Trypanosoma brucei genomics and the challenge of identifying drug and vaccine targets

Eddy C. Agbo[1,*], Phelix A.O. Majiwa[2], Philippe Büscher[3], Eric Claassen[4] and Marinus F.W. te Pas[1]

[1]Division of Animal Sciences, Institute for Animal Science and Health (ID-Lelystad), Edelhertweg 15, 8200 AB Lelystad, The Netherlands
[2]International Livestock Research Institute, PO Box 20709, Nairobi, Kenya
[3]Department of Parasitology, Institute of Tropical Medicine, Nationalestraat 155, B-2000 Antwerp, Belgium
[4]Department of Immunology, Erasmus University Rotterdam, Dr Molewaterplein 50, 3015 GE Rotterdam, The Netherlands

There have been tremendous advances in our knowledge of trypanosome biology, yet many aspects remain unclear. Currently, the genome of *Trypanosome brucei* is being sequenced and this, with other genome-wide analysis methods, could provide novel insights into the parasite and facilitate the development of effective controls. An important new challenge investigators face is how to exploit the information in studying a parasite with so many genetic peculiarities. Here, we summarize our current understanding of molecular genetics of *T. brucei* and attempt to link genome analysis to the prospects for identifying possible targets for vaccines, novel drugs and specific diagnostics. The value of newly developed genotyping approaches in accelerating these processes is discussed.

The tsetse-fly-transmitted trypanosomes cause an acute to chronic debilitating disease called sleeping sickness in humans and nagana in cattle. Molecular tools and other classical approaches have been used to show the taxonomic complexity of the species *Trypanosome brucei* s.l. (*sensu lato*, i.e. in the broad sense). These have permitted its classification into human-infective subspecies – *T. brucei gambiense* and *T. brucei rhodesiense* – and *T. brucei brucei*, which is not infective to humans because the subspecies is lysed by normal human serum [1]. There are ~500 000 cases of human disease per year in sub-Saharan Africa [2], and the animal disease severely limits livestock production in one-third of the continent.

Control of trypanosomosis in Africa has relied heavily on chemotherapy and vector-control strategies. Although therapy and prophylaxis have been the most widely applied, their effectiveness remains unsatisfactory [3,4]. Initial anti-trypanosomal drugs comprised the salts of three compounds: diamine, phenanthridine and phenanthridine-aromatic amide; none of these can cross the blood-brain barrier in sufficient quantity to prevent relapse of late-stage cases of sleeping sickness. This situation was markedly improved following the introduction of melarsoprol, but this often has severe adverse effects [5,6]. No less remarkable, however, is the versatility shown by the trypanosome in its ability to develop refractoriness to treatment of the disease with these drugs. During the past 35 years, cymelarsan (a trivalent water-soluble analogue of the arsenical melarsoprol) and D,L-α-difluoromethylornithine (DFMO) (Eflornithine®) are the only new trypanocides commercially available for veterinary and human use. Unfortunately, these are also toxic to humans, and trypanosomes are showing increasing resistance to these drugs [7]. Moreover, DFMO, the newest trypanocide, is ineffective against *T. brucei rhodesiense* and thus can be used in humans only to treat disease caused by *T. brucei gambiense*. Even then, the drug is less efficient in children than in adults [8] and has never achieved widespread use because it is difficult to administer under field conditions. Short-term measures such as combination therapy or chemical modification of existing compounds could have significant impact on disease control in the immediate future and, to a limited extent, have been previously evaluated [9,10].

The need for additional drug targets and new, less-toxic therapeutics is widely acknowledged [2]. Vaccination could provide the most effective means for the control of trypanosomosis but no vaccine is available. In addition, molecular tools (such as genetic markers) are also required for fine-scale identification and more detailed characterization of all *T. brucei* subspecies, and for epidemiological investigations of trypanosomosis. Such specific diagnostic tools will be necessary adjunct to the effective use of any new drug or vaccine. Recent efforts in developing specific markers through fine-scale trypanosome genome analysis indicate that specific molecular diagnosis of *T. brucei* subspecies is feasible [11–14]. There are high expectations that genomics will offer new leads to rational vaccine design and accelerate the discovery of novel drug targets and the development of robust diagnostic assays. In this

Corresponding author: Eddy C. Agbo (Eddy.Agbo@wur.nl).
* Address from 1 August 2003 will be: Department of Biological Chemistry, Johns Hopkins School of Medicine, 725 N. Wolfe Street, Baltimore, MD 21205, USA (e-mail: eagbo@jhmi.edu).

 0966-842X/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved. doi:10.1016/S0966-842X(03)00151-3

article, we describe the current genomics-based technologies and discuss their potential contribution to the control of trypanosomosis caused by *T. brucei*.

Trypanosome genomics

Traditionally, the search for novel genes required for survival or virulence of any organism was based on several genetic methods involving genetic crosses, gene knockouts and induced mutations, followed by screening for the relevant phenotype. In trypanosomes, these approaches have been hampered by the inability to generate sufficient numbers of progeny clones from genetic crosses and the lack of a fine genetic map. These problems have affected the ability to determine precisely the statistical significance of segregating phenotypes in relation to the number of loci or alleles determining the trait of interest [15–17]. There is now a significant investment in sequencing the genome of *T. brucei* (Box 1). At the current rate of progress, it is estimated that the full genomic sequence of this parasite will be available during 2004. The challenge then will be to convert this sequence information to the functions of the full gene complement of this organism and to assess their potential value as drug and/or vaccine targets.

The principal goal of genome sequence analysis is the discovery of molecules that control such parasite factors as pathogenicity, virulence, host immune evasion and drug response. Although the single trypanosome clone whose genome is being sequenced will permit a definition of the gene organization, this sequence alone will not define factors that govern differences among different clones or strains of this parasite. Target characterization based on genetic analysis will therefore be necessary to define the degree of variation within a gene encoding a molecule identified as a potential drug target. For several other organisms for which genome sequence is available, the search for these molecules is currently being pursued through bioinformatics, which has dramatically increased the speed with which genes can be identified. With improved algorithms for structure predictions, physical genomics data can be converted into protein structure and function prediction. Further challenge remains in the bioinformatics-based screening of compounds from predicted protein structure and the computerized docking of small molecules onto these structures as a means for high-throughput screening of lead targets. Although the algorithms underlying these programs have improved significantly, the predictions are not perfect.

Two complementary approaches are being followed in sequencing the trypanosome genome: genomic DNA sequencing and expressed sequence tag (EST) sequencing. Initial data emerging from the *T. brucei* shotgun genome sequence indicate that the overall sequence organization of chromosomes is distinct from that of other higher eukaryotic organisms. For example, trypanosome genes are very tightly packed on chromosomes [18], which might explain in part the high incidence of alternative splicing of pre-mRNAs and the generation of polycistronic RNAs. The transcripts appear to be unidirectional over long stretches of the chromosomes. However, there are several similarities with gene organization in the other eukaryotes, particularly the protozoan parasites [19,20]. Several housekeeping genes are arranged in tandem repeats, making the sequencing difficult because of ambiguity in distinguishing large DNA clones containing repetitive DNA clusters that may have originated from different chromosomes. Furthermore, size differences have been observed between chromosome homologues within a single strain of *T. brucei*; these are due primarily to differences in the number of repeat units within the repetitive elements in the different chromosome homologues. The size variations are more pronounced for sequences at the telomeric ends. Given that only a few specific genes are known in *T. brucei*, many targets might well remain undiscovered among genes of unknown function, underscoring the importance of starting functional characterization of the genome of this parasite, using the available sequence data as the starting point.

One of the most intriguing observations from most genome sequencing projects completed so far is that a significant proportion (at least 40%) of genes identified have unknown functions [21–24]. In other words, these genes have been identified solely by sequence determination and the fact that they are homologous to genes identified in other organisms. Where they have no homologues or their open reading frames (ORFs) encode proteins of unknown function, such genes will need to be characterized by genetic or biochemical analysis.

EST analysis provides a rapid means of gene discovery and comparative data on expression levels. An example of the power of EST sequencing is provided by the efforts to sequence ~4500 randomly selected cDNAs from *T. brucei*, a process that successfully identified more than 1000 unique gene transcripts [25–27]. Prior to this, there were only a handful of *T. brucei* genes in public databases. Thus, EST sequencing rapidly identified many genes for further investigations. About 10–15% of these ESTs are similar to a known gene of at least one other organism, and ~50% are novel [18] (Box 1). In addition, EST analysis offers the possibility of identifying the complete ORF by reconstructing an overlapping consensus sequence from multiple ESTs. Data from accurately and fully sequenced ESTs can be used in identifying alternatively spliced transcripts and

Box 1. Websites of interest

a) http://www.tigr.org/tdb/tgi/tbgi/GO.html
Functional classification based on the gene ontology assignments
b) http://www.tigr.org/tigr-scripts/tgi/est_ann.pl?db = tbest
EST Annotator
c) http://www.tigr.org/tigr-scripts/tgi/tc_ann.pl?db = tbest
TC annotator
d) http://www.genedb.org/genedb/GOprocess?organism = trypandkeywords = Browse
Provides links to fully annotated trypanosome genes that: (i) are involved in various biological processes ($n = 187$); (ii) encode cellular components ($n = 74$); (iii) have known molecular function ($n = 290$); and (iv) have derived products ($n = 2727$)
e) http://www.ncbi.nlm.nih.gov/dbEST/index.html
f) http://www.tigr.org/tdb/e2k1/tba1/
g) http://www.sanger.ac.uk/Projects/T_brucei/

gene families. In addition, an extensive comparative analysis of *T. brucei* ESTs and those of several other protozoan parasites might reveal interesting evolutionary or functional trends among the taxa. However, EST sequence data, because of their inherent inaccuracy, should be used with caution.

Although EST sequencing provides a rapid means of gene discovery and a range of markers for gene-based sequence tagged site (STS) mapping, it is biased, because the ESTs originate from cDNAs synthesized from mRNA present in a population of trypanosomes at a particular life-cycle stage. For *T. brucei*, most ESTs were generated from the bloodstream forms, a few from procyclic culture forms and none from the infective metacyclic forms [25–27], owing mainly to the technical difficulties of generating sufficient materials from parasites at these developmental stages. It would have been desirable to generate the EST data from either a mixture of trypanosomes of all life-cycle stages or, ideally but more technically challenging, all life-cycle stages individually. High fidelity amplified RNA from small amounts of material might be useful to overcome this limitation to study gene expression from low abundance stages. However, considering the size of *T. brucei* genome, many transcripts are probably not represented among the ~1000 unique ESTs currently available in the public domain databases. Even after all the possible ESTs of *T. brucei* have been described, there will still be a significant proportion that encodes proteins of unknown identity or function. Among these are likely to be the genes that are specific to the parasite mainly because they are essential to its parasitic lifestyle. Presumably, if such genes can be silenced *in vivo*, they or their products might be the best candidates for drug and/or vaccine design precisely because they are likely to be highly directed to processes unique to the parasite. Furthermore, because these genes probably have no human counterpart, the cognate drug or vaccine might be highly specific for the parasites. However, they would have to be highly expressible and/or accessible to the immune system.

Variations at multiple genetic loci have been analyzed for novel allelic differences among *T. brucei* subspecies using amplified fragment length polymorphism (AFLP), minisatellite marker system and restriction fragment length polymorphism (RFLP) [14,28,29]. The AFLP technique has also been used to identify several polymorphic markers in parental lines of *T. brucei* and to construct a genetic map [17] (http://www.gla.ac.uk/ibls/II/cmrt/resint.htm#Trypanosoma brucei genetic map). The genetic analysis of significant trait(s) requires that linkage analysis be performed using molecular markers that segregate with the traits in the progeny of a cross. Such markers might be relevant for the identification of targets for the development of drugs and/or vaccines.

To date, there has been a paucity of published data on the analysis of the inheritance of specific traits in trypanosomes, except a preliminary analysis of resistance to lysis by human serum [15]. It is now known that human serum resistance of *T. brucei gambiense* and some *T. brucei rhodesiense* strains involve a complex of genes, possibly from multiple loci and alleles. This could confound classical genetic analysis and urges for global genome-wide investigations to identify the genes involved. Furthermore, the functional complexity of trypanosomes far exceeds that indicated by the genome data of the single clone currently being sequenced. This complexity is contributed by the genome plasticity. Therefore, forging a link among the genomic sequences of the three trypanosomatids – *Trypanosoma* (*T. brucei*, *T. cruzi*), *Leishmania* and *Crithidia* – would broaden the list of target candidates that can then be rapidly validated by other means. The gene categories generated by such a comparative approach could enable a preselection of target candidates on a whole-genome scale. The targets can then be defined according to the characteristics of specific phenotypic trait(s) for possible broad-spectrum applications. Knowledge of not just the target but also the upstream and downstream pathway links can allow a more selective choice of targets. A valuable strategy could be to build functional assays for all of the genes in key families of targets and to screen diverse compound libraries to identify chemical tools that can be used to validate the therapeutic utility of the targets. Currently, identified disease-related genetic variations are relatively rare and there are multiple variables of complex traits and diseases, including the relationships between hereditary, somatic and environmental factors. Thus, gene expression studies present a principal step in giving some insight into clusters of alleles that might be linkable to diseases and phenotypes. Ongoing work on single nucleotide polymorphisms (SNPs) is promising, and powerful new methods for integrating data and detecting variants undetectable by current technologies will still be required.

Comparative analysis of genomes of such closely related organisms can also be used to formulate testable hypothesis of genes that might be responsible for differences in tropism and pathologies among parasite strains or subspecies. An example is the kinetoplast DNA (kDNA), a network of topologically interlocked DNA rings common to the kinetoplastida but not found in mammals. Comparative genetic studies indicate that all kDNAs are essentially the same, with only minor variations [30]. These characteristics have been guiding efforts to understand these unusual structures, which might reveal useful diagnostic and pharmacological targets. However, almost all information on kDNA structure and gene function is derived from studies of *Crithidia* and *Leishmania*, which belong to the same superfamily. Parallel studies of this structure in *T. brucei* could be most useful. Genomic regions that are significantly different among *T. brucei* variants, for example, can be rapidly identified and evaluated for relevance in the development of drugs and/or vaccines.

Such comparative studies are already gaining appeal as templates in homology modelling and molecular dynamics study, to elucidate the functional importance and therapeutic potential of target enzymes in trypanosomes [11,12]. However, a recent attempt to identify kDNA replicative polymerase in trypanosomes by searching for candidate genes in *T. brucei* genome sequence data highlighted the limitation of comparative genomics for trypanosome genetic studies [31]. The study also raised important questions regarding the

evolution (in trypanosomes) of particular gene families that have been fully characterized in other organisms, and their utility in trypanosome genome studies.

Genomics and diagnostic assay development

Successful control of trypanosomosis will depend upon reliable diagnosis of the parasites, which might be difficult, especially in closely related subspecies. However, genomics is unravelling genetic polymorphisms underlying subspecies differences, which can be used for the required diagnosis (Table 1). The development of subspecies-specific markers for molecular diagnosis will be discussed first.

SNPs are densely spaced nucleotide differences found throughout the genome (http://snp.cshl.org/). Analyses of SNPs promise to enable disease association studies even without defined family or pedigree studies. The SNPs are expected to be the mainstay of efforts to map disease associations and to identify the genetic determinants of the varying responses organisms show to different drugs or compounds. Many methods used in the recent past used a subset of SNPs detected by the appearance or disappearance of restriction enzyme sites. Therefore, an important adjunct to studies on genetic diversity will involve defining the correlation between observed SNP and specific phenotypes.

Restriction-endonuclease-based genome analysis methods have been used to search for DNA markers that might be used to simultaneously differentiate *T. brucei* subspecies [11,13,14,29]. The diagnostic value of each detected polymorphism can subsequently be evaluated in larger population studies, to validate its association with specific parasite traits. The recent development of a finer-scale multi-endonuclease-based genotyping approach that simultaneously accesses multiple independent allelic sites within the genome offers a tool for expanding the number of identifiable single-nucleotide and restriction-fragment polymorphisms [32]. These tools increase the chance of detecting subtle variations in restriction patterns and help to establish the parameters of any future large-scale efforts to accumulate SNPs in *T. brucei*. Fragments of DNA containing SNPs identified as candidate markers can be cloned and converted into PCR-based sequence-tagged sites. However, the development of suitable bioinformatics tools will be required to provide a link to the underlying genes and the interpretation of any associations these might have with relevant phenotypes.

One approach to the development of specific assays for *T. brucei* is the identification of genetic markers linked to traits specific to the respective subspecies. These could correspond to genes that are specific to a subspecies, and could thus be used for identification on the basis of the trait encoded by the gene. An approach that exploited differential screening of cDNA libraries led to the identification of the serum-resistant-associated (SRA) gene transcript [33]. This gene appears to be present in most *T. brucei rhodesiense* strains but is undetectable in *T. brucei brucei* and *T. brucei gambiense* [34]. Previous studies identified the lytic factor as haptoglobin-related protein found only in primates [35] but this was challenged by latter reports [36,37]. Recently, the human-specific serum protein apolipoprotein L-I was shown to be the trypanolytic factor [38]. Although the exact mechanism by which apolipoprotein L-I lyses *T. brucei rhodesiense* is not yet fully understood, it opens a possibility for developing *T. brucei-rhodesiense*- specific drug and/or vaccine targets. A recently developed PCR assay based on the SRA gene appears to be specific for the identification of human-infective *T. brucei rhodesiense* and could serve as a useful tool for studies on molecular epidemiology of the disease caused by this subspecies [39,40].

The SRA gene represents a potential target for developing a therapeutic against *T. brucei rhodesiense* because this subspecies does not respond to treatment with Erflonithine. The development of such a therapeutic agent will be enhanced by the identification of a specific ligand in normal human serum with which the protein product of this gene can interact to cause the lytic effect. Based on such ligands, *in vitro* assays can be designed in which compound libraries are screened for small molecules that competitively displace the putative ligand.

Table 1. Genomics-derived diagnostic tools for *T. brucei*

Genomic region	Refs
Serum-resistance-associated gene	[33]
Trypanosoma brucei gambiense specific glycoprotein (TgsGP)	[68]
Repetitive satellite DNA	[69–71]
Internal transcribed spacer 1 of rDNA	[72]
Variable surface glycoproteins: expression-site-associated genes	[73]

Defining gene function

With the genome sequencing initiatives providing putative targets in profusion, some notion of function beyond what can be discerned from sequence homology is crucial for the decision on how best to continue with a putative target. Knowing when and where a gene is expressed can be an important input into this process. In addition to categorizing genes solely by sequence comparisons, selection of putative targets can be more precisely addressed through direct studies of gene expression, at either the mRNA or protein level. Trypanosomes rely heavily on post-transcriptional or post-translational gene regulation. Thus, the level of complexity resulting from co- or post-translational modification events can only be dissected and understood through qualitative and quantitative studies of gene expression at the level of the functional proteins. This requires proteomics investigations of the parasite genome, a subject that is outside the scope of this article.

RNAi

Owing to its hostile environment with dangerous opportunities for unwanted gene expression, trypanosomes have a vehement response to double-stranded RNA [41,42]. This poses immense challenge to potential drug delivery machineries. Recent studies indicate an intimate correlation between RNA-triggered silencing and gene activity [43,44]. Procedures based on silencing of gene expression promise to be valuable for functional genomics studies in trypanosomes, providing valuable information about gene function even in cases where only a partial loss-of-function is generated. RNA interference (RNAi), a

natural process first identified in the nematode worm, involves the use of specific RNA molecules to silence a single gene while leaving others functionally intact. Using this powerful technique, it is now possible to switch off the expression of individual genes in *T. brucei* and to study the resulting phenotypes [45–48]. RNAi is now being extended to study the estimated 6000 different genes that make up the *T. brucei* genome. Obviously, the techniques that result in gene silencing can be used only in the study of genes whose expression is not essential for viability of the trypanosome.

RNAi has been used in trypanosomes to 'knock down' rather than 'knockout' gene activity [31,47]. Such targeted stable suppression of gene expression offers an effective strategy for defining the function of essential genes and the identification of candidate targets for drug development and vaccine design. Thus, the availability of the complete *T. brucei* genome sequence will permit its functional analysis by systematic RNA interference. Such RNAi-based screening has been previously used to target nearly 90% of predicted genes on *Caenorhabditis elegans* chromosome I [49]. However, assigning biological function to thousands of predicted genes will be a major challenge and should involve an international consortium of laboratories. Within such collaboration, efficiency could be enhanced by recombinational cloning of two or more genes of interest in one plasmid. Using this approach, Kleingbeil *et al.* [31] were able to knock down the activity of multiple DNA polymerase genes in *T. brucei*.

The assessment of differential gene expression based on DNA methylation has so far not been explored in *T. brucei*. The sequence data available so far on the trypanosome genome indicate that the genome of this parasite comprises a relatively small number of genes and a large proportion of repetitive DNA. From detailed knowledge of methylation processes in other taxa [50,51], it is tempting to speculate that most of these repetitive sequences are more heavily methylated relative to functional genes. It is therefore, foreseeable that *T. brucei* genes unique to a subspecies could be enriched-for by the exclusion of repetitive DNA sequences in genomic shotgun libraries.

Microarrays

The strength of microarray technologies lies in their ability to follow genome-wide gene expression patterns in order to identify several hundreds of genes that might be involved in specific phenotypic traits in an organism. One important factor in determining the value of microarrays is the selection of DNA probes. In a pilot study, inserts from a *T. brucei* genomic library were arrayed randomly to generate 'shotgun' microarrays [26]. To measure variation in expression of genes in the bloodstream and procyclic (insect vector) stages of the parasite, the arrays were probed with differently labelled cDNAs prepared from total RNA of each of the developmental stages. Although many clones displayed constitutive levels of expression in both stages of the parasite, a substantial number of the clones showed different gene expression.

A major limitation of transcriptome analysis in trypanosomes is, in addition to those associated with this approach in general, the extent of gene regulation at the mRNA level. Almost all the resistance genes isolated so far from many eukaryotic organisms are expressed at low levels and are not significantly induced upon infection. This implies that transcriptome profiling used to isolate abundant transcripts might not be applicable for isolating genes responsible for human serum or drug resistance in trypanosomes. Therefore, it is predictable that patterns of gene expression from DNA microarray technology might not be the ideal method for discovering important targets in all instances. Proteomics provides a handy alternative in such circumstances. The extensive co- and post-transcriptional modification events that occur in trypanosomes, result in a diversity of protein products from a single ORF. Probes for microarray are usually derived from the 3′ ends of cDNA fragments but a homologous probe seems necessary to detect representative expression profiles from each alternatively spliced transcript. For trypanosomes, where fully annotated genome sequence data are not yet available, such homologous probes cannot yet be generated and only one isoform of a gene might be detectable. Furthermore, telomeric sequences tend to be under-represented in standard genomic libraries, a situation that would limit the value of DNA microarrays for the study of antigenic variation in these parasites. Therefore, a concerted effort to clone telomeric sequences via other means has been proposed [52].

Following the identification of several potentially useful targets for rational drug or vaccine design, their downstream development into 'leads' could be greatly facilitated if efficient technologies for mode-of-action studies were available. For example, exploring drug-induced alterations in gene expression by microarray may provide insight into the induction of genes known to encode pathways affected by a potential drug target. The results should, at least, provide possible correlation between changes in gene expression and the mode of action of the drug. However, the use of microarray analysis need not be limited to gene expression. Oligonucleotide arrays could be used in the rapid screening for SNPs and, thus, for exploring subspecies variation in trypanosomes.

Role of genomics in defining vaccine targets

A *T. brucei* genome sequence (or locus) could be considered a potential vaccine target if its gene product is involved in substantial parasite immune evasion or control, or in the generation of antigenic variants. The development of a broad-spectrum vaccine against trypanosomes meets formidable obstacles primarily because of the biology of this parasite and immune responses exhibited by the hosts in the course of infection. A major obstacle here is that the well-characterized variable surface glycoprotein (VSG) gene loci known to be primarily involved in immune evasion are located at telomeric and sub-telomeric sites on multiple chromosomes [52,53]. Immune responses to the VSG (the most abundant surface antigen of the trypanosomes) are not protective because of the vast repertoire of immunologically distinct variants. Furthermore, the outcome of the disease in an individual host depends on its genetic background and that of parasite strains responsible for the infection. Although the possibility of anti-disease vaccine for trypanosomosis has been

suggested [54], it can be inferred that developing a protective broad-ranging anti-trypanosomosis vaccine using conventional systems is currently unfeasible.

The unique aspects of the biochemistry of trypanosomes make rational anti-parasite vaccine design an attractive approach, but targets must be selected carefully with the aid of the approaches described above. Trypanosome genomics will continue to have a major impact on identification of potentially useful targets, including some enzymes, transporters and metabolites that might be unique to the parasite [3,55] (Table 2). The potential application of these molecules in immuno-therapeutics is being pursued. For example, RNAi has been used to study the possible therapeutic value of the conserved glycolipid (glycosylphosphatidylinositol, GPI) anchor, which attaches the VSGs to the cell membrane [56]. However, it turns out that the high conservation of the VSG GPI signal sequence is not necessary for efficient expression and GPI attachment, suggesting that the GPI anchor alone might not be a useful vaccine target. Similarly, many surface molecules, including some expected to be major structural components of the coat, can be mutated or completely eliminated without altering the morphology or growth rate of trypanosomes in culture [56]. Although this might suggest that the GPI molecules should not be the first choice for vaccine development, they might offer unique genetic options for identifying genes involved in the GPI synthesis pathway.

Other key invariant surface glycoproteins, transporters and receptor complexes have been identified [57–59] but only a few have been tested for their utility in development of vaccines, diagnostics and therapeutics. For example, the VSG molecules themselves contain structurally conserved C-termini [60,61], but these are so few and are generally so small as to be immunologically insignificant. Also, the flagellar pocket harbours some invariant receptor molecules that could be targets for vaccine design [62–66]. Although some of these molecules are distributed over the entire cell surface, they are inaccessible to the host immune response. Analogues to myristate, the only fatty acid component of the glycolipid anchor, appeared to have trypanocidal activity *in vitro* [9] but the bioavailability of active compounds in the cell will need to be enhanced for sustained activity *in vivo*. A pragmatic model would be to traffic toxic molecules into these invariant target ligands. For example, RNA aptamers have been exclusively expressed in the flagellar pocket [67], making the approach attractive for rational vaccine design. Another possibility could be an agent interacting with the 35 nucleotide sequences at the extreme 5′ termini of all mRNAs in trypanosomes but not in mammalian host mRNA. From these studies, knowledge has also been gained on how inhibitory a compound needs to be in order to be an effective trypanocide.

Of special significance to future drug or vaccine development will be elucidating the role of genetic variation on the efficiency of patient responses to drug treatment or vaccine products (i.e. pharmacogenetics effect). Large-scale resequencing will be a valuable tool for the analysis of a known sequence in many individuals to identify variations in the sequence, to measure such genetic variation in populations. Also, efficient mutation-detection technologies are facilitating the identification and analysis of genetic polymorphisms, and the determination of their effects on cellular systems.

Table 2. Candidate immunotherapeutic targets

Molecular target	Cellular activity	Refs
Acyl carrier proteins	Fatty acid biosynthesis	[74–76]
Inhibitors of mitochondrial function	Glycolysis	[77–79]
Tubulin multigene family	Cytoskeleton	[80]
Cysteine proteinases	Proteolytic enzymes	[81–83]
Plasma membrane proteins	Cellular physiology	[84]
Cytosolic fractions	Cellular biochemistry	[85]

Conclusion

There are problems to consider and many hurdles to cross in the development of tools for a trypanosome vaccine, new therapeutics or specific diagnostics. Genome sequence data, differential gene expression analysis and comparative genomics are contributing enormously to identifying valuable candidate targets. Although the definition of these targets is being aided by the rapidly evolving functional analysis technologies, clearly even with trypanosome genome sequence, efficient and customized computational platforms and gene modelling tools will play a central role. Improvements in bioinformatics will enhance the analysis and interpretation of data generated from the genetics and genomics techniques. This tool will probably have a central role in the identification of new drug discovery targets, in rational drug design and in interpreting biological and clinical data.

A field that will play an increasing role in the rational drug and vaccine design is structural biology. Structural determinations through crystallography, nuclear magnetic resonance (NMR) and protein modelling represent an important interface in the drug discovery process. It is our opinion that the specific applications of the genomic technologies discussed in this article will accelerate the pace at which putative lead targets will be identified and validated. In addition, improving technologies for the identification of genetic locus (or loci) of interest through candidate gene approach implies that identification of genes of interest will be more efficient in the future.

References

1 Hawking, F. (1973) The differentiation of *Trypanosoma rhodesiense* from *T. brucei* by means of human serum. *Trans. R. Soc. Trop. Med. Hyg.* 67, 517–527
2 World Health Organisation (2001) African trypanosomiasis (Fact Sheet number 259), WHO Publications, Geneva
3 Keiser, J. *et al.* (2001) New drugs for the treatment of human African trypanosomiasis: research and development. *Trends Parasitol.* 17, 42–49
4 Anene, B.M. *et al.* (2001) Drug resistance in pathogenic African trypanosomes: what hope for the future? *Vet. Parasitol.* 96, 83–100
5 Kuzoe, F.A.S. (1993) Current situation of African trypanosomiasis. *Acta Trop.* 54, 153–162
6 World Health Organisation (1998) Control and surveillance of African trypanosomiasis. *Report of WHO Expert Committee, Technical Report Series* 881, I–Vi:1114
7 Milord, F. *et al.* (1993) Eflornithine concentrations in serum and

cerebrospinal fluid of 63 patients treated for *Trypanosoma brucei gambiense* sleeping sickness. *Trans. R. Soc. Trop. Med. Hyg.* 87, 473–477
8 Jennings, F.W. (1993) Combination chemotherapy of CNS trypanosomiasis. *Acta Trop.* 54, 205–213
9 Werbovetz, K.A. *et al.* (1996) Trypanocidal analogs of myristate and myristoyllysophosphatidyl-choline. *Mol. Biochem. Parasitol.* 81, 115–118
10 Doering, T.L. *et al.* (1994) Toxicity of myristic acid analogs toward African trypanosomes. *Proc. Natl. Acad. Sci. U. S. A.* 91, 9735–9739
11 Hide, G. *et al.* (1990) The identification of *Trypanosoma brucei* subspecies using repetitive DNA sequences. *Mol. Biochem. Parasitol.* 39, 213–225
12 Stevens, J.R. and Tibayrenc, M. (1995) Detection of linkage disequilibrium in *Trypanosoma brucei* isolated from tsetse flies and characterised by RAPD analysis and isoenzymes. *Parasitology* 110, 181–186
13 Agbo, E.C. *et al.* (2001) Measure of molecular diversity within the *Trypanosoma brucei* subspecies *Trypanosoma brucei brucei* and *Trypanosoma brucei gambiense* as revealed by genotypic characterisation. *Exp. Parasitol.* 99, 123–131
14 Agbo, E.C. *et al.* (2002) Molecular variation of *Trypanosoma brucei* subspecies as revealed by AFLP fingerprinting. *Parasitology* 124, 349–358
15 Gibson, W.C. and Mizen, W.H. (1997) Heritability of the trait for human infectivity in genetic crosses of *T. brucei* spp. *Trans. R. Soc. Trop. Med. Hyg.* 91, 236–237
16 Gibson, W.C. and Stevens, J. (1999) Genetic exchange in the Trypanosomatidae. *Adv. Parasitol.* 43, 1–45
17 Tait, A. *et al.* (2002) Genetic analysis of phenotype in *Trypanosoma brucei*: a classical approach to potentially complex traits. *Philos. Trans. R. Soc. London Ser. B* 357, 89–99
18 El-Sayed, N.M. *et al.* (2000) The African trypanosome genome. *Int. J. Parasitol.* 30, 329–345
19 Lanzer, M. *et al.* (1993) Transcriptional differences in polymorphic and conserved domains of a complete cloned *P. falciparum* chromosome. *Nature* 371, 654–657
20 Ravel, C. *et al.* (1996) Medium-range restriction maps of five chromosomes of *Leishmania infantum* and localisation of size-variable regions. *Genomics* 35, 509–516
21 The C. elegans sequencing Consortium, (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282, 2012–2018
22 Cole, S.T. *et al.* (2001) Massive gene decay in the leprosy bacillus. *Nature* 409, 1007–1011
23 Gardner, M.J. *et al.* (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419, 498–511
24 Hannaert, V. *et al.* (2003) Plant-like traits associated with metabolism of *Trypanosoma* parasites. *Proc. Natl. Acad. Sci. U. S. A.* 100, 1067–1071
25 El-Sayed, N.M. *et al.* (1995) cDNA expressed sequence tags of *Trypanosoma brucei rhodesiense* provide new insights into the biology of the parasite. *Mol. Biochem. Parasitol.* 73, 75–90
26 El-Sayed, N.M. and Donelson, J.E. (1997) A survey of the *Trypanosoma brucei rhodesiense* genome using shotgun sequencing. *Mol. Biochem. Parasitol.* 84, 167–178
27 Djikeng, A. *et al.* (1998) Generation of expressed sequence tags as physical landmarks in the genome of *Trypanosoma brucei*. *Gene* 221, 93–106
28 MacLeod, A.A. *et al.* (2000) Minisatellite marker analysis of *Trypanosoma brucei*: reconciliation of clonal, panmictic, and epidemic population genetic structures. *Proc. Natl. Acad. Sci. U. S. A.* 97, 13442–13447
29 Hide, G. *et al.* (1994) Epidemiological relationships of *Trypanosoma brucei* stocks from South-east Uganda: evidence for different populations structures in human infective and non-human infective isolates. *Parasitology* 109, 95–111
30 Lukeš, J. *et al.* (2002) Kinetoplast DNA network: evolution of an improbable structure. *Eukaryot. Cell* 1, 495–502
31 Klingbeil, M.M. *et al.* (2002) Multiple mitochondrial DNA polymerases in *Trypanosoma brucei*. *Mol. Cell* 10, 1–20
32 Agbo, E.C. *et al.* (2003) Multiplex-endonuclease genotyping approach (MEGA): a tool for the fine-scale detection of unlinked polymorphic DNA markers. *Chromosoma* 111, 518–524
33 De Greef, C. *et al.* (1989) A gene expressed only in serum-resistant variants of *Trypanosoma brucei rhodesiense*. *Mol. Biochem. Parasitol.* 36, 169–176
34 Rifkin, M.R. *et al.* (1994) Human serum-sensitive *Trypanosoma brucei rhodesiense*: a comparison with serologically identical human serum-resistant clones. *Mol. Biochem. Parasitol.* 66, 211–220
35 Smith, A.B. *et al.* (1995) Killing of trypanosomes by haptoglobin-related protein. *Science* 268, 284–286
36 Raper, J. *et al.* (2001) Trypanosome lytic factors: novel mediators of human innate immunity. *Curr. Opin. Microbiol.* 4, 402–408
37 Hatada, S. *et al.* (2002) No trypanosome lytic activity in the sera of mice producing human haptoglobin-related protein. *Mol. Biochem. Parasitol.* 119, 291–294
38 VanHamme, L. *et al.* (2003) Apolipoprotein L-I is the trypanosome lytic factor of human serum. *Nature* 422, 83–87
39 Welburn, S.C. *et al.* (2001) Identification of human-infective trypanosomes in animal reservoir of sleeping sickness in Uganda by means of serum-resistance-associated (SRA) gene. *Lancet* 358, 2017–2019
40 Gibson, W. *et al.* (2002) The human resistance associated gene is ubiquitous and conserved in *Trypanosoma brucei rhodesiense* throughout East Africa. *Infect. Genet. Evol.* 1, 207–214
41 Ngo, H. *et al.* (1998) Double-stranded RNA induced mRNA degradation in *Trypanosoma brucei*. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14687–14692
42 Drozdz, M. *et al.* (2002) RNA interference in trypanosomes transfected with sense and antisense plasmids. *Mol. Biochem. Parasitol.* 121, 149–152
43 Ullu, E. *et al.* (2002) RNA interference: advances and questions. *Philos. Trans. R. Soc. London Ser. B* 357, 65–70
44 Hannon, G.J. (2002) RNA interference. *Nature* 418, 244–251
45 Wirtz, E. *et al.* (1999) A tightly regulated inducible expression system for dominant negative approaches in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* 99, 89–101
46 Wang, Z. *et al.* (2000) Inhibition of *Trypanosoma brucei* gene expression by RNA interference using an integratable vector with opposing T7 promoters. *J. Biol. Chem.* 275, 40174–40179
47 Wang, Z. *et al.* (2002) Asymmetrical division of the kinetoplast DNA network of the trypanosome. *EMBO J.* 21, 4998–5005
48 Wang, Z. and Englund, P.T. (2001) RNA interference of a trypanosome topoisomerase II causes progressive loss of mitochondrial DNA. *EMBO J.* 20, 4674–4683
49 Fraser, A.G. *et al.* (2000) Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* 408, 325–330
50 Martienssen, R.A. and Richards, E.J. (1995) DNA methylation in eukaryotes. *Curr. Opin. Genet. Dev.* 5, 234–242
51 Martienssen, R. (1998) Transposons, DNA methylation and gene control. *Trends Genet.* 14, 263–264
52 Rudenko, G. (2000) The polymorphic telomeres of the African trypanosome – *Trypanosoma brucei*. *Biochem. Soc. Trans.* 28, 536–540
53 Rudenko, G. *et al.* (1998) Changing the end: antigenic variation orchestrated at the telomeres of African trypanosomes. *Trends Microbiol.* 6, 113–116
54 Authie, E. (1994) Trypanosomiasis and trypanotolerance in cattle: a role for congopain? *Parasitol. Today* 10, 360–364
55 Barrett, M.P. *et al.* (1999) Recent advances in identifying and validating drug targets in trypanosomes and leishmanias. *Trends Microbiol.* 7, 82–88
56 Bohme, U. and Cross, G.A.M. (2001) Mutational analysis of the variant surface glycoprotein GPI-anchor signal sequence in *Trypanosoma brucei*. *J. Cell Sci.* 115, 805–816
57 Seyfang, A. and Duszenko, M. (1993) Functional reconstitution of the *Trypanosoma brucei* plasma-membrane D-glucose transporter. *Eur. J. Biochem.* 214, 593–597
58 Borst, P. *et al.* (1998) Control of VSG gene expression sites in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* 91, 67–76
59 Pays, E. and Nolan, D.P. (1998) Expression and function of surface proteins in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* 91, 3–36
60 Carrington, M. *et al.* (1991) Variant specific glycoprotein of *Trypanosoma brucei* consists of two domains each having an independently conserved pattern of cysteine residues. *J. Mol. Biol.* 221, 823–835
61 Carrington, M. and Boothroyd, J. (1996) Implications of conserved

structural motifs in disparate trypanosome surface proteins. *Mol. Biochem. Parasitol.* 81, 119–126
62 Da Cunha e Silva, N.L. *et al.* (1989) Isolation and characterisation of a highly purified flagellar membrane fraction from trypanosomatids. *Mol. Biochem. Parasitol.* 37, 129–136
63 Engman, D.M. *et al.* (1989) A novel flagellar Ca^{2+}-binding protein in trypanosomes. *J. Biol. Chem.* 264, 18627–18631
64 Paindavoine, P. *et al.* (1992) A gene from the variant surface glycoprotein expression site encodes one of several transmembrane adenylate cyclases located on the flagellum of *Trypanosoma brucei*. *Mol. Cell. Biol.* 12, 1218–1225
65 Nozaki, T. *et al.* (1996) Characterisation of the *Trypanosoma brucei* homologue of a *Trypanosoma cruzi* flagellum-adhesion glycoprotein. *Mol. Biochem. Parasitol.* 82, 245–255
66 Landfear, S.M. and Ignatushchenko, M. (2001) The flagellum and flagellar pocket of trypanosomatids. *Mol. Biochem. Parasitol.* 115, 1–17
67 Homnann, M. and Ulricht, G.H. (2001) Uptake and intracellular transport of RNA aptamers in African trypanosomes suggest therapeutic 'piggy-back' approach. *Bioorg. Med. Chem.* 9, 2571–2580
68 Radwanska, M. *et al.* (2002) The serum resistance-associated gene as a diagnostic tool for the detection of *Trypanosoma brucei rhodesiense*. *Am. J. Trop. Med. Hyg.* 67, 684–690
69 Sloof, P. *et al.* (1983) Size fractionation of *Trypanosoma brucei* DNA: localisation of the 177-bp repeat satellite DNA and a variant surface glycoprotein gene in a mini-chromosomal DNA fraction. *Nucleic Acids Res.* 11, 3889–3901
70 Imboden, M. *et al.* (1995) Repetitive proteins from the flagellar cytoskeleton of African trypanosomes are diagnostically useful antigens. *Parasitology* 110, 249–258
71 Masake, R.A. *et al.* (2002) The application of PCR–ELISA to the detection of *Trypanosoma brucei* and *T. vivax* infections in livestock. *Vet. Parasitol.* 105, 79–89
72 Desquesnes, M. *et al.* (2001) Detection and identification of Trypanosoma of African livestock through a single PCR based on internal transcribed spacer 1 of rDNA. *Int. J. Parasitol.* 31, 610–614
73 Kabiri, M. *et al.* (1999) Detection of *Trypanosoma brucei gambiense* in sleeping sickness suspects by PCR amplification of expression-site-associated genes 6 and 7. *Trop. Med. Int. Health* 4, 658–661
74 Yasu, S.M. *et al.* (2000) Specialised fatty acid synthesis in African trypanosomes: myristate for GPI anchors. *Science* 288, 140–143
75 Price, H.P. *et al.* (2003) Myristoyl-CoA:protein *N*-myristoyltransferase, an essential enzyme and potential drug target in kinetoplastid parasites. *J. Biol. Chem.* 278, 7206–7214
76 Coichi, N. *et al.* (2002) Trypanosome alternative oxidase as a target of chemotherapy. *Biochim. Biophys. Acta* 1587, 234–239
77 Lakhdar-Ghazal, F. *et al.* (2002) Glycolysis and proteases as targets for the design of new anti-trypanosome drugs. *Curr. Top. Med. Chem.* 2, 439–456
78 Omura, S. *et al.* (2001) An anthelmintic compound, nafuredin, shows selective inhibition of complex I in helminth mitochondria. *Proc. Natl. Acad. Sci. U. S. A.* 98, 60–62
79 Balaban, N. *et al.* (1995) Intracellular antigens (microtubule-associated protein copurified with glycosomal enzymes) as possible vaccines against trypanosomiasis. *J. Infect. Dis.* 172, 845–850
80 Lubega, G.W. *et al.* (2002) Immunisation with a tubulin-rich preparation from *Trypanosoma brucei* confers broad protection against African trypanosomosis. *Exp. Parasitol.* 102, 9–22
81 Caffrey, C.R. *et al.* (2000) Cysteine proteinases of trypanosome parasites: novel targets for chemotherapy. *Curr. Drug Targets* 1, 155–162
82 Troeberg, L. *et al.* (1999) Cysteine proteinase inhibitors kill cultured bloodstream forms of *Trypanosoma brucei brucei*. *Exp. Parasitol.* 91, 349–355
83 Eakin, A.E. *et al.* (1990) Amplification and sequencing of genomic DNA fragments encoding cysteine proteases from protozoan parasites. *Mol. Biochem. Parasitol.* 39, 1–8
84 Murray, M. *et al.* (1985) Suppression of Trypanosoma congolense, *T. vivax* and *T. brucei* infection in tsetse flies maintained on goats immunised with uncoated forms of trypanosomes grown *in vitro*. *Parasitology* 91, 53–66
85 Beat, G.A. *et al.* (1984) Non-variant antigens limited to blood-stream forms of *Trypanosoma brucei brucei* and *Trypanosoma brucei rhodesiense*. *J. Protozool.* 31, 541–548

Trends in Microbiology online – making the most of your personal subscription

High-quality printouts (from PDF files)
Links to other articles, other journals and cited software and databases

All you have to do is:

Obtain your subscription key from the address label of your print subscription.
Then go to http://www.trends.com, click on the Claim online access button and select *Trends in Microbiology*.
You will see a BioMedNet login screen. Enter your BioMedNet username and password. If you are not already a BioMedNet member, please click on the Register button. Once registered, you will be asked to enter your subscription key.
Following confirmation, you will have full access to *Trends in Microbiology*.

If you obtain an error message please contact Customer Services (info@current-trends.com) stating your subscription key and BioMedNet username and password. Please note that you only need to enter your subscription key once; BioMedNet 'remembers' your subscription.
Institutional online access is available at a premium.
If your institute is interested in subscribing to print and online, please ask them to contact ct.subs@qss-ik.com