation. Integration efficiencies are transformation efficiencies divided by the proportion of cells positive for GFP at 16 h post-electroporation and are expressed as number of stable transformants per 1000 transfected (GFP-positive) cells.

that in *T.brucei* targeting sequences of 50, 75 or 90 bp were able to induce gene targeting with no significant differences in efficiency.

Transformation efficiencies for bloodstream form cells

The nuclear environment of the parasite in the bloodstream stage of its life cycle is different from that of the insect midgut stage. Prominent among the transcriptional changes are the well-documented ones associated with expression of the major surface proteins (reviewed in 42). Multiple sites of procyclin transcription are down-regulated in favour of a single subtelomeric VSG expression site (VSG-ES). Tight regulation ensures that only one of the potential VSG-ESs is expressed at one time, while reciprocal (cross-over) and non-reciprocal (gene conversion) recombination events are used as ways of switching the active *VSG*. There are also changes in chromatin

Table 2. Lengths of targeting sequences used in constructs for the stable transformation of *T.brucei*

Targeting sequence	Length of targeting sequence (kb)[a]	
	LHS	RHS
β-Tubulin	0.48	0.42
rDNA spacer	0.26	0.38
177 bp repeat[b]	0.09	0.09
3 × 177 bp repeat	0.27	0.27
VSG G4	0.36	0.53
VSG S8	0.83	0.28

[a]Lengths of targeting sequence on the left-hand side (LHS) and right-hand side (RHS) of the restriction site used for construct linearisation.
[b]Representative of both Type 1 and Type 2 repeats.

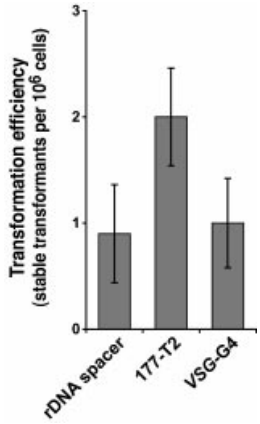

Figure 5. Efficiency of transformation of BSF cells when targeting genomic sequences occurring at different copy numbers. Cells were transformed with the plasmids shown in Figure 1. Targeting sequences were: rDNA spacer; Type 2 177 bp repeat (177-T2); *VSG* G4. Data shown are from four independent transformations performed for each targeting sequence. Transformation efficiencies are expressed as number of stable transformants per 10^6 cells at 16 h post-electroporation.

structure (43). For these reasons it cannot be automatically assumed that data relating to homologous recombination in PCF cells hold true in BSF cells. This statement is perhaps particularly applicable when considering the MCs, which are intimately linked with the process of VSG switching. However, when we performed BSF transformation experiments analogous to those performed for PCF cells, we found the same independence of transformation efficiency and target copy number. Three sites were targeted for BSF cells: the rDNA spacer, the 177 bp repeat and *VSG* G4 (see Table 1). All three sites showed the same frequency of gene targeting (Fig. 5), although the average frequency of targeting was much lower than for PCF cells, as is expected for BSF cells.

DISCUSSION

The effects of target copy number on the frequency of gene targeting in *T.brucei* were assessed. Similar assessments have been made in budding yeast (15), mammalian cells (44) and *Leishmania* (14) using isogenic cell lines in which particular genes had been amplified. Since in such cells the targeting sequence remains the same while the target copy number varies, this method removes artefacts that could result from

differences in targeting efficiency inherent to a particular donor or target. However, studies using this method must contend with the possibility that the act of gene amplification might alter the expression levels of proteins involved in the homologous recombination pathway (see 14). Here, we used the inverse approach: keeping the cell line constant and hoping to control for possible consequences of a changing target sequence.

Equal frequencies of gene targeting were observed in PCF cells for five genomic targets at copy numbers of one (*VSG* S8), four (*VSG* G4), ~20 (rDNA spacer), ~40 (β-tubulin gene) and ⩾30 000 (177 bp repeat). The same was true in BSF cells for a smaller number of targets (*VSG* S8, rDNA spacer and 177 bp repeat). As seen in *S.cerevisiae* (15), targeting did not seem to be strongly influenced by chromosomal context, since sites on MBCs were as efficient targets as both low and high copy sites on MCs. There were also no differences between chromosome-internal and subtelomeric sites. We do not believe that the targeting frequency of the high copy repetitive element was being significantly biased by heterogeneity occurring in the repeat population which might reduce the effective target copy number. The data supporting this conclusion were two-fold: (i) database mining revealed a high degree of homology between the repeats; (ii) changing the targeting sequence to a version of the repeat occurring at a different copy number did not affect the targeting frequency. Nor do we think that repeat targeting was aberrantly low because of short targeting flanks on the vectors, since increasing the lengths of the targeting regions had no affect on targeting efficiency. Indeed, we found no relationship between the length of plasmid targeting sequence and frequency of targeting for flanks ranging from ~80 to ~400 bp, adding to a similar observation made for targeting sequences of 50–90 bp in length (41). We interpret these data as demonstrating a true independence of frequency of gene targeting and target copy number for *T.brucei*.

In budding yeast, the frequency of gene targeting has a roughly linear dependence on target copy number, at least over a limited window (15). Interestingly, the same copy number dependence is seen when tandemly repeated genes on linear amplicons of the protozoan parasite *Leishmania* are targeted (although, interestingly, not for genes on circular amplicons) (14). For these organisms, the search for homology is rate limiting to gene targeting. The homologous recombination machinery must find a suitable locus for integration before the input DNA loses its potency to some interfering process, presumably degradation by nucleases. In contrast, and despite having ~100-fold more genomic DNA to explore, mammalian homologous recombination does not appear to be rate limited by the search for homology: similar frequencies of gene targeting are observed when targeting dihydrofolate reductase genes present at 2 or 800 copies (44). This conclusion is supported by an independence of homologous recombination with respect to input DNA dosage (45,46). Our data suggest that the protozoan *T.brucei* is not rate limited in gene targeting by the search for a homologous sequence. This is a facet *T.brucei* shares with mammalian cells and one that is not shared by related parasites of the *Leishmania* genus.

It is unlikely that the pathways for gene targeting in *Saccharomyces*, *Leishmania*, *Trypanosoma* and mammals are fundamentally different. A more plausible explanation is that differences exist in the levels and activities of proteins that promote or interfere with the process of homology searching during recombination. Factors that increase either the longevity of input DNA or efficacy of searching would allow the recombination machinery adequate time to explore a sufficient portion of the genome to prevent this step being rate limiting. *Trypanosoma brucei* is an exclusively extracellular parasite that evades the mammalian host immune response by periodically changing the identity of its cell surface (reviewed in 42,47). This antigenic variation is achieved by switching of the expressed VSG gene, predominantly through recombination events occurring within the VSG expression site (48). In contrast, the predominantly intracellular parasite *Leishmania* does not undergo antigenic variation and does not rely on homologous recombination for a known mechanism of pathogenicity. Given these different dependencies on homologous recombination, it is perhaps not surprising that, of the two closely related parasites, *T.brucei* should be the one to exhibit target copy number independence.

One other aspect of the data is worthy of note: that of minimal target length. The data presented here, along with those described by others (41,49), show that only short lengths of shared homology between input DNA and plasmid are necessary for efficient gene targeting in *T.brucei*, well below those required in *Leishmania* (14). Blundell *et al.* (39) have noted a paradox regarding trypanosomal gene conversion: there is a preference for a high degree of homology between donor and target DNA when transforming trypanosomes, yet important recombinational activity around the VSG genes is mediated by sequences that are rather divergent (reviewed in 47). At the 5′ border, homology between most *VSG*s is restricted to regions of very heterogeneous 70 bp repeats, while homology at the 3′ border consists of limited subtelomere identity. However, it is possible that an organism might fulfil both criteria if homologous recombination required only small blocks of very high homology, as appears to be the case for *T.brucei*.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Philippe Bastin (Laboratoire de Biophysique, Muséum National d'Histoire Naturelle, Paris, France) for donating the PTP cell line and Dr Miguel Navarro (Departamento de Biologia Molecular, Instituto de Parasitologia y Biomedicina, Granada, Spain) for his suggestions and helpful discussions on the experiments. This work was supported by a Wellcome Trust programme and equipment grant and a BBSRC postgraduate studentship to B.W.

REFERENCES

1. Lee,M.G.S. and Van der Ploeg,L.H.T. (1990) Homologous recombination and stable transfection in the parasitic protozoan *Trypanosoma brucei*. *Science*, **250**, 1583–1587.
2. Ten Asbroek,A.L.M.A., Ouellette,M. and Borst,P. (1990) Targeted insertion of the neomycin phosphotransferase gene into the tubulin gene cluster of *Trypanosoma brucei*. *Nature*, **348**, 174–175.

3. Cruz,A. and Beverley,S.M. (1990) Gene replacement in parasitic protozoa. *Nature*, **348**, 171–173.
4. Laban,A., Tobin,J.F., de Lafaille,M.A.C. and Wirth,D.F. (1990) Stable expression of the bacterial *neo*^{r} gene in *Leishmania enriettii*. *Nature*, **343**, 572–574.
5. Hariharan,S., Ajioka,J. and Swindle,J. (1993) Stable transformation of *Trypanosoma cruzi*: inactivation of the Pub12.5 polyubiquitin gene by targeted gene disruption. *Mol. Biochem. Parasitol.*, **57**, 15–30.
6. Metzenberg,S. and Agabian,N. (1994) Mitochondrial minicircle DNA supports plasmid replication and maintenance in nuclei of *Trypanosoma brucei*. *Proc. Natl Acad. Sci. USA*, **91**, 5962–5966.
7. Patnaik,P.K., Kulkarni,S.K. and Cross,G.A.M. (1993) Autonomously replicating single-copy episomes in *Trypanosoma brucei* show unusual stability. *EMBO J.*, **12**, 2529–2538.
8. Vasquez,K.M., Marburger,K., Intody,Z. and Wilson,J.H. (2001) Manipulating the mammalian genome by homologous recombination. *Proc. Natl Acad. Sci. USA*, **98**, 8403–8410.
9. Jasin,M. and Berg,P. (1988) Homologous integration in mammalian cells without target gene selection. *Genes Dev.*, **2**, 1353–1363.
10. Mansour,S.L., Thomas,K.R. and Capecchi,M.R. (1988) Disruption of the proto-oncogene *Int-2* in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature*, **336**, 348–352.
11. Sedivy,J.M. and Sharp,P.A. (1989) Positive genetic selection for gene disruption in mammalian cells by homologous recombination. *Proc. Natl Acad. Sci. USA*, **86**, 227–231.
12. Eid,J. and Sollnerwebb,B. (1991) Stable integrative transformation of *Trypanosoma brucei* that occurs exclusively by homologous recombination. *Proc. Natl Acad. Sci. USA*, **88**, 2118–2121.
13. Rothstein,R. (1991) Targeting, disruption, replacement and allele rescue—integrative DNA transformation in yeast. *Methods Enzymol.*, **194**, 281–301.
14. Papadopoulou,B. and Dumas,C. (1997) Parameters controlling the rate of gene targeting frequency in the protozoan parasite *Leishmania*. *Nucleic Acids Res.*, **25**, 4278–4286.
15. Wilson,J.H., Leung,W.Y., Bosco,G., Dieu,D. and Haber,J.E. (1994) The frequency of gene targeting in yeast depends on the number of target copies. *Proc. Natl Acad. Sci. USA*, **91**, 177–181.
16. Wirtz,E. and Clayton,C. (1995) Inducible gene-expression in trypanosomes mediated by a prokaryotic repressor. *Science*, **268**, 1179–1183.
17. Wirtz,E., Leal,S., Ochatt,C. and Cross,G.A.M. (1999) A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **99**, 89–101.
18. Biebinger,S., Wirtz,L.E., Lorenz,P. and Clayton,C. (1997) Vectors for inducible expression of toxic gene products in bloodstream and procyclic *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **85**, 99–112.
19. Brun,R. and Schönenberger,M. (1979) Cultivation and *in vitro* cloning of procyclic culture forms of *Trypanosoma brucei* in a semi-defined medium. *Acta Trop.*, **36**, 289–292.
20. Hirumi,H. and Hirumi,K. (1994) Axenic culture of African trypanosome bloodstream forms. *Parasitol. Today*, **10**, 80–84.
21. Wirtz,E., Hoek,M. and Cross,G.A.M. (1998) Regulated processive transcription of chromatin by T7 RNA polymerase in *Trypanosoma brucei*. *Nucleic Acids Res.*, **26**, 4626–4634.
22. Bellofatto,V. and Cross,G.A.M. (1989) Expression of a bacterial gene in a trypanosomatid protozoan. *Science*, **244**, 1167–1169.
23. Kimmel,B.E., Samson,S., Wu,J., Hirschberg,R. and Yarbrough,L.R. (1985) Tubulin genes of the African trypanosome *Trypanosoma brucei rhodesiense*: nucleotide sequence of a 3.7 kb fragment containing genes for alpha and beta tubulins. *Gene*, **35**, 237–248.
24. White,T.C., Rudenko,G. and Borst,P. (1986) Three small RNAs within the 10 kb trypanosome rRNA transcription unit are analogous to domain VII of other eukaryotic 28S rRNAs. *Nucleic Acids Res.*, **14**, 9471–9489.
25. Sloof,P., Bos,J.L., Konings,A.F.J.M., Menke,H.H., Borst,P., Gutteridge,W.E. and Leon,W. (1983) Characterization of satellite DNA in *Trypanosoma brucei* and *Trypanosoma cruzi*. *J. Mol. Biol.*, **167**, 1–21.
26. Alsford,S., Wickstead,B., Ersfeld,K. and Gull,K. (2001) Diversity and dynamics of the minichromosomal karyotype in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **113**, 79–88.
27. Seebeck,T., Whittaker,P.A., Imboden,M.A., Hardman,N. and Braun,R. (1983) Tubulin genes of *Trypanosoma brucei*: a tightly clustered family of alternating genes. *Proc. Natl Acad. Sci. USA*, **80**, 4634–4638.
28. Thomashow,L.S., Milhausen,M., Rutter,W.J. and Agabian,N. (1983) Tubulin genes are tandemly linked and clustered in the genome of *Trypanosoma brucei*. *Cell*, **32**, 35–43.
29. Ersfeld,K., Asbeck,K. and Gull,K. (1998) Direct visualisation of individual gene organisation in *Trypanosoma brucei* by high-resolution *in situ* hybridisation. *Chromosoma*, **107**, 237–240.
30. van Leeuwen,F., Kieft,R., Cross,M. and Borst,P. (2000) Tandemly repeated DNA is a target for the partial replacement of thymine by beta-D-glucosyl-hydroxymethyluracil in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **109**, 133–145.
31. Zomerdijk,J.C.B.M., Kieft,R. and Borst,P. (1992) A ribosomal RNA gene promoter at the telomere of a minichromosome in *Trypanosoma brucei*. *Nucleic Acids Res.*, **20**, 2725–2734.
32. El-Sayed,N.M., Hegde,P., Quackenbush,J., Melville,S.E. and Donelson,J.E. (2000) The African trypanosome genome. *Int. J. Parasitol.*, **30**, 329–345.
33. Weiden,M., Osheim,Y.N., Beyer,A.L. and Van der Ploeg,L.H.T. (1991) Chromosome structure—DNA nucleotide-sequence elements of a subset of the minichromosomes of the protozoan *Trypanosoma brucei*. *Mol. Cell. Biol.*, **11**, 3823–3834.
34. Deng,C.X. and Capecchi,M.R. (1992) Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Mol. Cell. Biol.*, **12**, 3365–3371.
35. Te Riele,H., Maandag,E.R. and Berns,A. (1992) Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. *Proc. Natl Acad. Sci. USA*, **89**, 5128–5132.
36. Shen,P. and Huang,H.V. (1986) Homologous recombination in *Escherichia coli*—dependence on substrate length and homology. *Genetics*, **112**, 441–457.
37. Dewind,N., Dekker,M., Berns,A., Radman,M. and Riele,H.T. (1995) Inactivation of the mouse *Msh2* gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination and predisposition to cancer. *Cell*, **82**, 321–330.
38. Modrich,P. (1994) Mismatch repair, genetic stability and cancer. *Science*, **266**, 1959–1960.
39. Blundell,P.A., Rudenko,G. and Borst,P. (1996) Targeting of exogenous DNA into *Trypanosoma brucei* requires a high degree of homology between donor and target DNA. *Mol. Biochem. Parasitol.*, **76**, 215–229.
40. Jinks-Robertson,S., Michelitch,M. and Ramcharan,S. (1993) Substrate length requirements for efficient mitotic recombination in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **13**, 3937–3950.
41. Shen,S.Y., Arhin,G.K., Ullu,E. and Tschudi,C. (2001) *In vivo* epitope tagging of *Trypanosoma brucei* genes using a one step PCR-based strategy. *Mol. Biochem. Parasitol.*, **113**, 171–173.
42. Pays,E. and Nolan,D.P. (1998) Expression and function of surface proteins in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **91**, 3–36.
43. Schlimme,W., Burri,M., Bender,K., Betschart,B. and Hecker,H. (1993) *Trypanosoma brucei brucei*: differences in the nuclear chromatin of bloodstream forms and procyclic culture forms. *Parasitology*, **107**, 237–247.
44. Zheng,H. and Wilson,J.H. (1990) Gene targeting in normal and amplified cell lines. *Nature*, **344**, 170–173.
45. Rommerskirch,W., Graeber,I., Grassmann,M. and Grassmann,A. (1988) Homologous recombination of SV40 DNA in COS7 cells occurs with high frequency in a gene dose independent fashion. *Nucleic Acids Res.*, **16**, 941–952.
46. Thomas,K.R., Folger,K.R. and Capecchi,M.R. (1986) High frequency targeting of genes to specific sites in the mammalian genome. *Cell*, **44**, 419–428.
47. Borst,P., Rudenko,G., Taylor,M.C., Blundell,P.A., van Leeuwen,F., Bitter,W., Cross,M. and McCulloch,R. (1996) Antigenic variation in trypanosomes. *Arch. Med. Res.*, **27**, 379–388.
48. Robinson,N.P., Burman,N., Melville,S.E. and Barry,J.D. (1999) Predominance of duplicative VSG gene conversion in antigenic variation in African trypanosomes. *Mol. Cell. Biol.*, **19**, 5839–5846.
49. Gaud,A., Carrington,M., Deshusses,J. and Schaller,D.R.G. (1997) Polymerase chain reaction-based gene disruption in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, **87**, 113–115.