3310–3317 Nucleic Acids Research, 1995, Vol. 23, No. 16

Structure of a frequently rearranged rRNA-encoding chromosome in *Giardia lamblia*

Guoying Hou, Sylvie M. Le Blancq[1], Yaping E[2], Huixiang Zhu[3] and Mary Gwo-Shu Lee*

Department of Pathology, New York University Medical Center, 550 First Ave, New York, NY 10016, USA, [1]Division of Environmental Health Sciences, School of Public Health, Columbia University, 630 West 168th Street, New York, NY 10032, USA, [2]Department of Dermatology and [3]Center for Neurology and Behavior, Columbia University, 630 West 168th Street, New York, NY 10032, USA

Received March 15, 1995; Revised and Accepted July 11, 1995

ABSTRACT

It has been shown previously that the rRNA encoding chromosomes in *Giardia lamblia* undergo frequent rearrangements with an estimated rate of ~1% per cell per division (Le Blancq *et al.*, 1992, *Nucleic Acids Res.*, 17, 4539–4545). Following these observations, we searched for highly recombinogenic regions in one of the frequently rearranged rRNA encoding chromosomes, that is chromosome 1, a small, 1.1 Mb chromosome. Chromosome 1 undergoes frequent rearrangements that result in size variation of 5–20%. We analyzed the structure of chromosome 1 in clonal lineages from the WB strain. The two ends of chromosome 1 comprise telomere repeat [TAGGG] arrays joined to a truncated rRNA gene and a sequence referred to as '4e', respectively. Comparison of the structure of four polymorphic versions of chromosome 1, resulting from independent rearrangement events in four cloned lines, located a single polymorphic region to the variable rDNA-telomere domain. Chromosome 1 is organized into two domains: a core region spanning ~850 kb that does not exhibit size heterogeneity among different chromosome 1 and a variable region that spans 185–450 kb and includes the telomeric rRNA genes, referred to as the variable rDNA-telomere domain. The core region contains a conserved region, spanning ~550 kb adjacent to the telomeric 4e sequence, which is only present in the 4e containing chromosomes and a 300 kb region of repetitive sequences that are also components of other chromosomes as well. Changes in the number of rDNA repeats accounted for some, but not all, of the size variation. Since there are four chromosomes that share the core region of chromosome 1, we suggest that the genome is tetraploid for this chromosome.

INTRODUCTION

The protozoan parasite *Giardia lamblia*, (also called *G.intestinalis* and *G.duodenalis*), causes a broad spectrum of acute to chronic gastrointestinal symptoms including diarrhea and malabsorption (1,2). Many infected individuals, however, are asymptomatic and the factors that determine the pathogenesis are unclear (3,4). *Giardia lamblia* trophozoites parasitize the small intestine where they proliferate and encyst to form infective cysts that are shed in the feces. Infection occurs when cysts are ingested and excystation occurs after passage through the stomach (5).

Giardia lamblia is one of the most primitive extant eukaryotes, according to a comparative analysis of small subunit rRNA gene sequences (6). Each *G.lamblia* trophozoite contains two transcriptionally active nuclei which are equal in size and contain the same amount of DNA (7). The accurate size of the genome is unknown and estimates of the genome size range from 1.2 to 8×10^7 base pairs (bp) per cell (8–11). Up to nine discrete size classes of chromosomes, ranging from 0.7 to ~4 megabasepairs (Mb), can be separated by pulsed-field gradient gel electrophoresis (PFGE) (12–14). But, the precise number of chromosomes is not known, with estimates ranging from 8 to 50 per trophozoite (7,12). The ploidy of the genome is unclear, as is the distribution of chromosomes between the two nuclei. However, recent data suggest that the genome is polyploid and that some components may be aneuploid (15,16).

Giardia lamblia exhibits considerable genomic variability (18). Different isolates from restricted geographic areas show extensive heterogeneity in the number and sizes of their chromosomes (14,17) and cloned lines maintained *in vitro* exhibit increased karyotype heterogeneity over time (14,19,20). A substantial amount of the variability in the clonal lineages that have been analyzed can be attributed to rearrangements of rRNA-encoding chromosomes (14,19–23). The rRNA-encoding chromosomes are hypervariable, with an estimated rearrangement rate of ~1% per cell per division cycle in the WB strain (20).

There are ~60 copies of a 5.6 kb rRNA gene (rDNA) repeat unit in the *G.lamblia* genome (9,24,25). These rRNA genes are arranged in short tandem arrays located on at least six telomeres (referred to as rDNA-telomere domains) (19,22,26). The rDNA-telomere domains, including a large area of as yet uncharacterized sequence(s) located upstream of the rDNA repeats, range in size from 25 to 500 kilobasepairs (kb) and are bounded by a series of restriction enzyme sites. The size of each rDNA-telomere domain can vary and this variation contributes to chromosomal size heterogeneity (19,23).

* To whom correspondence should be addressed

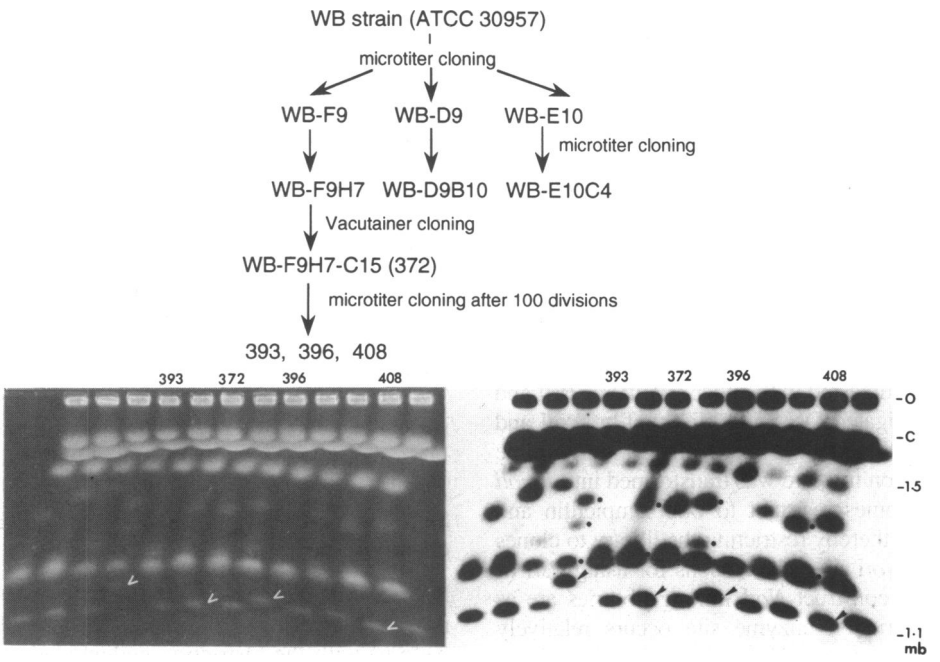

Figure 1. The lineage of cloned lines of the *G.lamblia* WB strain. Left panel: A PFGE gel stained with ethidium bromide comparing the chromosome-sized DNA of ten clones derived from WB-F9-H7-C15 [372] (lanes, from left to right, 1–5 and 8–12) and 372 itself (lanes 6, 7). The electrophoresis conditions were 650 s pulse time, 7.5 V/cm for 6 days. Right panel: Southern analysis of the PFGE gel showing the hybridization pattern with a rRNA gene probe (9). The numbers on the top of the gel indicate the clones chosen for the analysis of chromosome 1. Arrows point to the position of chromosome 1. Black dots indicate other rearranged rRNA-encoding chromosomes in these clones which were also visualized by chromosome 1 derived probes. O, origin; C, compression zone.

We have analyzed the structure of a ~1.1 Mb, rRNA-encoding chromosome, referred to as chromosome 1, in order to locate the polymorphic region(s) and thereby gain insight into the nature of the rearrangements that lead to karyotype heterogeneity in *G.lamblia*. Chromosome 1 undergoes frequent rearrangements that result in size variations of ~5–20%. We present here that (i) chromosome 1 is a truncated version of a larger chromosome that is present in several copies per cell. The truncation has the hallmarks of a chromosome breakage and healing event, mediated by telomere addition; (ii) comparison of the physical maps of several independently rearranged versions of chromosome 1 identifies a single variable region at the rDNA-telomere domain; and (iii) changes in the number of rDNA repeats account for some, but not all, of the chromosomal size differences.

MATERIALS AND METHODS

Cell lines and culture

Giardia lamblia strain WB (ATCC 30957) was obtained from the American Type Culture Collection. Trophozoites were grown in modified TYI-S-33 medium (27). Limiting dilution cloning was performed through the use of either microtiter plates or vacutainers (19,20). WB-F9H7, WB-D9B10 and WB-E10C4 are subclones in separate clonal lineages derived from the WB strain (Fig. 1) (19). WB-F9H7-C15 is a clone derived from WB-F9H7, referred to as 372, in which the chromosomal banding pattern remained unchanged (20). Clone 372 was propagated for 100 division cycles and this population was sampled by limiting dilution cloning. Three clones, 393, 396, 408 from this sample were chosen for this study because of the obvious size differences of their chromosome 1.

Southern genomic blot analysis

The cultured trophozoites were harvested as previously described (19). The genomic DNA was prepared as described (28). Following the restriction endonuclease digestions, the DNA was separated on a 0.8% agarose gel, transferred to nitrocellulose or nylon filters and hybridized with ^{32}P labeled probes. The final post-hybridizational wash was performed in 0.1 × SSC, 0.1% SDS at 65°C. For some of experiments, the hybridization signal was quantitated by a Betagen Betascope 603 Blot analyzer.

PFGE

Chromosome sized *Giardia* DNA was prepared as previously described (19,20,29,30). Conditions for restriction endonuclease digestion of agarose blocks containing chromosome-sized DNA was as described (19,30). The PFGE running conditions for each gel are described in the figure legends. DNA separated by PFGE was transferred to nylon or nitrocellulose filters and hybridized with ^{32}P-labeled probes. The final post-hybridizational wash was performed in 0.1 × SSC, 0.1% SDS at 65°C.

Purification of chromosome specific DNAs

Chromosome-sized *Giardia* DNA was obtained by PFGE through preparative 1% low melting point agarose gels. Agarose strips containing each chromosome band were sliced from the gel and stored in 0.5 M EDTA pH 9 and 1% Sarkosyl. Band-specific DNA in the agarose strips was either subject to direct restriction

endonuclease digestions, as described previously (19) or it was purified by either digestion with β-agarase or extraction with hot phenol.

Construction of chromosome 1 specific libraries

Chromosome 1 DNA was either (i) partially digested with DNase I, blunt-ended with T4 DNA polymerase I and ligated with *Hin*dIII linkers. The DNA was then cloned into pUC18; or (ii) digested with *Bam*HI and ligated into pBluescript SK+. The ligated plasmids were transformed into *Escherichia coli,* using standard protocols. Libraries were also constructed to select DNA fragments that contain *Not*I sites (i.e. *Not*I linking libraries), as follows. Chromosome 1 DNA was digested with either *Hin*dIII or *Pst*I and ligated into pUC18. The ligated DNA was digested by *Not*I and ligated with a 2 kb *Not*I fragment containing a tetracycline resistance gene. This ligation mixture was transformed into *E.coli* HB101 and bacterial colonies resistant to both ampicillin and tetracycline were selected, thereby restricting the library to clones containing inserts with a *Not*I site. The reasons for using *Pst*I or *Hin*dIII digested DNA to construct *Not*I linking libraries are as follows: (i) the *Pst*I restriction enzyme site occurs relatively frequently in the genome of *G.lamblia*; and (ii) the *Hin*dIII site does not occur in the rDNA-telomere domain. Therefore, *Not*I linking clones derived from rDNA repeats can be eliminated.

Cloning of the telomere of chromosome 1

Chromosome 1-derived DNA from WB-F9H7 was treated with BAL-31 nuclease for 60 s followed by phenol extraction, ethanol precipitation and digestion with *Bam*HI. The DNA was ligated into Bluescript SK$^+$ which had been digested with *Hin*cII and *Bam*HI. The ligation mixture was transformed into DH5α *E.coli* and bacterial colonies were screened with a telomere repeat probe [TAGGG]n (26). Theoretically, the library will be enriched for telomeric fragments, since the BAL-31 treated telomere supplies the blunt end and the cohesive end is generated by the *Bam*HI digestion.

Description of probes

The rRNA gene probe and the telomere repeat probe [TAGGG]n were as described by Boothroyd *et al.* (9) and Le Blancq *et al.* (26) respectively. The 4e probe was derived from the non-telomere repeat region of the clone CH1T4e. Probes B17, B27, B86, B97, B115, B143 and B158 are derived from a chromosome 1-*Bam*HI library. Probes P1, P8, P9, P10, P16, P21 and P27 are derived from a chromosome 1-partial DNase I library. Probes PT17, PT89, PT93, HT5 and HT32 are derived from *Not*I linking libraries.

RESULTS

We compared the structure of chromosome 1 in several clonal lineages derived from the WB strain of *G.lamblia*. These clones were obtained from previous experiments (19,20; see Materials and Methods). Figure 1 shows the lineages of the clones used for this study. The smallest 1.1 Mb rRNA-encoding chromosome that is referred to as chromosome 1 was first visualized by PFGE in the subclone WB-F9H7 (19). In the parental clone WB-F9, chromosome 1 was larger and migrated close to other chromosomes (see following sections). The WB-F9H7 subclone was re-cloned to give WB-F9H7-C15, referred to as 372 throughout this paper, in which the chromosomal banding pattern remained unchanged. Clones prepared from 372 showed considerable heterogeneity in the sizes of the rRNA-encoding chromosomes (Fig. 1, 20). Four or five size classes of rRNA-encoding chromosomes, ranging in size from 1 to 1.5 Mb, were examined. Larger chromosomes were not resolved under these PFGE conditions, and the strong hybridization of the rRNA probe at the compression zone represents rRNA-encoding chromosomes >1.5 Mb in size. The size of chromosome 1 varied by ~50–250 kb in three clones derived from 372, namely, 393, 396 and 408, to give an ~5–20% size variation (Fig. 1, indicated by arrowheads). The sizes of other rRNA-encoding chromosomes also varied in these clones (indicated by the black dots in Fig. 1), but subsequent analysis enabled us to distinguish chromosome 1 (see below). Therefore, we were able to perform a direct comparison of the structure of an rRNA-encoding chromosome before and after DNA rearrangements that resulted in the size polymorphisms.

Sequences adjacent to the telomere repeat array in chromosome 1

Digestion of chromosome 1 DNA from WB-F9H7 line (please note: WB-F9H7 and 372 have the same chromosomal banding pattern) with the restriction endonuclease *Bam*HI released two telomeric restriction fragments of ~0.7–0.8 and ~4.5 kb (data not shown). Clones of these two telomeric fragments were isolated from a chromosome 1, telomere-specific library. A 0.7 kb telomere clone, CH1Tr, comprised 162 bp of DNA sequence homologous to the published rRNA large subunit sequence (nucleotide position 2660–2821) followed by a 10 bp sequence (acagacagag), that included a 4 bp direct repeat (acag) and an array of ~110 copies of the telomere repeat [TAGGG]. The 10 bp oligomer is not homologous to any region of the rRNA gene repeat or to the cloning vector pBluescript SK$^+$ and was observed in several clones (data not shown). It is unlikely, therefore, that it is a cloning artifact. The distribution of this 10 bp sequence in the genome remains to be examined. The second clone, CH1T4e, of 4.8 kb comprised a 600 bp array of telomere repeats and 4.2 kb of previously unidentified sequence, referred to as 4e. This clone did not contain any obvious candidate for subtelomere repeats. The sequence CAGAG was found at the junction of the [TAGGG] repeat in both CH1Tr and CH1T4e clones (Table 1).

Chromosomal location of 4e sequences

To demonstrate whether the CH1T4e clone represents a telomeric fragment of chromosome 1, we analyzed the chromosomal location of the 4e sequence. Genomic DNA from clones derived from WB-F9 lineage (372, 393, 396 and 408) and clones from another independent clonal lineage, WB-D9, contained two size classes of *Bam*HI restriction fragments that hybridized with the 4e probe (Fig. 2A). These were a broad band of ~4.5–5 kb and bands of >9.6 kb. Surprisingly, only the >9.6 kb *Bam*HI fragments were detected in the other lineage examined, that is the WB-E10 series. Furthermore, the uncloned WB strain gave a very weak signal in the 4.5–5 kb region compared to that of the >9.6 kb bands. The uncloned WB strain is a heterogeneous population and the 4.5–5 kb 4e *Bam*HI fragment is only present in a subpopulation of cells. Southern blot analysis of chromosome specific DNA from clones 372, 393, 396 and 408 digested with *Bam*HI showed that the ~4.5–5 kb 4e was derived from chromosome 1 (data not shown) while the >9.6 kb 4e fragments were components of other chromosomes (see below).

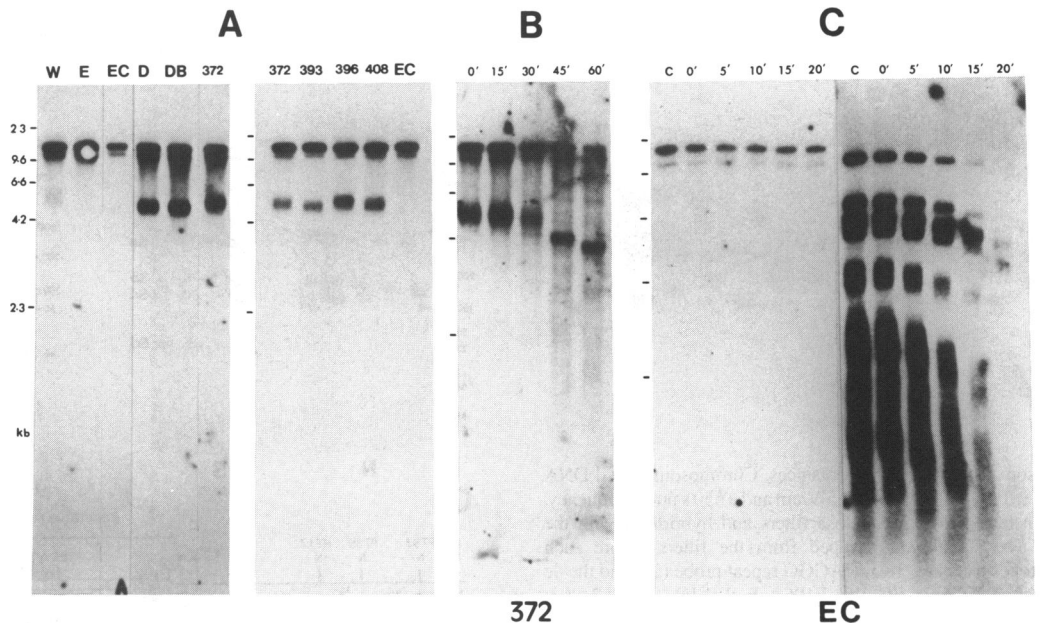

Figure 2. Southern genomic blot analysis of the 4e sequence. (**A**) Comparison of the banding pattern of 4e sequences among different WB clones. Genomic DNA derived from different clones, indicated at the top of each panel, was digested with *Bam*HI, separated in a 0.8% agarose gel and transferred onto nitrocellulose filters. These filters were hybridized with the 4e probe. W, the original WB strain; E, WB-E10; EC, WB-E10C4; D, WB-D9; DB, WB-D9B10. 372, 393, 396 and 408 are WB clones from the WB-F9H7 series, see Figure 1. (**B**) BAL-31 sensitivity of 4e sequences in clone 372. Genomic DNA of the 372 cloned line was treated with BAL-31 for different periods of time (indicated on the top of each lane) and then digested with *Bam*HI. Following the digestion, the DNA samples were separated in a 0.8% agarose gel, transferred onto nitrocellulose filters and hybridized with the 4e probe. (**C**) The BAL-31 sensitivity of 4e fragments in WB-E10C4. Genomic DNA of the WB-E10C4 cloned line was treated with BAL-31 for different periods of time (indicated on the top of each lane) and then digested with *Bam*HI. The left panel is the filter hybridized with the 4e probe. The right panel is the same filter hybridized with a telomere repeat [TAGGG]n probe.

The telomeric location of the 4.5–5 kb 4e restriction fragment was determined by the sensitivity of DNA to digestion with BAL-31 nuclease (Fig. 2B). Figure 2B shows that the length of the 4.5–5 kb restriction fragment gradually decreases with increasing BAL-31 digestion, while the lengths of the >9.6 kb fragments are unchanged after treatment with BAL-31. This result indicates that the 4.5–5 kb fragment is telomeric while the >9.6 kb fragments are located internally in the chromosome. Interestingly, all of the 4e *Bam*HI fragments (>9.6 kb) in the WB-E10C4 were resistant to treatment with BAL-31 (Fig. 2C, left panel). In contrast, telomeric fragments showed a rapid decrease in length after the incubation with BAL-31 (Fig. 2C, right panel). Thus, in WB-E10C4 clone, all of the 4e sequences are internally located.

Figure 3 shows the hybridization of a PFGE blot with the telomere repeat probe (left panel) and the 4e probe (right panel). The telomere repeat probe identifies all of the chromosomes. Interestingly, the 4e sequences are present in 3 or 4 sizes of chromosomes in clones of the WB-F9 lineage. These chromosomes are also detected by the rRNA probe (Fig. 1). Four bands can be distinguished in clones 393 and 396 (lanes 2 and 3); while the relative signal intensities in clones F9, 372 and 408 (lanes F, 1 and 4) suggest that the second band from the front may contain two co-migrating 4e chromosomes, to give a total of four chromosomes. We suggest, therefore, that there are four copies of the 4e chromosome in trophozoites of the WB-F9 lineage. Figure 3 also shows that chromosomes of a size equivalent to chromosome 1 is absent in WB-E10C4 clone. This explains that the telomeric 4e sequence can not be detected in WB-E10C4 clone.

These results show that 4e is present at a telomere in chromosome 1 and at an internal position in three other chromosomes. We hypothesize that chromosome 1 may have been created by a breakage in a chromosome-internal copy of the 4e sequence, followed by the addition of telomere repeats. The truncation might occur in a trophozoite in the original WB strain that subsequently formed a subpopulation of cells. Attempts to identify a novel chromosome 1 in the WB-E10 lineage after prolonged growth (200 generations; unpublished data) have failed. This result indicated that chromosome breakage at the 4e sequence most likely occurred at a lower frequency than the rearrangements of chromosome 1. Therefore, we assume that the polymorphic versions of chromosome 1 in the F9 lineage were generated by independent rearrangements of the original chromosome 1. In order to compare the structure of chromosome 1 and hence to identify the polymorphic regions, we prepared physical maps of the four polymorphic versions of chromosome 1.

Physical mapping of the conserved region of chromosome 1

Agarose blocks containing chromosome 1 DNA from *G.lamblia* clones 372, 393, 396 and 408 were digested by *Not*I and *Sfi*I, followed by fractionation in PFGE gels. Recombinant clones from chromosome 1-specific libraries were used as probes in the analysis of these PFGE gels. Four *Not*I and four *Sfi*I fragments were identified in the 550 kb region adjacent to the telomeric 4e sequence (Fig. 4A; please note that in panel A the telomeric *Not*I fragment is not shown and the band at the compression zone is uncut DNA). The sizes of these restriction fragments were identical in chromosome 1 DNA from the four *G.lamblia* clones. The probes used to identify each *Not*I and *Sfi*I fragments are indicated in the physical map (Fig. 4C). The banding pattern identified by each probe was first individually verified with a series of PFGE blots (data not shown). To provide a summarized

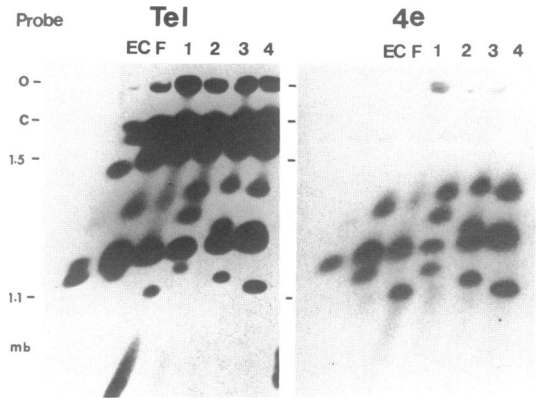

Figure 3. Chromosomal location of 4e sequences. Chromosome sized DNA was separated in a 1% PFGE for 6 days at 7.5 V/cm and a 650 s pulse frequency. The DNA was transferred to nitrocellulose filters and hybridized with the following probes. The signal was stripped from the filters before each re-hybridization. Probes used are: Tel, a TAGGG repeat probe (26) and the 4e probe. The lanes were as follows: EC, WB-E10C4; F, WB-F9, 1, 372; 2, 393; 3, 396; 4, 408; O, origin; C, compression zone.

result, the blots shown in Figure 4 were sequentially hybridized with indicated probes and the signals were superimposed. The relative positions of the *Not*I and *Sfi*I sites were assigned based on: (i) overlapping hybridization patterns of *Not*I and *Sfi*I digested chromosome 1 specific DNA and (ii) hybridization patterns of DNA partially digested with *Not*I or *Sfi*I. The alignment of the *Not*I sites was confirmed by hybridization with *Not*I-linking probes (i.e. PT93, PT89 and HT17).

Each probe from the 550 kb region adjacent to the 4e sequence of chromosome 1 identified three other chromosomes that also contain the 4e sequence (data not shown). Furthermore, these probes also detected the same sized *Not*I fragments, when all of the chromosomes from each clone were analyzed together (Fig. 4, compare panels A and B). These results showed that extensive homology existed in this 550 kb region among the four 4e-containing chromosomes within each clone and also among the clones. We refer to this 550 kb region as the conserved region.

Physical mapping of the repetitive region of chromosome 1

The conserved region comprises about half of chromosome 1. The other half, ranging from 450 to 750 kb in different versions of chromosome 1, is composed of repetitive sequence elements, including the rRNA genes (rDNA repeats). Most recombinant probes derived from this region detected multiple *Not*I and *Sfi*I restriction fragments in chromosome 1 DNA and the repetitive nature of this region precluded the alignment of the *Not*I and *Sfi*I sites. Therefore, the positions of these sites do not appear in the physical map, but the sizes of the various restriction fragments and probes obtained from this region are listed below the physical map in Figure 4C. We refer to this 450–750 kb region as the repetitive region. In contrast to sequences in the conserved region that are only present in the 4e-containing chromosomes, sequences from the repetitive region are present in many other chromosomes as well (data not shown). Approximately 300 kb of the repetitive region immediately adjacent to the conserved domain is the same size in the four versions of chromosome 1. This region is followed by a region that is polymorphic in size (i.e. the variable region)

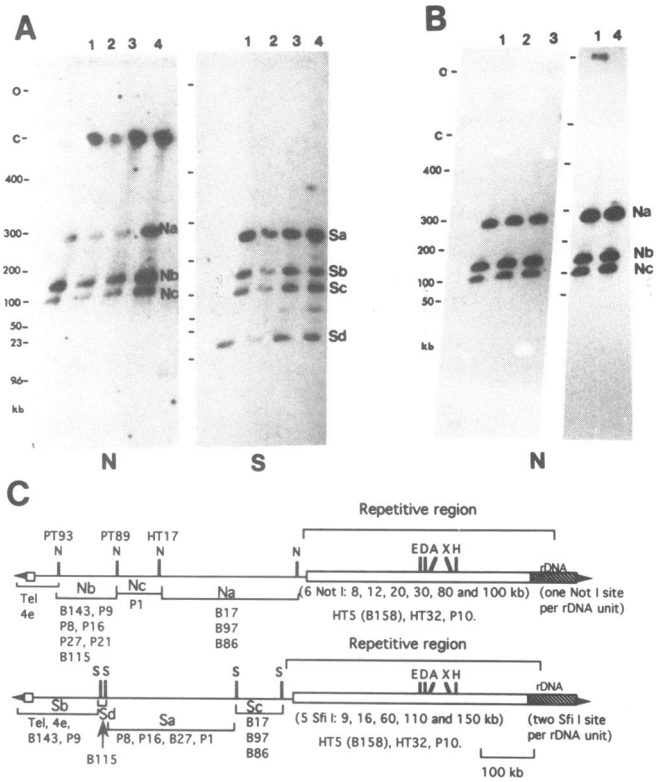

Figure 4. Construction of the physical map of chromosome 1. (**A**) PFGE analysis of restriction enzyme digested chromosome 1 specific blocks. N: *Not*I digestion; S: *Sfi*I digestion. (**B**) PFGE analysis of total chromosome blocks digested with *Not*I. The PFGE running conditions were 16 V/cm and a pulse frequency of 35 s for 30 h. Lanes 1, 2, 3 and 4 are *Giardia* clones 372, 393, 396 and 408, respectively. The panels with *Not*I digested DNA (N) in A and B were sequentially hybridized with probes B143, P1 and B97 and the signals were superimposed. The panel with *Sfi*I digested DNA (S) in A was sequentially hybridized with probes B143, B115, P1 and B97 and the signals were superimposed. O, origin; C; compression zone. The signal at the compression zone is uncut DNA. (**C**) Physical map. Top: locations of *Not*I sites. Bottom: locations of *Sfi*I sites. Na, Nb, Nc, Sa, Sb, Sc and Sd are *Not*I and *Sfi*I restriction fragments mapped in the conserved region as indicated in the physical maps. Probes which detected each fragment are listed underneath the corresponding fragment. Tel is a [TAGGG]n probe. PT93, PT89 and HT17 are *Not*I fragment linking probes which detect two adjacent *Not*I fragments at the positions indicated in the map. Two black arrows, telomere; open box, 4e; hatched box rDNA repeats. The positions of *Not*I and *Sfi*I sites were not assigned in the repetitive regions (the boxed region). P10, HT5 and HT 32 are derived from the repetitive region. There are at least six *Not*I and five *Sfi*I fragments which are detected by P10, HT5 and HT 32 probes and do not exhibit size variations and the sizes of these fragments are listed under the map. Each unit of the rDNA repeats (thick hatched box) contains one *Not*I and two *Sfi*I sites. N, *Not*I; S, *Sfi*I; E, *Eco*RI; X, *Xho*I; D, *Dra*I; A, *Acc*I; H, *Hin*dIII. The only *Eco*RI, *Xho*I, *Dra*I, *Acc*I, *Hin*dIII sites shown are those closest to the rDNA repeats (see Fig. 5).

and which accounts for the size differences in chromosome 1 (see the next section).

Physical mapping of the variable region of chromosome 1

We identified a single variable region in the four polymorphic versions of chromosome 1 (Fig. 5A). This region is the rDNA-telomere domain located at the other end of chromosome 1. The rDNA-telomere domains in WB strain of *G.lamblia* are delineated internally by a group of restriction sites that do not occur in the rDNA repeat (19). Figure 5A shows that the restriction fragments

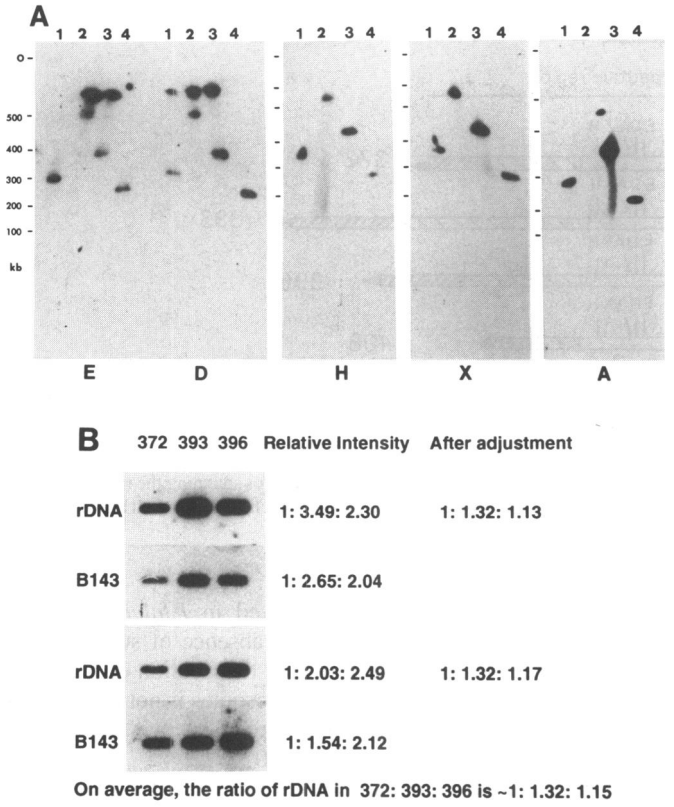

Figure 5. Identification of the variable region of chromosome 1. (**A**) PFGE analysis. Restriction enzyme digested chromosome 1 blocks derived from clones 372 (lane 1), 393 (lane 2), 396 (lane 3) and 408 (lane 4) were separated in 1% PFG gels for 30 h at 16 V/cm and a pulse frequency of 35 s. DNA was transferred onto nitrocellulose filters and then hybridized with ^{32}P-labeled rRNA gene probe. Each panel shows digestion with the restriction enzyme indicated below (see Fig. 4 for abbreviations). (**B**) Quantitation of the relative amount of rDNA repeats in clones 372, 393 and 396. Chromosome 1 specific DNA was digested with *Bam*HI (top two panels) or *Pst*I (bottom two panels), separated in 0.8 % agarose gel, transferred to nylon membranes and then hybridized with rDNA and B143 probes as indicated. The amount of signal in each panel was quantitated by a Betagen Betascope and the ratio of the relative intensity was calculated.

that contain the rDNA differ in size among the polymorphic versions of chromosome 1. For example, the *Hin*dIII fragments in chromosome 1 from 372, 393, 396 and 408 are ~250, 450, 330 and 185 kb, respectively. These polymorphic rDNA fragments were also identified by the telomere repeat probe, confirming the linkage of the rDNA and the telomere (data not shown). We therefore conclude that the polymorphic restriction fragments identified in chromosome 1 of clones 372, 393, 396 and 408 are the rDNA-telomere domains. Furthermore, the size differences of the rDNA-telomere domains account for the overall length differences (~5–20 % size variation) of different versions of chromosome 1.

To determine whether changes in the rDNA copy number contributed to the size polymorphism of the variable region, a comparison of the relative number of rDNA repeats in chromosome 1 was performed. Since, the rDNA repeats are present in many chromosomes which undergo frequent rearrangements, the comparison can only be conducted with chromosome specific DNA. Southern blots of chromosome 1 specific DNA from clones 372, 393 and 396 were hybridized with the rRNA probe and several other probes from the conserved region (408 was not available for this study). We assumed that the copy number of the sequences in the conserved region of chromosome 1 does not change in the different *G.lamblia* clones and hence that the relative amount measured reflected the relative amount of DNA loaded per sample. Figure 5B shows an example of the hybridization signal intensities of the rRNA probe and B143 probe which identified the Nb-*Not*I fragment indicated in Figure 4A. The relative amount of rRNA in 372: 393 :396 is 1: 3.49: 2.30; while that of B143 is 1: 2.65: 2.04 (Fig. 5B). The relative amount of rRNA repeats in 372: 393: 396 is ~1: 1.32: 1.13 after adjustment for the difference in amounts of DNA loaded. The ratio of 1: 1.32: 1.17 was obtained in a second experiment. On average, the ratio of rDNA in 372: 393: 396 is ~1: 1.32: 1.15. This calculation was performed with several conserved region probes and similar results were obtained (data not shown). The *Hin*dIII fragments that encompass the variable regions of chromosome 1 in clones 372, 393 and 396 are ~250, 450 and 330 kb, respectively. The exact copy number of rDNA repeats in chromosome 1 of each *Giardia* clone is unknown. If we assume that the 250 kb *Hin*dIII-rDNA fragment in chromosome 1 of the clone 372 contains the maximum possible number of 45 copies of rDNA repeat (5.6 kb per unit), then ~59 and 51 copies of rDNA repeat would be present in clones 393 and 396 respectively, since the ratio of chromosome 1 derived rDNA in 372: 393: 396 is 1: 1.32: 1.15. This would only account for ~39 and 42% of the size difference in the variable region of chromosome 1 in clones 393 and 396 respectively since only ~60 copies of rDNA repeat unit are present in the genome of *G.lamblia* and they are distributed on at least six chromosomes (19,22), the actual number of rDNA repeats in chromosome 1 must be less than that calculated above. Therefore, the changes of copy number of the rDNA repeat probably contributed a relatively small amount to the size variations observed in chromosome 1. The nature of the other DNA sequences in the variable region remains to be determined. However, the absence of many restriction enzyme sites indicates that the variable region may be composed of simple sequence(s).

Our results are summarized in Figure 6, where chromosome 1 is divided into several regions based on DNA complexity. Chromosome 1 is composed of a core region whose size does not change and a variable rDNA-telomere domain. The core region consists of a conserved region that identifies the four 4e containing chromosomes and a portion of the repetitive region whose sequences are also present in many other chromosomes.

DISCUSSION

Chromosomes encoding rRNA genes in *G.lamblia* are subject to frequent rearrangements that result in size changes ranging from 10 to 100s of kb (19,20,22,23). We have analyzed the structure and composition of four polymorphic versions of a small rRNA-encoding chromosome (chromosome 1) from a clonal lineage of the WB strain. This comparative analysis has shed light on several issues concerning genomic organization in *G.lamblia*. We show that the chromosome polymorphisms in a frequently rearranged rRNA-encoding chromosome are located in the subtelomeric, rDNA domain; that the genome is tetraploid for at least one chromosome; and that rRNA gene loci and other sequences in the repetitive region of chromosome 1 are distributed to non-homologous chromosomes.

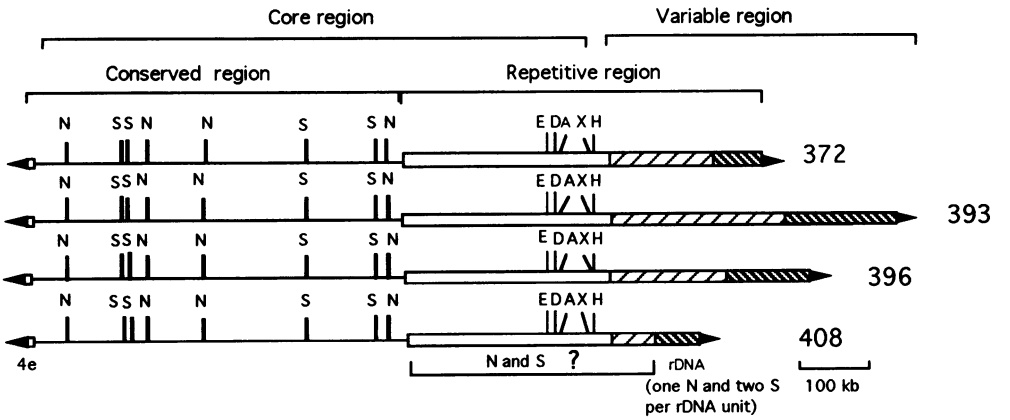

Figure 6. Physical maps of chromosome 1. Abbreviations for restriction enzyme sites and symbols are as described in Figure 4. The sizes spanning the rDNA repeats region (thick hatched boxes) were arbitrarily assigned and the exact copy numbers of rDNA repeat in each chromosome were unknown. The light hatched box represent as yet uncharacterized sequence in the variable region.

We have identified a single variable region, the rDNA-telomere domain, that accounts for the length polymorphisms of the different versions of chromosome 1. An increase in the copy number of the rDNA repeats accounted for a small amount of the size difference in two versions of chromosome 1, the remainder being contributed by uncharacterized sequences. Previous studies had shown that the rDNA-telomere domains contribute to chromosome polymorphism in *G.lamblia* (19,23), but they did not compare the same chromosome before and after a DNA rearrangement. In this study we have not only compared a chromosome before and after a DNA rearrangement, but we have also examined three independent rearrangements of the same chromosome 1. Our data identify the rDNA-telomere domain as the only variable region in three independent rearrangements. The possibility of small rearrangements elsewhere in the chromosome cannot be ruled out, but if they have occurred their contribution to the overall size polymorphism is negligible. The mechanism underlying the rearrangements in the rDNA-telomere domains is unknown. The overall amounts of loss or gain of rDNA per genome before and after rearrangements has not yet been determined, although large scale deletion or amplification has not been seen (19). It is possible that expansion or contraction of large tandemly arranged direct repeats, such as the rDNA repeats or putative simple sequence elements, may be involved. Interchromosomal transposition of a large subtelomeric segment via homologous recombination, has been observed in *P.falciparum* (36,37) and *T.brucei* (38). In *G.lamblia*, the presence of numerous homologous subtelomeric regions, such as the rDNA-telomere domains, in both homologous and non-homologous chromosomes, suggest that interchromosomal interactions may also play a role in chromosome length polymorphisms in this organism.

The compartmentalization of chromosomes into a conserved core region and variable subtelomeric domains is characteristic of chromosome organization in parasitic protozoa as diverse as *Plasmodium* sp., *Trypanosoma* sp. and *Leishmania* sp. (31). *Giardia lamblia* chromosomes appear to be organized along similar lines (16,23; this paper), although chromosome 1 only contains one variable subtelomeric region, the rDNA-telomere domain. Chromosome 1 is a truncated version of a larger chromosome and probably originated through a breakage and healing event mediated by telomere repeat addition. Breakage and healing events have been observed in *P.falciparum* and *T.brucei* and are characterized by the absence of subtelomeric sequences (32–34). The frequency and distribution of breakage and healing events in *G.lamblia* chromosomes is not known and their contribution to the overall rearrangement rate of ~1% per cell per division cycle may be small (20).

The rDNA-telomere domain of chromosome 1 provides the fourth independent example of cloned telomeres in *G.lamblia* in which the telomere repeat array is juxtaposed to a truncated rDNA repeat (22,26). The rDNA repeat sequences end at different points in each rDNA-telomere clone. A comparison of the sequences spanning the transition to the telomere repeats in all of the telomere clones identifies GGG or GAG residues at the junction with the first [TAGGG] (Table 1). It is possible that these residues facilitate an interaction with a putative *G.lamblia* telomerase that results in the addition of telomere repeats (35). A previous estimation showed that there are at least six rDNA-telomere loci per genome (26). Conventional subtelomeric repeats have not been found in *G.lamblia* thus far. However, only a limited number of telomeric regions have been analyzed, with all five of the cloned telomeres coming from the rRNA-encoding chromosomes. The structure of subtelomeric domains in other chromosomes remains to be examined.

Table 1. Comparison of the nucleotide sequence at the internal junction of the telomere repeat array in five telomere clones

		Reference
pGC1:	CGAGACGCCCCGGG[TAGGG]n	26
pG1T:'	CAGTACGCCCCGGAG[TAGGG]n	22
pG2T:	CCCCCGGACGGCaggg[TAGGG]n	22
Ch1Tr:	ATGCCTacagacagag[TAGGG]n	This paper
Ch1T4e:	CCTTTATCCCCCAGAG[TAGGG]n	This paper

Sequences derived from rDNA are underlined. [TAGGG]n is the telomere repeat.

Three or four discrete rRNA-encoding chromosomes with the same conserved region as chromosome 1 are present in some *G.lamblia* clones. These chromosomes are polymorphic and

differ in size by up to 40%. However, their structural similarity suggests that it is reasonable to consider them to be homologues and therefore that there are at least four copies per cell. Hence, we tentatively conclude that the genome is tetraploid for this chromosome. Interestingly, Yang and Adam (15) have data suggesting that there are 4–5 copies of a large chromosome that encodes a variant surface protein (VSP) gene. Homologous chromosomes in *T.brucei* can differ up 2-fold in size (30), while more modest size differences occur in *P.falciparum* with most polymorphisms restricted to the subtelomeric regions (36,39). A large domain of the internal region is conserved in the four chromosome 1 homologues examined in *G.lamblia*, this suggests that the polymorphisms of homologous chromosomes may also be concentrated in the subtelomeric regions. The karyotype variability seen in field isolates of *G.lamblia* is not restricted to rRNA-encoding chromosomes (14). The contribution of regions other than rDNA-telomere domains to chromosome size variation remains to be determined. The polyploid nature and the presence of repetitive sequences complicates the study of the *G.lamblia* genome and extra caution should be taken to establish the origin of DNA clones if YAC libraries are used for *Giardia* genome mapping.

Chromosomal hypervariability could affect the regulation of gene expression especially of genes that are proximal to or within the variable domains. DNA rearrangements are involved in the control of antigenic variation in *T.brucei* (38). To what extent chromosomal rearrangements in *G.lamblia* affect the regulation of gene expression is not known. However, *G.lamblia* undergoes antigenic variation of the VSPs, presumably mediated by differential regulation of VSP expression (15,40–43). The molecular mechanisms controlling *Giardia* antigenic variation remain unclear, and recombinational events affecting this process have not been identified. Future studies will focus on the biological consequences of genotypic heterogeneity in *G.lamblia* and its role in pathogenesis.

ACKNOWLEDGEMENTS

We thank Drs C. Berger, M. Briones, H. M. Chung, K. Gottesdiener and A. Sultanand for critical reading of this manuscript. This work was supported by a grant AI26497 from the National Institutes of Health. M.G-S.L. was supported by a grant AI28953 from National Institutes of Health. M.G-S.L. is a Burroughs Wellcome Fund New Investigator in Molecular Parasitology.

REFERENCES

1 Nash, T.E., Herrington, D.A., Losonsky, G.A. and Levine, M.M. (1987) *J. Infect. Dis.*, **156**, 974–984.
2 Heyworth, M.F. (1992) *J. Infect. Dis.*, **199**, 465–472.
3 Farthing, M.J.G. (1989) *Q. J. Med. New Ser.*, **70**, 191–204.
4 Ish-Horowicz, M., Korman, S.H., Shapiro, M., Har-Even, U., Tamir, I., Strauss, N. and Deckelbaum, R.J. (1989) *Pediatr. Infect. Dis. J.*, **8**, 773–779.
5 Meyer, E.A. (1985) *Parasitol. Today*, **1**, 101–105.
6 Sogin, M.L., Gunderson, J.H., Elwood, H.J., Alonso, R.A. and Peattie, D.A. (1989) *Science*, **243**, 75–77.
7 Kabnick, K.S. and Peattie, D.A. (1990) *J. Cell. Sci.*, **95**, 353–360.
8 Nash, T.E., McCutchan, T., Keister, D., Dame, J.B., Conrad, J.D. and Gillin, F.D. (1985) *J. Infect. Dis.*, **152**, 64–73.
9 Boothroyd, J.C., Wang, A., Campbell, D.A. and Wang. C.C. (1987) *Nucleic Acids Res.*, **15**, 4065–4084.
10 Fan, J.B., Korman, S.H., Cantor, C.R. and Smith, C.L. (1991) *Nucleic Acids Res.*, **19**, 1905–1908.
11 Adam, R.D. (1991) *Microbiol. Rev.*, **55**, 706–732.
12 Adam, R.D., Nash, T.E. and Wellems,T.E. (1988) *Nucleic Acids Res.*, **16**, 4555–4567.
13 Upcroft, J.A., Boreham, P.F.L. and Upcroft, P. (1989) *Int. J. Parasitol.*, **19**, 519–527.
14 Korman. S.H., Le Blancq, S.M., Deckelbaum, R.J. and Van der Ploeg, L.H.T. (1992) *J. Clin. Invest.*, **89**, 1725–1733.
15 Yang, Y. and Adam, R.D. (1994) *Nucleic Acids Res.*, **22**, 2102–2108.
16 Chen, N., Upcroft, J.A. and Upcroft, P. (1994) *Chromosome Res.*, **2**, 307–313.
17 Sarafis, K. and Isaac-Reutou, J. (1993) *Am. J. Trop. Med. Hyg.*, **48**, 134–144.
18 Carnaby, S., Katelaris, P.H., Naeem, A. and Farthing, M.J.G. (1994) *Infect. Immun.*, **62**, 1875–1880.
19 Le Blancq, S.M., Korman, S.H. and Van der Ploeg, L.H.T. (1991) *Nucleic Acids Res.*, **16**, 4405–4412.
20 Le Blancq, S.M., Korman, S.H. and Van der Ploeg, L.H.T. (1992) *Nucleic Acids Res.*, **17**, 4539–4545.
21 De Jonckheere, J.F., Majewska, A.C. and Kasprak, W. (1990) *Mol. Biochem. Parasitol.*, **39**, 23–28.
22 Adam, R.D., Nash, T.E. and Wellems, T.E. (1991) *Mol. Cell. Biol.*, **11**, 3326–3330.
23 Adam, R.D. (1992) *Nucleic Acids Res.*, **20**, 3057–3061.
24 Edlind, T.D. and Chakraborty, P.R. (1987) *Nucleic Acids Res.*, **15**, 7889–7901.
25 Healey, A., Mitchell, R., Upcroft, J.A. and Boreham, P.F. (1990) *Nucleic Acids Res.*, **18**, 4006.
26 Le Blancq, S.M., Kase, R.S. and Van der Ploeg, L.H.T. (1991) *Nucleic Acids Res.*, **19**, 5790.
27 Keister, D.B. (1983) *Trans. R. Soc. Trop. Med. Hyg.*, **77**, 487–488.
28 Van der Ploeg, L.H.T., Bernards, A., Rijswijk, F.A.M. (1982) *Nucleic Acids Res.*, **10**, 593–609.
29 Van der Ploeg, L.H.T., Smith, C.L., Polvere, R.I. and Gottesdiener, K. (1989) *Nucleic Acids Res.*, **17**, 3217–3227.
30 Gottesdiener, K., Garcia-Anoveros,J., Lee, M.G.S. and Van der Ploeg, L.H.T. (1990) *Mol. Cell. Biol.*, **10**, 6079–6083.
31 Lanzer, M., Fischer, K. and Le Blancq, S.M. (1995) *Mol. Biochem. Parasitol.*, **70**, 1–8.
32 Pologe, L.G. and Ravetch, J.V. (1988) *Cell*, **55**, 869–874.
33 Scherf, A., Carter, R., Paterson, C., Alano, P., Nelson, R., Ailcana, M., Mattei, O., Pereira da Silva, L. and Leech, J. (1992) *EMBO J.*, **11**, 2293–2301.
34 Zomerdijk, J.C.B.M., Kieft, R. and Borst, P. (1992) *Nucleic Acids Res.*, **20**, 2725–2734.
35 Blackburn, E.H. (1990) *J. Biol. Chem.*, **265**, 5919–5921.
36 Corcoran, L.M., Thompson, J.K., Walliker, D. and Kemp, D.J. (1988) *Cell*, **53**, 807–813.
37 Hinterberg, K., Mattei, D., Wellems, T.E. and Scherf, A. (1994) *EMBO J.*, **13**, 4174–4180.
38 Borst, P. and Greaves, D.R. (1987) *Science*, **235**, 658–667.
39 Kemp, D. J., Corcoran, L.M., Coppel, R.L., Stahl, H.D., Bianco, A.E., Brown, G.V. and Anders, R.F. (1985) *Nature*, **315**, 347–350.
40 Adam, R.D., Aggarwal, A., Lal, A.A., De La Cruz, V.F., MuCutchan, T. and Nash. T.E. (1988) *J. Exp. Med.*, **167**, 109–118.
41 Nash, T.E., Aggarwal, A., Adam, R.D., Conard, J. T. and Merritt, J.W. Jr (1988) *J. Immunol.*, **141**, 636–641.
42 Nash, T.E., Conard, J.T. and Merritt, J.W. (1990) *Mol. Biochem. Parasitol.*, **42**, 125–132.
43 Nash, T.E. (1992) *Parasitol. Today*, **8**, 299–234.