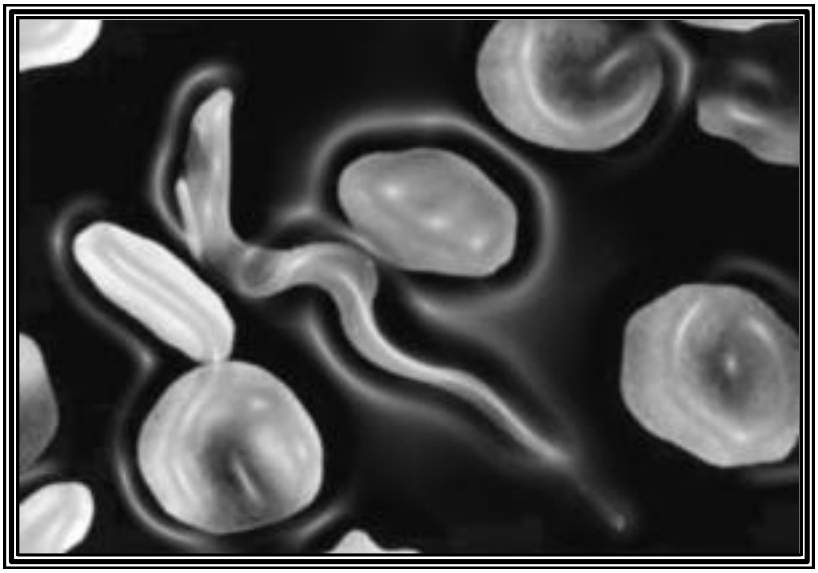
UNIVERSITÉ DE NEUCHÂTEL **FACULTÉ DES SCIENCES**

INSTITUT DE ZOOLOGIE

ANALYSIS OF HISTONE H1 GENES AND GENE PRODUCTS IN *TRYPANOSOMA BRUCEI*



PAR

ERIC GRUETER

THÈSE PRÉSENTÉE À LA
FACULTÉ DES SCIENCES DE L'UNIVERSITÉ DE NEUCHÂTEL
POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

NEUCHÂTEL 2000

IMPRIMATUR POUR LA THESE

Analysis of histone H1 genes and gene products in *Trypanosoma brucei*

de M. Eric Grueter

UNIVERSITE DE NEUCHATEL

FACULTE DES SCIENCES

La Faculté des sciences de l'Université de Neuchâtel sur le rapport des membres du jury,

MM. B. Betschart (directeur de thèse),
J.-M. Neuhaus, N. Fasel (Lausanne) et
H. Hecker (Bâle)

autorise l'impression de la présente thèse.

Neuchâtel, le 11 janvier 2001

Le doyen:

J.-P. Derendinger

Contents:

Scroll the document or use the "**go to page**" function according to the table below!

Summary

Summary (English version):

For many years, trypanosomes were thought to be deficient in histone H1, explaining in some way the observed absence of chromatin condensation during cell division. However, histone H1 proteins displaying different biochemical properties if compared to their counterparts in higher eucaryotes were highlighted successively in different protozoan. Especially in kinetoplastids, H1 proteins are of astonishingly small size and they correspond mainly to the C-terminal tail of classic H1s. Finally, H1 proteins were also showed to be expressed in the African trypanosome, *Trypanosoma brucei*, the causative agent of African sleeping sickness afflicting thousands of people in central Africa and also of Nagana in cattle which causes severe economic loss. These H1s stayed undiscovered for long time because of their electrophoretic mobility which made them migrate with the core histones and thus making them to be hidden by these latter. Special gel systems allowed to highlight them, revealing four fast migrating histone H1 variant and/or posttranslational modifications in front of the core histones. Differences were also seen by comparing the two stages of the life-cycle and one variant appeared to be over-expressed in the bloodstream form of the parasite. Study of histone H1 and core histones biochemical properties, amino acids composition, modifications and implications in chromatin condensation by in vitro reconstitution experiments showed that trypanosomes display an original nuclear organization and that their DNA processes greatly differ from those found in higher eucaryotes. All these differences made some authors emit the hypothesis that histones may provide new potential targets for trypanocidal drugs that would have no side effects on the host. In addition histones which are usually known to be among the most conserved proteins in evolution do present high rates of divergences among different protists. Therefore, it was also put forward that knowledge of histone sequences may proof to be a good tool for reexamining the phylogenetic relationships among protozoa.

Histone H1 sequences of several lower eucaryotes have been published in data banks and some partial peptide fragments were also obtained from African trypanosomes. However, to this date, the gene sequences coding for histone H1 in *Trypanosoma brucei* were still missing. The main purpose of the present work was to isolate and to characterize H1 genes in this organism.

In this work, a gene family composed of several closely related genetic variants which differ in size from each other by some deletions which do not affect their general primary structure is described. The obtained results are in good correlation with former studies on African trypanosomes and also with the situation described in *Trypanosoma cruzi*, the causative agent of Chagas disease in southern America. However, there are also great divergences between African and American trypanosomes H1 which rely mainly on different sequence motifs which can have important implications for posttranslational modification. On the other hand, the general structure, amino acid composition, number of variants and genomic organization are comparable.

Different subspecies of African trypanosomes, *Trypanosoma brucei brucei* and *Trypanosoma brucei gambiense*, do exhibit divergent H1 genes. Four H1 size-variants were isolated in the former subspecies and at least seven in the latter. Each size class showed to be composed of several genes which differ from each other by few nucleotides which do or do not have an effect on the gene products. In addition, evidence that this heterogeneity of size classes seems to be present even by comparing different strains of the same subspecies is also reported.

The deduced amino acid composition appears to be very simple, giving rise to highly basic proteins with theoretical molecular weights of 6 to 8 kDa migrating along the 14 kDa band marker in normal 15% SDS PAGE gels.

The genes are organized in a head to tail fashion in the genome. In addition, the isolation and sequencing of several tandemly arranged genes showed all typical features seen in trypanosomes in regard to the sequence elements which are located between genes of a polycistronic transcription unit. This intergenic sequence appears to be very conserved and quite identical among all H1 tandems that were analyzed.

One H1 variant does present a different N-terminal amino acid sequence, and the gene coding for it is located in front of all other genes in the cluster. The other variants do present another N-termini which is conserved among all of them.

The transcripts, once they are processed by the cell, are polyadenylated and trans-spliced. The potential implication of sequence motifs located in the intergenic region on expression regulation are discussed, but the global transcription rates between the two life stages of the parasite did not reveal to be significantly different.

Recently, an additional H1 variant which exhibits an additional different N-terminus was isolated. This gene is preceded by a different non-coding sequence and it is not closely linked in the genome to the other H1 genes. The proposal is made that there are two

different H1 clusters located on the same chromosome. The first cluster is composed of several gene copies from which the first in the row codes for an additional specific N-terminal region of 12 amino acids. The second cluster contains only two copies from which the first does code for a variant that exhibits a pair of threonines at position 4-5 that differentiates this variant from the variants in the other cluster.
It may well be that these two clusters are differentially regulated over the cell cycle and/or between the parasitic life stages. This could explain the differences in the expression of one variant which was described elsewhere. However, transcripts coding for this new variant are shown to be present in both stages of the life cycle.
Finally, antibodies against recombinant H1 proteins were raised in mice. The obtained serum showed strong specificity for H1 in *T. brucei* and also cross-reacted with calf thymus histone H1, probably recognizing its C-terminal tail. Confocal analysis showed H1 to be confined in the nucleus and to be well linked to DNA distribution over the whole cell cycle in both the procyclic culture form and the bloodstream form.

Knowledge of the gene sequences, their general organization in the genome and the availability of a very specific antibody should now allow further investigations which were not possible before. The implication in chromatin condensation as well as potential other functions of these primitive H1 can now be envisaged in more details and the fine localization of H1 in trypanosomes also appears easier.

Résumé (French version):

Trypanosoma brucei est l'agent de la maladie du sommeil qui touche des milliers de personnes en Afrique intertropicale ainsi que du Nagana qui touche le bétail et cause d'énormes pertes économiques dans ces régions. Le cycle de vie de ce parasite est dixène, passant alternativement par un hôte vertébré dans lequel il se développe dans le sang et le système nerveux et par un vecteur qui est un diptère hématophage (Glossina sp.). L'absence de condensation de la chromatine dans le cycle cellulaire des trypanosomes fut expliqué par l'apparente absence d'histone H1, toutefois, des H1 démontrant des propriétés biochimiques différentes de celles connues chez les eucaryotes supérieurs furent décrits successivement pour divers protozoaires. Ces H1, et plus particulièrement ceux trouvés chez les kinetoplastidés, sont d'étonnante petite taille et ne correspondent qu'à la partie C-terminale d'un histone H1 classique. Finalement, H1 fut aussi mis en évidence chez le trypanosome africain. Ces protéines restèrent inaperçues très longtemps à cause de leur co-migration avec les histones nucléosomiques (H2A/H2B/H3/H4). Un groupe de quatre variants et/ou modifications posttraductionnelles migrant en tête des autres histones sur gel Triton-urée a été décrit pour les deux stades de vie du parasite. Toutefois, la comparaison des bandes obtenues pour les deux stades du cycle de vie du parasite révèlent des différences et un des quatre variants et/ou modifications posttraductionnelles est sur-exprimé dans la forme sanguine.

L'étude des H1 ainsi que des histones en général au niveau de leurs propriétés biochimiques, leurs compositions en acides aminés, leurs modifications et leurs implications dans la condensation de la chromatine démontre que les trypanosomes possèdent une organisation nucléaire originale qui pourrait avoir une incidence sur les mécanismes de régulation génétique. Toutes ces différences poussèrent certains auteurs à émettre l'hypothèse que les histones pourraient se révéler être de nouvelles cibles potentielles pour le développement de trypanocides qui n'auraient pas d'effets secondaires sur l'hôte. De plus, les histones, considérés comme étant très conservés dans l'évolution, présentent des degrés élevés de divergence chez les protozoaires. C'est pourquoi l'idée selon laquelle la connaissance des séquences des histones pourrait constituer un outil puissant pour réexaminer les relations phylogénétiques parmi les protozoaires à également été avancée.

Plusieurs séquences codants pour H1 chez divers protozoaires et quelques fragments peptidiques correspondant aux H1 de trypanosomes africains ont été analysés. Toutefois, les séquences génétiques codants pour H1 chez *Trypanosoma brucei* ne sont pas connues. Les principaux buts de ce travail étaient donc d'isoler et de caractériser les gènes codants pour H1 dans cet organisme.

Une famille de gènes composée de différents variants qui diffèrent entre eux par des délétions qui n'affectent pas leur structure primaire générale est décrite. Les résultats sont en bonne corrélation avec les études précédentes qui ont porté sur les H1 de *Trypanosoma brucei* ainsi qu'avec la situation qui a été décrite pour *Trypanosoma cruzi*, l'agent de la maladie de Chagas en Amérique du sud. Cependant, il existe de grandes divergences entre histone H1 de trypanosomes africains et américains au niveau de motifs d'acides aminés pouvant avoir de grandes implications dans les modifications posttraductionnelles. D'un autre côté, la structure générale, la composition en acides aminés ou le nombre de variants et leur organisation sont plutôt comparables.

Ces divergences se retrouvent aussi en comparant les gènes de deux sous-espèces de trypanosomes africains, *Trypanosoma brucei brucei* et *Trypanosoma brucei gambiense*. Dans la première, les différents variants peuvent être groupés en 4 classes de tailles alors que dans la deuxième, au moins 7 classes ont été isolées. De plus, chaque classe de taille est composée de plusieurs gènes qui diffèrent entre eux par quelques substitutions nucléotidiques qui ont ou n'ont pas d'incidences sur le niveau de la protéine. A ceci s'ajoute le fait que cette hétérogénéité s'observe même entre différentes souches de la même sous-espèce.

La composition en acides aminés de ces histones s'avère très simple, produisant des protéines hautement basiques avec des poids moléculaires théoriques entre 6 et 8 kDa. Toutefois, sur des gels SDS PAGE de 15%, elles migrent aux environs du marqueur de 14 kDa.

Les gènes sont organisés en unités polycistroniques dans le génome et l'isolation et le séquençage de plusieurs tandems ont montré que les régions intergéniques présentent des particularités communes à d'autres gènes déjà décrites chez les trypanosomes. Ces séquences non codantes sont très conservées si l'on considère tous les tandems qui ont été analysés.

L'un des variants de H1 décrit diffère des autres par une courte séquence N-terminale et le gène codant pour ce dernier est localisé en tête de l'unité polycistronique. Tous les autres variants possèdent une autre région N-terminale conservée.

Après épissage et maturation, les transcrits portent une queue poly-A^+ et le mini-exon (spliced leader).

Les implications potentielles de différents motifs de séquences sur la régulation de l'expression sont discutées, mais l'analyse globale de la transcription des H1 entre les deux stades de vie du parasite n'a pas montré de différence significative.

Récemment, un nouveau type de gène qui présente un troisième motif N-terminal a été isolé. Ce gène est précédé par une séquence non codante différente de celles déjà connues pour les deux autres types et aucun lien topographique entre ce nouveau type et les deux autres n'a pu être vérifié. L'hypothèse est émise que les gènes codants pour H1 sont groupés en deux unités sur le même chromosome. La première est composée de plusieurs gènes très similaires dont le premier code pour une séquence N-terminale additionnelle de 12 acides aminés. La seconde ne contient que deux copies dont la première présente une paire de thréonines en positions 4 et 5 qui la différentie nettement de l'autre unité.

Il se peut que ces deux unités soient régulées différemment au cours du cycle cellulaire ou entre les deux stades du cycle de vie parasitaire. Ceci pourrait expliquer la différence d'expression de l'un des variant qui a été décrite ailleurs. Toutefois, des transcrits codant pour ce nouveau variant ont été mis en évidence dans les deux stades du parasite.

Finalement, la production d'anticorps anti-H1 dans des souris à l'aide de protéines recombinantes a permit d'obtenir un sérum très spécifique qui reconnaît également les histones H1 de thymus de veau. Des analyses au microscope confocal ont montré que les H1 sont localisés exclusivement dans le noyau et qu'ils suivent très bien la distribution de l'ADN tout au long du cycle cellulaire aussi bien dans la forme procyclique que dans la forme sanguine du parasite.

La connaissance des gènes codants pour H1, de leur organisation dans le génome ainsi que la disponibilité d'un anticorps spécifique devraient maintenant permettre des études qui n'étaient pas possibles auparavant comme, par exemple, l'implication de ces histones H1 primitifs dans la condensation de l'ADN ainsi que leur(s) éventuelle(s) fonction(s) additionnelle(s).

I. INTRODUCTION

GENERAL INTRODUCTION

Current situation in African trypanosomiasis:

African trypanosomiasis (sleeping sickness) is a severe affection of human caused by the hemoflagellate *Trypanosoma brucei*, a protozoan parasite belonging to the phylum Sarcomastigophora. *T. brucei* is a member of the order Kinetoplastida and is attributed to the family Trypanosomatidae. This family contains several of the main causative agents of diseases found in man and livestocks such as Chagas disease in south America or leishmaniasis. In addition, Nagana, another form of trypanosomiasis affecting mainly ungulates in intertropical Africa by causing chronic anemia, reduced growth and diminished fertility as well as rapid death of unadapted species is also found.

Human trypanosomiasis exists in two forms which differ from each other mainly by the time course of the disease but which are both fatal without appropriate treatment. An acute form caused by *Trypanosoma brucei rhodesiense* which runs a rapid course leading to death in a few weeks is found in east and south Africa while in west and central Africa, *Trypanosoma brucei gambiense* causes a chronic form of the disease which can last several month or even years until severe neurological symptoms appear [Kuzoe, 93].

Transmission occurs through the bite of an haematophagous insect, the tsetse fly (*Glossina spp*) (**Fig. 1**) and the geographical distribution of the disease follows ecological requirements of the vector (**Fig. 2**, inset). The favored habitat of glossines is the vegetation along watercourses, forest edges and gallery forests in the western regions with *Glossina palpalis* as a main representant and extends to vast areas of scrub savanna in the eastern regions with xerophile tsetse species like *Glossina morsitans*.

The epidemiological situation also shows differences at the level of possible reservoirs which can have great impact on efficiency of control programs. In the west form of the disease, only man and perhaps some domestic animals are reservoir hosts whereas in the east form, a large number of domestic and wild animals can be infected and cause maintenance of the parasite.

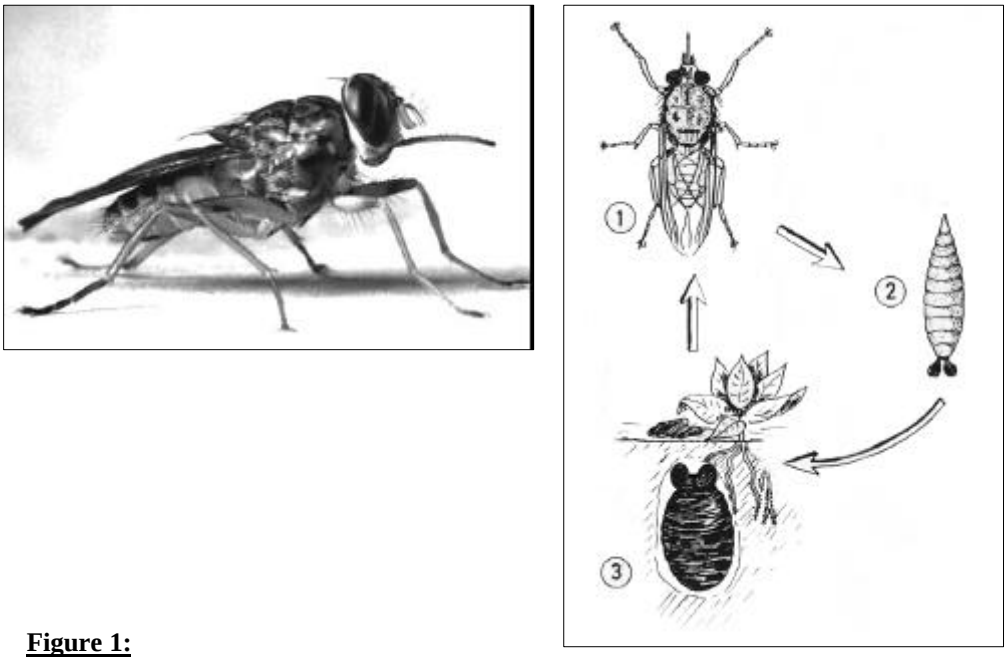


Figure 1:
General view of tsetse fly. Both sexes are bloodfeeding. During its life, the female gives birth to completely developed larvae, (one at a time) which buries itself under a few centimeters of loose soil. Hardening and darkening of the tegument to form the puparium occurs rapidly after larvi-position. The adult emerges within 2 to 4 weeks. (Redrawn from Golvan Y.J., 1983; photography from WHO website).

Human African trypanosomiasis covers vast areas and is rather a rural and focal problem. It is estimated that 60 million people living in 36 sub-Saharan countries (**Fig. 2**) are exposed to the risk of becoming infected but only about 45’000 new cases were reported in 1999 (WHO website). This represents clearly an underestimation since less than 10% of the people living at risk are under active surveillance or have access to health centers where reliable diagnosis is available. The real epidemiological situation is much more dramatic and the estimation of at least 300'000 to 500’000 infected persons was made (WHO website). By the 1960s, trypanosomiasis was brought under good control, but since 1970 the situation has deteriorated with recrudescence of old foci and a spread over new areas principally in countries which have not maintained surveillance at a sufficient level. This was caused by several factors of different orders like civil wars, political disorders and economic problems all linked to massive population displacements. The emergence of other burdens like AIDS which monopolized attention or the progressive disinterest of developed countries which allowed less financial or logistic support are certainly also involved [Kuzoe, 93]. The relatively small number of declared infections per year if compared to other parasitic diseases [Hirst, 00] clearly gives trypanosomiasis a

misestimated priority and recent epidemiological programs clearly show that the disease is in expansion reaching the epidemic situation prevalent in the 1930s (EpiTryp program, see link from WHO website).

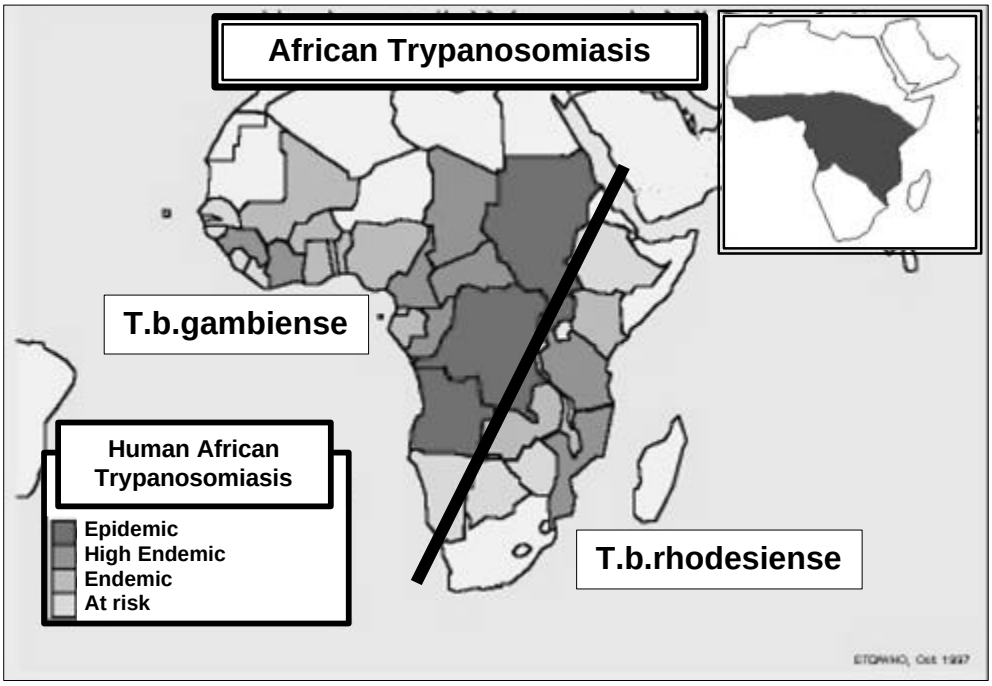


Figure 2: Distribution of African trypanosomiasis and indication of epidemics level. The black line shows the frontier between west and east forms of the disease. Upper right inset, global geographical distribution of tsetse fly. (Adapted from WHO website)

Both, human and animal trypanosomiasis have considerable impact on the economic development of rural regions or countries by making whole fertile areas being abandoned, by reducing labor force and by making farming and cattle production impossible. Some authors estimated already 15 years ago that the global beef loss was 1,5 x 10^6 tons per year [Allsopp, 1985] and 10 million square kilometers of potential grazing land are rendered unsuitable for livestock breeding [Kuzoe, 93].

It appears evident that African trypanosomiasis remains one of the main problems for long term development in sub-Saharan countries which clearly depend on external resources. So far, no vaccine is available and major research plans in this direction were abandoned in recent years principally because of the astonishing ability of trypanosomes to perform antigenic variation (see later). The programs that are currently active for disease control rely on different targets from which the main ones are regular medical surveillance of the population at risk involving early diagnosis and systematic treatment of patients as well as

tsetse fly control principally by the use of traps containing different attractants [Kuzoe, 93]. However, whereas the fly control is giving good results in certain regions, the available trypanocidal drugs do not constitute ideal treatments for third world countries. They do all present severe side effects, are not always efficient against both forms or are only active on early stages of infection. The treatment needs qualified medical staff and patient surveillance and, in addition, they are expensive (WHO website). This clearly shows that new and safe targets are needed for future drug developments. A large number of target molecules and metabolic pathways have been pointed out, however, the number of molecules based on selective interference with these targets or pathways is very small (WHO website). Even if an astonishing knowledge of metabolic pathways, virulence, immunity, differentiation or gene regulation and antigenic variation was gained over the past decades, sleeping sickness is still in expansion [Denise, 99].

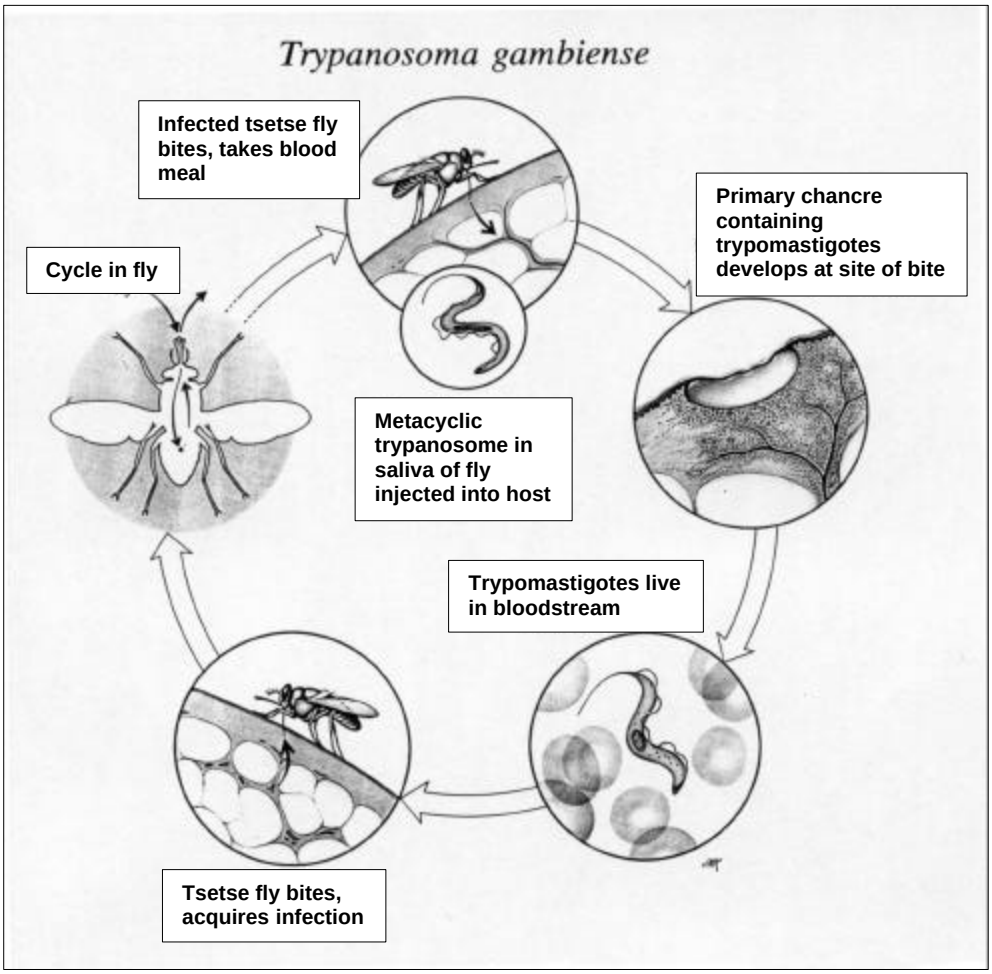


Figure 3: Parasitological life cycle of African trypanosomes.(After Karapelou JW; copied in Katz, Despommier and Gwadz, 89)

Biology of trypanosomes:

Like most members of the Family Trypanosomatidae, *T. brucei* is heteroxenous. During one stage it multiplies in the blood, lymph and central nervous system of its vertebrate host and during the other stage, it passes through a bloodfeeding insect vector (**Fig. 3**) [Schmidt, 85]. During the whole life cycle, *T. brucei* undergoes several morphological and metabolic changes which are well documented. The cyclic transmission begins with the bloodmeal of an infected tsetse fly which inoculates infective trypanosomes with the saliva into the dermal connective tissue. Trypanosomes start to multiply by binary longitudinal fission and an inflammatory chancre develops. This passage is also linked to the inactivation of the mitochondrion and the energy transformation now relies only upon glycolysis. Slender trypomastigote forms (see cell structure) then invade the bloodstream and lymph and continue to multiply. Biochemical analysis showed that at this stage, the whole surface of the cells is covered with a homogeneous coat of tightly juxtaposed glycoproteins [Barry, 97; Pays, 98]. This coat thwarts nonspecific immunity and denies access of antibodies to invariable antigens of the parasite like transporters or receptors. Since the coat itself is very immunogenic, the host immune system reacts during this phase and brings the exploding trypanosome population under control. However, the surface coat can be switched and some parasites displaying yet another surface evade the immune system and start to multiply giving rise to a new parasitaemia peak. This will drive a new immunoglobulin to be expressed by the host, a new switch to a surface coat that was not experienced by the host yet and so on. It was estimated that approximately 1000 genes are coding for the so called variant surface glycoprotein (VSG) [Van der Ploeg, 82] and the parasite, taking advantage of this VSG switching mechanism, will always be one step in front of the immune system. In addition, each peak appears to be composed of several variable antigen types (VAT) and switching rate is very high (10^{-2} switches/cell/generation) [reviewed in Hide, Mottram, Coombs and Holmes, 1997]. Early observations of parasitaemia in patient already showed the wave like development of trypanosomes in the blood (**Fig. 4**).

In natural infections, *Trypanosoma brucei* tends to be pleomorphic. The two main morphologically distinct forms seen in bloodstream are the long and rapidly dividing slender form with a long free flagellum which is replaced by short non-dividing stumpy forms with no free flagellum as cell density becomes high and when peaks decline as a

result of immune response (**Fig. 5**). As infection goes on, the parasite gains tissue fluids and finally invades the central nervous system. The patient will suffer from generalized pain, cramps, headache and weakness. Later, coordination problems, apathy, disinclination to do anything, somnambulism or sudden changes in behavior will soon be followed by coma and death. These neurological symptoms are rare in *T. b. rhodesiense* infections because the host usually dies before they can develop.

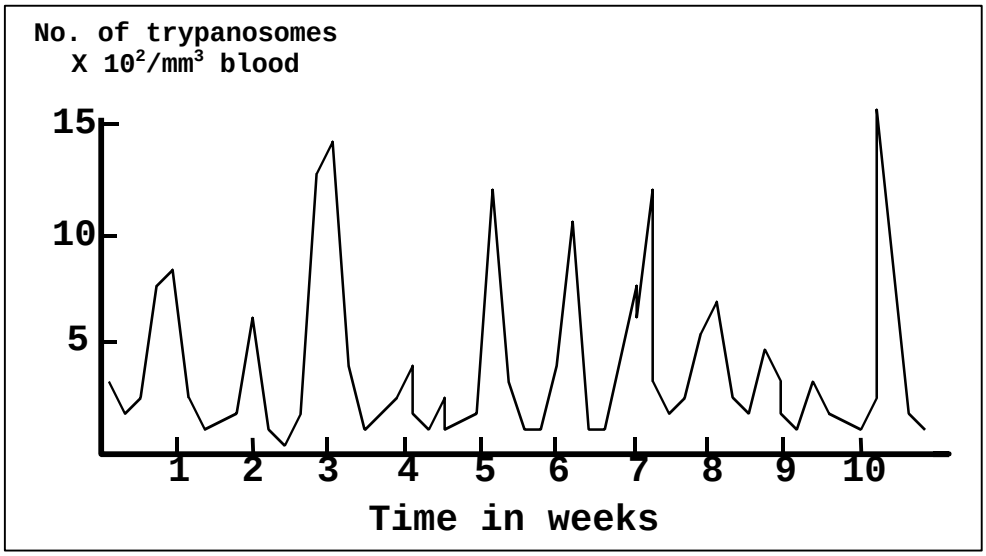


Figure 4: Course of parasitemia in a patient with *T.b.rhodesiense*. Each peak represents the development of a new B-VATs set.
(After Vickerman K., as redrawn from Ross and Thomson, Proc R Soc London [Biol] Sci 82: 411-415, 1910. Copied from Katz, Despommier and Gwadz, 1989)

For the cycle to continue, intervention of a tsetse fly which is going to ingest infected blood is necessary [Schmidt, 85]. As it can be expected, for the preparation to pass again into the invertebrate host in which the environment is much poorer in oxygen and glucose, the stumpy forms show a certain reactivation and development of the mitochondrion, but no cytochrome are present yet. Once they are taken up with the bloodmeal, the stumpy trypanosomes reactivate readily the oxydative pathway in the mitochondrion and will first multiply for several days as trypomastigotes called the procyclic forms in the midgut of the fly (**Fig. 5**). These slender cells will then migrate forward through foregut, esophagus, pharynx and finally reach the salivary glands and transform into epimastigote forms (see cell structure) which, after several generations, transform into the metacyclic trypomastigote. This is a small stumpy cell lacking a free flagellum and is the only infective stage to the vertebrate host (**Fig. 5**).

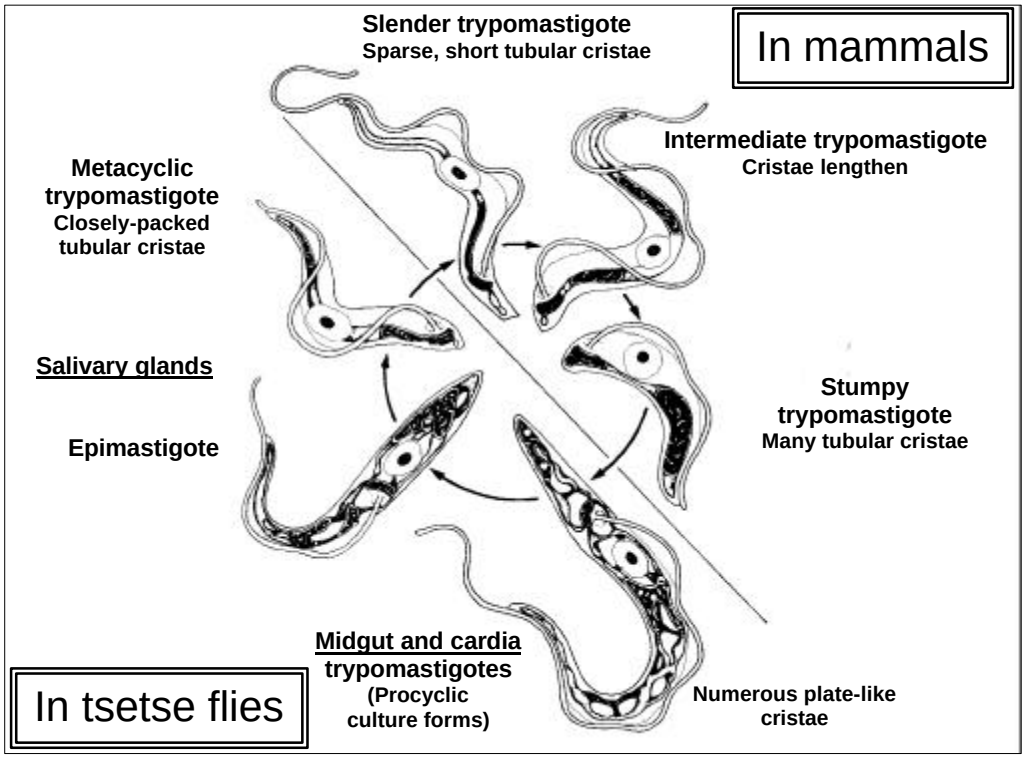


Figure 5: Diagram to show morphologic changes throughout the life cycle of *Trypanosoma brucei*. Note changes of mitochondrion, position and length of flagellum and general shape of the cells. (After Vickermann, K. 1971; copied from Schmidt and Roberts, 1985).

At the entry into the vector, the cells loose rapidly their VSG coat which is replaced by another surface protein called procyclin or PARP (procyclic acidic repetitive protein) [Roditi, 89] protecting against the proteolytic activity of the flies gut. After multiplication and end localization in the salivary glands, the metacyclic cells get a new VSG surface coat pre-adapting them to their transmission into the next vertebrate host. These VSGs, often called M-VSGs differ from the VSGs that are expressed by bloodstream forms (B-VSGs) mainly by their different genetic environments, regulatory mechanisms and expression sites (ES). The population of metacyclic forms injected into the new host will be composed of a mixture of M-VATS (variable antigenic types), probably in order to prevent elimination by the immune system especially in the case of previous infections of the new host. The expression of M-VSGs appear to be predictable and it seems that only a specific subset of VSG-coding genes can be used at this stage. This phenomenon first gave hope for a “cocktail” vaccine development but further studies showed that there was also a certain turnover of this subset of genes [Barry, 83].

The mechanisms involved in regulation of the hierarchical and very precisely tuned switching events between procyclin and VSGs [Roditi, 96] as well as those leading to single VSG expression at a time in blood are progressively becoming unraveled [reviewed by Barry, 97].

Cell structure:

Trypanosomes are elongated unicellular organisms well adapted to their environment [Schmidt, 85]. The flagellum arises from the kinetosome and is attached along the cell body by an undulating membrane until it becomes free at the anterior end (**Fig. 6**). When well developed, this structure propels the trypanosome in the fluid.

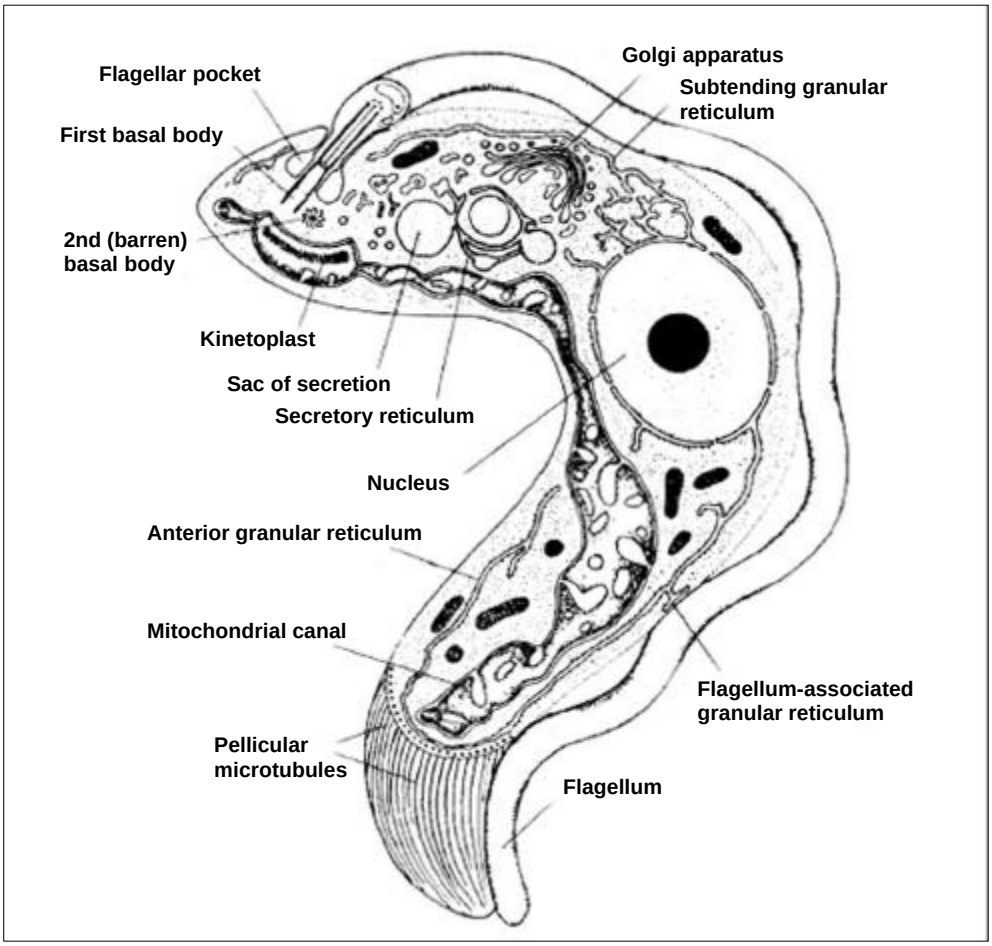


Figure 6: Sagittal section showing principal structures revealed by electron microscope in the bloodstream trypomastigote form of an African trypanosome. (After Vickermann K., copied from Schmidt and Roberts, 85).

The kinetoplast containing the mitochondrial DNA (kDNA) is always in close association with the kinetosome and the unique mitochondria arises from it. This structure is unique to trypanosomes and its location, associating the kinetosome, varies upon life stages. In African trypanosomes, only the trypomastigote form displaying a very posterior location of flagellum origin/kinetoplast and the epimastigote form in which the structure is pushed in front of the nucleus are present. Other kinetoplastids like *Trypanosoma cruzi* or *Leishmania sp.* will show other or additional forms during their life cycles in which this structure lies at the anterior end with a free flagellum along its whole length (promastigote form) or the cells are going to become rounded with only a tiny and quite internalized flagellum (amastigote form). This latter form is seen when the parasite displays an intracellular lifestyle.

The mitochondrion is much larger and shows well developed lamellar cristae in the insect stage of the parasite while in bloodstream forms, these structures regress [Schmidt, 85]. This is very well linked to the metabolic changes that occur by passing from one host to the other (**Fig. 5**). At the base of the flagellum, the flagellar pocket surrounds the kinetosome and is responsible for nutrition. Finally, the last remarkable feature of trypanosomes if compared to other more classic eucaryotic cells is a pellicular microtubule layer just beneath the cell membrane that gives resistance to deformation of the cell body [Kohl, 98].

Nucleus and genetic material:

The nucleus of trypanosomes does not reveal special features at first glance. At interphase, it is approximately 3µm in diameter and the nucleoplasm is dominated by a single large nucleolus of 0,5-1µm which persists upon mitosis and elongates concomitantly with nuclear division [Ersfeld, 99]. However, it was observed that no chromosomes are seen at any stage of cell division and the nuclear membrane persists upon the whole mitotic process (**Fig. 7**). Dense chromatin appears to be peripheral and attached to the nuclear membrane. Translucent areas surround it, but it is not known if this represents hetero- and euchromatin nor if there is any functional difference between these two chromatin states in terms of genetic activity [Ersfeld, 99]. In bloodstream forms which have a lot of peripheral chromatin, it takes the form of clumps which at least in the earlier stages of division remain adherent to the nuclear envelope, whereas in procyclic forms, these masses are

more diffuse and line the inside of the membrane. It was proposed that the nuclear membrane could serve as a carrier for genetic material to be separated into the daughter cells, like it happens in dinoflagellates or procaryotes [Vickerman, 70]. Fine structural analysis revealed the presence of some kind of microtubular spindle as well as structures appearing as dense plaques which are going to arrange on an equatorial location and are linked to the poles of the nucleus by microtubules. These plaques subsequently are going to split into two halves when nucleus elongation starts and continuous microtubules form between them. The number of plaques was proposed to correspond to the number of chromosomes and to play a kinetochore-like function, but chromatin remains dispersed and chromosomes discrete throughout division [De Souza, 74; Solari, 80]. For review, see [Ersfeld, 99 and references therein].

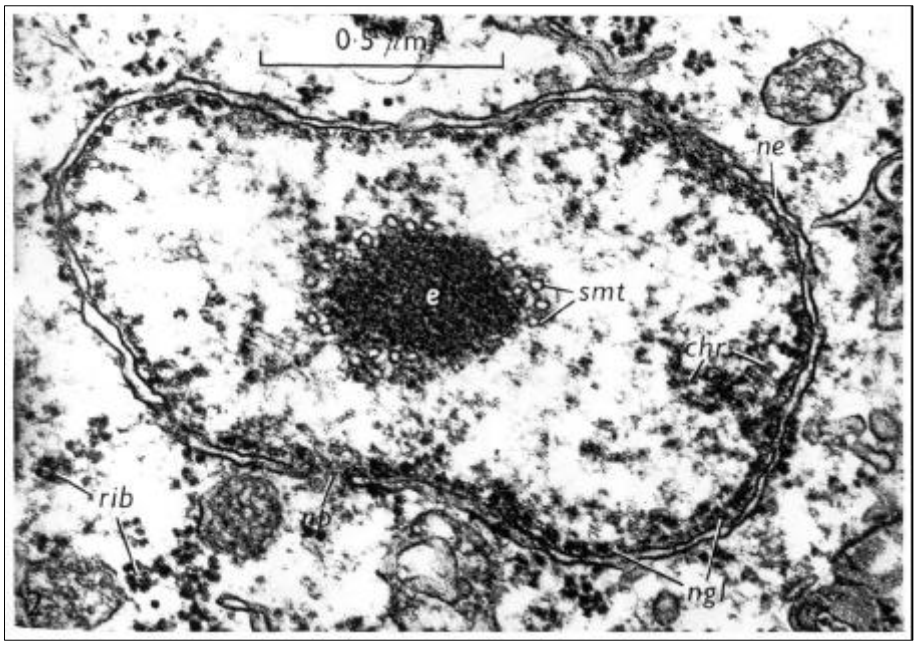


Figure 7: Transverse section of dividing nucleus with microtubular elements (smt) surrounding endosomal material (e). Nuclear envelope (ne), pores (np). Loosely clumped granules which may represent chromatin (chr) and a layer of such granules is lying beneath the persistant nuclear envelope throughout division (ngl), ribosomes (rib). (After Vickerman K and Preston TM, J Cell Sci. 6, 365-383 (1970)

Another intriguing feature of the trypanosome biology is the poorly understood sexual mechanism. The finding of non parental phenotypes after mixed infection with two different marked clones of trypanosomes and full development in the tsetse fly vector shows that gene exchanges occur [Jenni, 86]. The found hybrids cannot result from some spontaneous rearrangements since either the parental or the hybrids were shown to be

stable in terms of phenotype and genotype after individual tsetse fly transmission. The involved mechanisms as well as the location in the vector where these mating events take place are still unresolved but would well fit some kind of mendelian meiosis followed by fusion of a haploid stage [Jenni, 86]. It was also shown that mating is not obligatory, since some emerging trypanosomes were of parental genotypes and phenotypes. More recently, using several markers differing between several trypanosome stocks, it was shown that self fertilization occurs in trypanosomes, but also that the presence of dissimilar clones is needed to initiate it since no such recombinants were found when a single clone was used to infect tsetse flies [Tait, 96].

The genetic material of trypanosomes is divided into two compartments, the nucleus and the kinetoplast (mitochondrion). Replication and segregation to daughter cells of both genomes are coordinated, the kinetoplast dividing first.

The nuclear genome consists of approximately 3,5 x 10^7 bp per haploid genome and the chromosomes are divided into three classes according to their mobility in PFGE [reviewed in Ersfeld, 99].

About 100 linear minichromosomes of 50-100 kb are found and 90% of their sequence consists in repeated stretches of unknown function. Many of these minichromosomes contain a silent copy of a VSG gene near one or both telomeric ends. The telomeric regions contain conserved regions also found on larger chromosomes and therefore these regions are susceptible to recombination events enabling VSG genes to be translocated into an expression site (ES). They were shown to be inherited stably and no other genes were reported for minichromosomes.

Little is known about the function of intermediate chromosomes whose size ranges between 200 and 900 kb. They are known to contain VSG expression sites. They differ from the other two classes by the absence of the repeated sequence motifs of 177 bp which are typical for the two other classes and can represent as far as 90% of minichromosome sequence. Their size can vary considerably, probably by recombination events. No specific gene marker hybridizes exclusively with these chromosomes and it is possible that they serve as reservoirs of ESs for VSGs. Their ploidy is not evident and the absence of specific markers makes this difficult to assess.

Finally, megabase chromosomes of 1-6 Mbp were difficult to numerate because of their size divergence between homologue chromosomes which can also show different migration patterns between different trypanosome stocks. To this date, 11 chromosomes

per haploid genome were highlighted in *T. brucei* (numbered I to XI from the smallest to the largest) by the use of 500 specific cDNA probes. Such a plasticity in chromosome size was also observed in *T. cruzi* [Henriksson, 96]. Megabase chromosomes carry most active genes and some of these are arranged in tandem repeats which may contribute to size polymorphism. All telomeric regions harbour a VSG expression site.

CHROMATIN STRUCTURE AND FUNCTION

General overview:

In higher eucaryotes, DNA is organized through several levels of compaction which lead to chromosome formation upon cell division. The first step of chromatin higher order structure is achieved by spooling the DNA around a set of basic proteins called histones. DNA is first wrapped around an octamer of core histones (H2A, H2B, H3 and H4) which is formed by a heterotetramer of H3-H4 to which are added two H2A-H2B dimers. The structure is then stabilized by the fixation of the linker histone H1 that interacts with the DNA at the entry and exit of the formed nucleosome particle as well as with the linker DNA that runs towards the next nucleosome (**Fig. 8**).

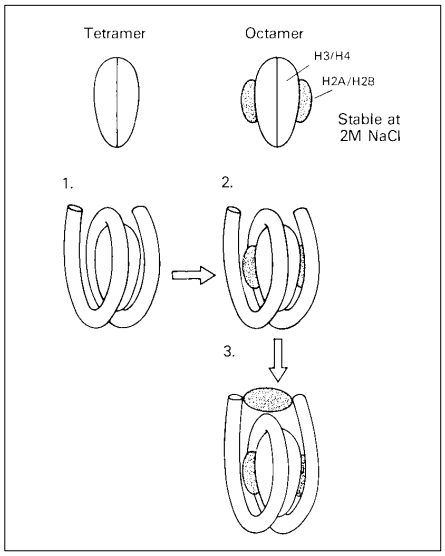


Figure 8: Schematic model of nucleosome assembly. H3 and H4 heterotetramer assemble first (1) and two H2A/H2B dimers join the structure and stabilize approximately 146 bp of DNA (2). H1 closes the two turns of DNA and protects an additional 20 bp of DNA (3).
Note that H1 is placed symmetrically on the axis of the core particle and contacts the coil of DNA. Recent studies tend to show an asymmetric localization of the globular domain of H1. (After Wolffe A, 1992)

Structurally, histone H1 is also controlling the regular spacing of the nucleosomes along the DNA and the formation of the typical 30nm chromatin fiber called solenoid. Experimentally, soluble chromatin can be prepared at different salt concentration leading to the stepwise condensation pattern up to the 30 nm fiber [Thomas, 84; van Holde, 89] (**Fig. 9**).

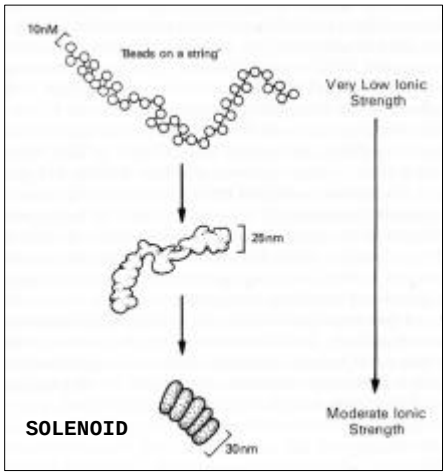


Figure 9: Schematic representation of chromatin condensation. Experimental preparation for electron microscopy under low salt concentration appears as a regular zigzag structure often compared to beads on a string. Under increased salt concentration, the structure folds and forms the typical 30nm fiber called solenoid.
(After Wolffe A, 1992)

Experimentally, different structures can be obtained in terms of constituents and length of DNA that remains associated upon micrococcal nuclease digestion. The so called nucleosome core particle and the chromatosome have long been considered as the fundamental repeating components of chromatin (**Fig. 10a** and **b**). The core particle obtained after extended digestion, consists in the association of core histones with 146 bp of DNA while the chromatosome which protects a further 20 bp of DNA with the addition of the linker histone is obtained by shorter digestion periods. However, the length of the DNA wrapped around the histone octamer now appears as a highly variable and dynamic feature. Therefore, in a functional view, it was proposed that the nucleosome (**Fig. 10c**) in which the DNA stretches entering and exiting the structure are present should now be regarded as the fundamental repeating unit of chromatin structure [reviewed in van Holde, 99].

All core histones are remarkably conserved in length and amino acid composition through evolution. Histone H3 and H4 are the most conserved; for example, H4 of calf and pea differ only by 2 residues out of 102. H2A and H2B are slightly less conserved and H1 appears to be more subject to substitutions. This reflects the organizational sequence of chromatin folding with the most central histone being under higher evolutionary pressure

in terms of function and structure. Core histones are small basic proteins of 11-16 kDa which all have a similar secondary structure consisting of a globular domain which plays a central role in nucleosome formation by histone-histone and histone-DNA interactions and charged N-terminal tails which contain high proportions of lysine and arginine. Histone H1 is slightly larger than core histones (20 kDa) and also differs from the others by the presence of highly charged tails at both the amino and carboxyl terminus. They are highly basic, being particularly rich in lysine. H1 is the least tightly bound histone and dissociates first from chromatin, followed by H2A/H2B and finally H3/H4 when preparing chromatin fibers at decreasing ionic strength [reviewed in Wolffe, 92].

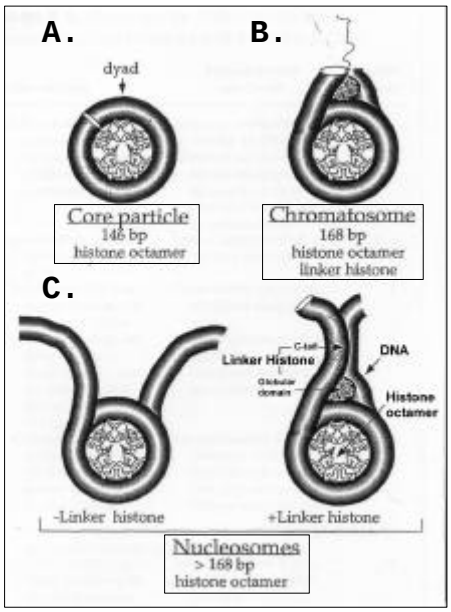


Figure 10: Schematic representation of the major structural elements of the chromatin fiber: core particle (A), chromatosome (B) and nucleosome containing or lacking the linker histone (C). The globular domain of linker histone is depicted as an oval and its C-terminal tail as an irregular curve. (After van Holde K and Zlatanova J, 1999)

Histones have long been regarded strictly as structural elements which participate in the folding of DNA and until recently, nucleosome formation was viewed as a static chromatin building block preventing access to the DNA template for any factor implicated in DNA processes. In such a model, remodeling of gene activity would largely depend on competition between histones and *trans*-acting factors upon replication, a system which clearly appears insufficient. With many studies which showed not only that nucleosomes were influencing gene regulation globally, but also that histones and chromosomal proteins displayed much subtle and even specific functions in gene activity, evidence accumulated that the importance of chromatin structure had to be reconsidered [Felsenfeld, 92; Grunstein, 92; Wolffe, 94].

Fine analysis of histone-histone as well as histone-DNA interaction revealed a very complex anatomy of the nucleosome. The interfaces between histones are very specific but also allow flexibility and conformational changes. It appears that any modification at any location of one constituent of the nucleosome will potentially affect the whole structure and will have important effects on chromatin folding. This, in turn, will influence regulatory mechanisms or replication by means of *trans*-acting factors access to DNA [Pruss, 95]. Furthermore, chromatin is not a stable, static structure. Proteins, including the histones, continually equilibrate in and out and modifications of its constituents and interactions with other structural and non-structural proteins lead to a wide range of reversible changes in chromatin.

It is known that all histones can undergo posttranslational modifications (e.g. acetylation or phosphorylation) which have an effect on nucleosome-DNA stability and weaken constraint of linker histone. In addition, expression of different alleles during development and/or cell cycle have been reported. Histone genes are present in multiple copies which reiteration varies from two copies to several hundreds. In some organisms, different forms of histones are transcribed at precise periods in development suggesting that DNA can be compacted in a wide variety of ways [Wolffe, 92]. All these variations have important consequences for chromatin structure and strongly suggest that different chromatin fibers have distinct functional roles during embryogenesis, in somatic cells or between divergent evolutionary levels [reviewed in Pruss, 95].

The way by which chromatin folds is still a question of debate and many models have been proposed which are sometimes contradictory. However, at the time, a huge number of data become convergent and show that chromatin structure is intimately linked to regulation processes [van Holde, 96b].

Linker histone appears to be necessary for the proper folding of the chromatin fiber and even if experiments showed that condensation also happens without it, the obtained structure is quite different. The C-terminal tail of H1 make the DNA entering and exiting the nucleosome pulled together, influencing the steric arrangement of nucleosome chains. When only the globular domain of H1 is associated or in absence of H1, the structure is much wider (**Fig. 10c**). On the other hand, the N-terminal tails of core histones do not seem to be implicated in nucleosome formation but they protrude out of it. Their interactions with linker DNA is well established and contact with H1 has also been suggested. It was shown that in addition to the central domain of linker histone, three-

dimensional fiber folding needed either the tails of linker histone or the N-terminal tail of H3 [van Holde, 96a].
On a functional level, access to the genetic information for the transcriptional machinery clearly needs at least a certain unfolding of the chromatin fiber and this may be accomplished by linker histone depletion and/or modifications of histone tails. These two events may well be linked since core histone tails seem to interact with linker histone tails and acetylation of these first could provoke the destabilization or ejection of the second [van Holde, 96a]. It was also reported that transcriptionally active regions of chromatin show a reduced H1 content [van Holde, 96a]. Furthermore, it becomes evident that many transcription factors or complexes resemble histone structures or even whole nucleosome configurations. This could be interpreted in terms of facilitation for the integration of functional proteins into the chromatin environment [van Holde, 96b].

DNA becomes folded about 7 fold when associated with nucleosomes and another 7 fold by the formation of the 30nm solenoid. Further steps of metaphase chromosome formation is still less well understood, but an additional 250 fold compaction takes place. Again, researchers have proposed different models. One consisting of large loops of the 30nm fiber anchored on a discrete central axis formed by non-histone proteins that continually assemble and disassemble. Another one proposes further hierarchy of higher order chromatin folding patterns for the 30nm fiber, leading stepwise to 50nm, 100nm and 130nm structures. However, the chromatin fiber is organized into large domains potentially separated through interactions with the nuclear scaffold or matrix. The association of lamins, which normally form a protein meshwork on the inside of the nuclear envelope, with the chromosome scaffold fraction led to the idea that lamina anchors interphase chromatin to the nuclear envelope and thereby influences higher order chromosome structures [reviewed in Wolffe, 92].

Some other scaffold proteins, especially topoisomerase II have been shown to be associated with chromatin. Topoisomerase II allows double strand breaks in DNA that are unravelling knots generated during replication and transcription. If topoisomerase is inactivated *in vivo*, the cell dies because it cannot separate the chromosomes at the end of mitosis. Antibodies to topoisomerase II locate it along the axial region of chromosomes. Although, topoisomerase II does not seem to have an essential structural role as a building block of chromosomes. However, *in vitro* nuclear assembly systems indicate that

chromosome condensation is closely correlated with the level of this protein. H1 and topoisomerase II are believed to preferentially bind to AT-rich DNA sequences that in turn have been suggested to interact with specific nuclear scaffold attachment regions [Adachi, 89; Izaurralde, 89; Ivanchenko, 92]. However, even if the presence of residual DNA fragments in matrix or scaffold preparation remain controversial, substantial evidence exists to suggest that at some time in the cell cycle, all DNA will have some attachment to the nuclear matrix.

Chromatin structures can also be modified by selective association of abundant non-histone proteins that interact with DNA-histone complexes. The fractionation of histone H1 by perchloric acid extraction also solubilizes several other proteins, e.g. Low and High Mobility Group proteins (LMG and HMG). HMG 14 and 17 show preferential binding to nucleosomal DNA where it enters and exits the nucleosome, like H1 does. By competition with H1 or by acetylation and phosphorylation of the core histones, the interaction at this site is likely to modify higher order structure of chromatin. In contrast, HMG 1 and 2 do prefer naked DNA. These proteins could cause a local destabilization of nucleosome structure, perhaps by competition with the histones for the interaction with DNA [Wolffe, 92; Zlatanova, 96].

Status of histone H1:

Histone H1 is one of the most abundant proteins in the nucleus. However, its exact positioning on the nucleosome as well as the question whether it has primarily a structural function or rather gene regulation implications or both is still controversial.

The exact positioning of H1 in the nucleosome is not well understood. H1 appears to be able to bind DNA by two binding surfaces [Pruss, 95; Widom, 98]. When H1 is removed from the nucleosome the linker DNA strands entering and exiting the nucleosome appear separated while in presence of H1, only one region of entry/exit is seen. In addition, the globular domain of H1 alone can provide protection of the 168 bp of DNA in the chromatosome [Pruss, 95], showing that the tails are not implicated. All this suggested that H1 is located symmetrically on the axis of the histone octamer [van Holde, 99].

However, recent studies have shown that the globular domain of the linker histones would rather be located asymmetrically, and thereby change greatly the models of how H1 can be implicated structurally and functionally in chromatin [Widom, 98; van Holde, 99].

Depending on the experimental procedure involving either natural chromatosomes or fully reconstituted structures, different locations were proposed, involving either both DNA binding surfaces or only one. In the latter case, the binding surface remaining free is proposed to be able to join adjacent nucleosomes and thereby to give a possible polarity to the chromatin fiber [Widom, 98].
However, when analyzing lower eucaryotes, differences appear in histone properties and structure. Most differences involve histone H1 which is known to be less conserved and which differs from the H1 of higher eucaryotes not only by some substitutions but also by appearing truncated. In the ciliated protozoan *Tetrahymena sp*., H1 lacks the globular domain entirely and in this organism, compaction models requiring this domain cannot be applicable. Stabilization and compaction of the nucleosomal array would then simply rely on electrostatic interaction between this histone and the phosphodiester backbone of linker DNA [Wolffe, 97].

The functional aspect of linker histone was also to be reexamined since several recent studies showed that H1 is not essential. H1 gene knockout experiments in the ciliate *Tetrahymena thermophila* have shown that H1 is dispensable for cell survival and growth even if the nuclear size appears increased about twofold [Shen, 95]. In *Xenopus laevis*, H1 depletion still allows chromosome condensation similar to extracts containing the linker histone [Ohsumi, 93; Dasso, 94]. On the other hand, the great number of developmentally regulated and/or tissue specific variants of H1 described to this date [Wolffe, 92], their differential expression during the cell-cycles and/or life-cycles of several protozoan [Burri, 94; Sabaj, 97; Noll, 97], the ability to influence positively or negatively specific gene transcription control rather than have a global effect on transcription [Shen, 96], the variety of timed post-translational modifications they can undergo [van Holde, 89] and the nonrandom distribution of H1 variants within the nucleus [Schulze, 93; Trieshmann, 97] are all indications that histone H1 plays an important role in the nucleus [reviewed in Wolffe, 97].

Chromatin in trypanosomes:

While histones and nucleosome structures are absent in procaryotes, they appear to have evolved in order to enable DNA compaction in most eucaryotic cells. However, in

dinoflagellates, histones seem to be absent and in some protozoan, the set of usual eucaryotic histones appears incomplete [van Holde, 89].

Thin sections of nuclei observed by electron microscopy clearly show that only faint chromatin condensation takes place in *T. b. brucei* if compared to rat liver cells [Hecker, 85]. However, despite this absence of chromatin compaction, trypanosomatids do exhibit the full set of histones, but their amino acid composition, their biochemical properties and primary structures often differ from those of higher eucaryotes [reviewed in Galanti, 98]. The general features that differentiate trypanosomatids from higher eucaryotes on the level of histone genes can be summarized as follows [Galanti, 98]:

A) The genes coding for each histone appear to be spread over different chromosomes.

B) Their transcripts are polyadenylated and the sequences which are usually very strongly conserved through evolution show substantial differences, especially for the N-terminal tails.

C) Histones H1 are of astonishing small size and do correspond only to the C-terminal region of higher eucaryote H1 histones.

During the course of detailed investigations on chromatin structure in trypanosomes, several peculiarities concerning stability, protein interactions as well as biochemical properties of their histones were highlighted.

The first studies on chromatin compaction patterns showed a very faint condensation in trypanosomes if compared to higher eucaryotes under standard conditions [Hecker, 85]. Differences are also observed when comparing African and American trypanosomes, *T. cruzi* showing a higher degree of compaction than *T. brucei* but a typical 30nm fiber (solenoid) was never seen. In addition, reconstitution of trypanosome chromatin with rat liver histone H1 increased the compaction of *T. cruzi* chromatin but had no influence on *T. brucei* chromatin. However, it was also recognized that the DNA-protein interactions in the nucleosome are less stable in *T. brucei* and can easily be destabilized by standard experimental procedures developed for the isolation of chromatin from higher eucaryotes [Hecker, 89].

SDS-polyacrylamide gel electrophoresis revealed no H1 comparable proteins in both trypanosomes. Comparison of the core histone band patterns of both species as well as of rat liver histones clearly showed electrophoretic mobility differences (**Fig. 11**). Differences can also be seen in the number of core histone bands and their mobility between the two life stages of the parasite [Schlimme, 93].

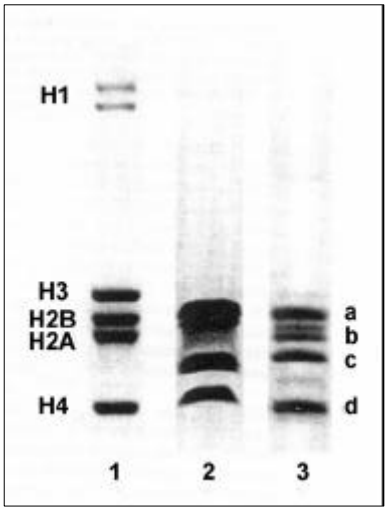


Figure 11: Histones of calf thymus (1), *T.cruzi* (2) and *T.b.brucei* (3) separated on SDS-Tricine PAGE according to molecular weight. Bands a, b, c and d are respectively the counterparts of H3, H2A, H2B and H4 in higher eucaryotes. Note the absence of bands in the H1 region.
(After Hecker H, 1994)

However, micrococcal nuclease digestion showed that chromatin of trypanosomes is composed of nucleosome filaments which lead to DNA ladders similar to those of rat liver chromatin even if the chromatin of trypanosomes is digested more rapidly (**Fig. 12**). Differences in digestion can also be seen between the two life stages of the parasite, chromatin of bloodstream forms being more resistant than in procyclic form [Schlimme, 93]. It was argued that the differences in core histones, the absence of a typical H1 and the less condensed chromatin leave the linker DNA more exposed and accessible to the nuclease.

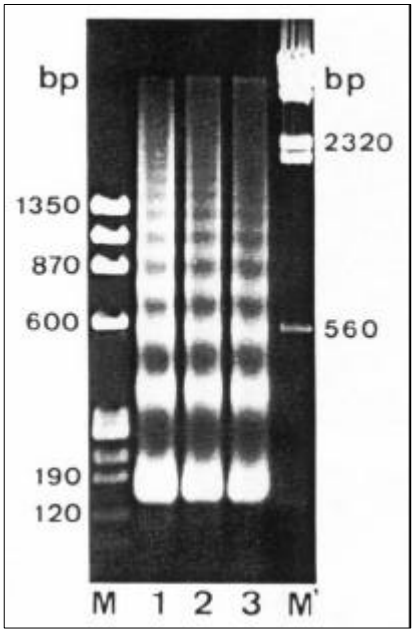


Figure 12: Micrococcal nuclease digestion pattern of nuclear chromatin of *T.b.brucei* and rat liver DNA under adjusted conditions. M: *Hae*III fragments of ϕX174RF DNA; M ': *Hind*III fragments of λ DNA. Lane 1: 20 A_{260} of rat liver chromatin digested with 0,4 U enzyme at 37°C for 10 min; lane 3: 20 A_{260} of *T.b.brucei* chromatin digested with 0,2 U enzyme at 30°C for 4 min; lane 2: mixed preparation of 1 and 3.
(Reproduced from Hecker H, 1989)

Despite these differences, gel electrophoresis, amino acid content and partial amino acid sequences allowed to assign the different core histones to their higher eucaryote counterparts [Bender, 91].

By analyzing the disruption sequence of the core histones in increasing salt concentration, it was observed that H3 and H4 were already released from DNA at 1M NaCl while a higher ionic strength of 1,2-2M NaCl is needed to dissociate them in higher eucaryotes [Bender, 92b]. In addition, digestion with immobilized trypsin revealed similarities with higher eucaryote core particle digestion but also clear differences particularly for H4 which is degraded rapidly in *T. brucei*, indicating that it is located on the outside of the core particle. This contrasts greatly to the strong cleavage resistance to immobilized trypsin and the internal location of higher eucaryote H4 [Bender, 92c].

Sequencing of core histone fragments of *T. brucei* demonstrated that the usual extreme conservation of histones through evolution was not applicable in lower eucaryotes. Differences of 21 to 48% between *T. brucei* and higher eucaryote histones as well as substitutions with amino acids with different properties indicated that the nature of histone-histone and/or histone-DNA interactions is different and strongly suggested an alternative method of organization and processing of the genetic information in trypanosomes [Bender, 92a]. Therefore, the divergences in histone properties between higher eucaryotes which are potential hosts and trypanosomes which are both pathogens led to the idea that new specific targets for the action of trypanocidal drugs might be highlighted [Hecker, 85]. In addition, sequence differences between African and American as well as among several other protozoan suggested that the knowledge of whole histone sequences might prove a useful tool for phylogenetic investigations of lower eucaryotes [Hecker, 93].

Meanwhile, evidence accumulated that a histone H1 protein, presenting small size and high electrophoretic mobility, was also present. Finally, H1-like proteins showing similar biochemical properties than histones of *Tetrahymana fasciculata* [Johmann, 76] were described in *T. cruzi*, *Leishmania major* and *Crithidia fasciculata* [Toro, 88, 90, 93; Fasel, 94; Duschak, 90; Espinoza, 96]. These proteins were either lost or hidden among the core histones by the use of experimental procedures, explaining their late discovery. The use of Triton acid-urea gels which separate proteins according to their hydrophobicity allowed to separate *T. brucei* H1 and revealed a fast migrating group of bands in front of the core histones [Schlimme, 93]. Differences in the number of bands, the position in the gel as well as in the relative amount were seen between the two stages of the parasite (**Fig. 13a**).

Differences between *T. brucei* and *T. cruzi* were also visible (**Fig. 13b**) [reviewed in Hecker, 94].

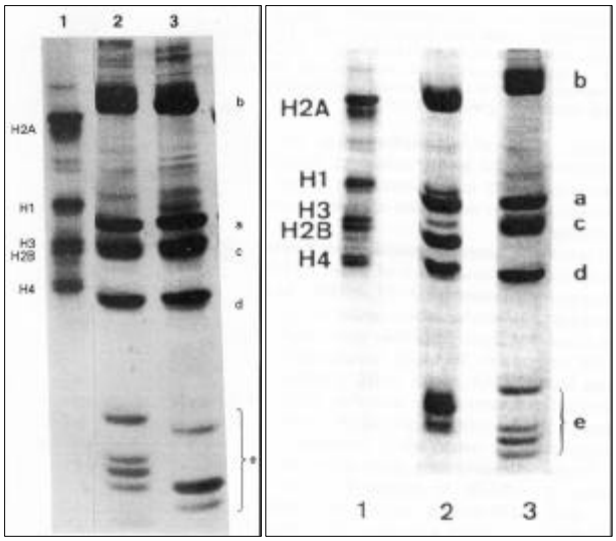


Figure 13: Histone separation on Triton acid-urea PAGE according to hydrophobicity. A: comparison of the two life cycle stages of *T.b.brucei*. Lane 1: calf thymus histones; lane 2: procyclic culture forms; lane 3: bloodstream forms. B: differences between American and African trypanosomes. Lane 1: calf thymus histones; lane 2: *T.cruzi*; lane 3: *T.b.brucei*. Histones a, b, c and d, as in Fig.11; e: histone H1 variants. (Reproduced from Schlimme W, 1993 and Hecker H, 1994)

By adaptating the procedure for the soluble chromatin preparation to the weaker stability found in trypanosomes, it was shown that studies preceeding the discovery of H1 were made with damaged chromatin depleted in H1. Indeed, when the preparation was carried out according to the adapted procedure, a regular spacing of nucleosomes could be seen even in procyclic forms in which chromatin is known to be the most unstable [Burri, 93]. Reinvestigation of salt-dependent condensation of chromatin using the adapted procedure showed that the condensation failed in previous studies because of the loss of H1 during chromatin preparation, however, the 30 nm solenoid structure typical for higher eucaryotes was not obtained [Burri, 93] (**Fig. 14**).

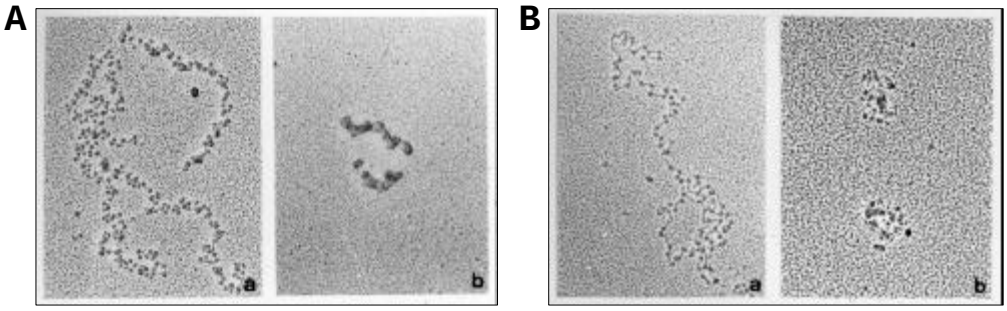


Figure 14: Chromatin of rat liver and *T.b.brucei* procyclic culture forms prepared for electron microscopy at different salt concentrations. A: Rat liver chromatin at 10 mM NaCl (a) and 100 mM NaCl (b). *T.b.brucei* chromatin at 10 mM NaCl (a) and 100 mM NaCl (b). (After Burri M, 1995)

The use of two-dimensional separation gels in which different protein properties are relevant, four variants and/or posttranslational modifications of H1 were found in both procyclic and bloodstream forms of *T. brucei* [Burri, 93; 94]. However, in bloodstream forms, one variant and/or posttranslational modification appeared to be over-expressed relatively to the other three (**Fig. 15**), while in procyclic forms, all H1 are expressed to a similar level [Burri, 94].

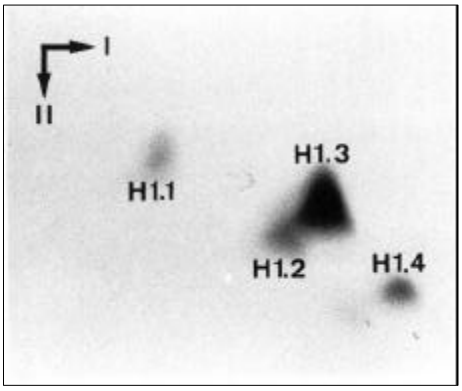


Figure 15: Two dimensional gel electrophoresis of FPLC separated histone H1 of *T.b.brucei* bloodstream forms. I: Triton acid-urea gel separation according to hydrophilic / hydrophobic properties; II: SDS-Tricine gel separation according to molecular weight. Four variants and/or post-translational modifications were obtained. Note the increased level of H1.3 if compared to the other.
(After Burri M, 1994)

Topographic analysis of H1 in the nucleosome filament by immobilized trypsin showed that in an open structure at low ionic strength, H1 was readily degraded by the enzyme whereas at increased ionic strength, no digestion occurred [Burri, 95]. This would be in line with the solenoid model in which H1 is internalized during higher order structure formation [Thoma, 79].

In addition, similarly to the situation encountered in *T. cruzi* [Åslund, 94] and in trypanosomatids in general, partial amino acid sequences showed that all *T. brucei* H1 fragments that were sequenced were related to the C-terminal region of higher eucaryote H1 [Burri, 95].

Astonishingly, reconstitution experiments revealed that dephosphorylated H1 proteins of *T. brucei* procyclic forms were able to trigger compaction almost to the level of 30 nm fibers. When heterologously reconstituted onto rat liver chromatin, purified H1 of procyclic forms brought condensation to a higher level than the one seen in native trypanosome chromatin. When dephosphorylated, structures resembling solenoids were obtained, indicating that not only the particular properties of H1 are implicated in the faint chromatin condensation found in *T. brucei*, but that the special features of core histones are also involved [Burri, 95]. The enhanced condensation obtained with dephosphorylated H1 is

probably due to the increase of the net positive charge leading to stronger interaction with linker DNA. Such kind of modifications were extensively studied and may have important effects on chromatin structure and gene regulation. However, contradictory observations were made which makes it difficult to draw a general effect of H1 phosphorylation. First, hyperphosphorylation of H1 was observed during mitosis, leading to the proposal that phosphorylation is causally linked to mitotic chromosome condensation [Bradbury, 74]. However, several other systems in which phosphorylation is uncoupled from mitosis were later described. It is now generally admitted that phosphorylation weakens chromatin structure and allows access to the transcriptional machinery [Roth, 92]. The observed effect of dephosphorylated H1 on trypanosome chromatin is also in line with this latter theory, considering that the used cells were in logarithmic phase in which intense transcription and replication occurs [Burri, 95].
Similarly, reconstitution experiments made in *T. cruzi* also revealed the ability of its small histone to trigger chromatin compaction [Schlimme, 95].

Finally, it can be said that clear differences exist in the global histone properties (band patterns in various gel systems, relative positioning in the nucleosome, histone-histone and histone-DNA interactions, primary structure) between higher eucaryotes and trypanosomes or lower eucaryotes in general. Furthermore, even by the use of experimental procedures especially adapted to the weaker stability of trypanosome chromatin, differences in compaction and stability persist between *T. brucei* procyclic culture forms and bloodstream forms as well as between *T. brucei* and *T. cruzi*. The salt-dependent condensation is more pronounced in metacyclic *T. cruzi* than in epimastigote forms of the same species followed by bloodstream forms of *T. brucei* and finally procyclic culture forms of this latter species in which chromatin barely condenses [Hecker, 95]. All this suggests that a very complex mechanism involving differential expression of histone variants or differential regulation of postsynthetic modifications or probably both of them are used by trypanosomes in organization and processing of their genetic information. The implication of histones in gene regulation also suggest that the divergent chromatin structure of trypanosomes, if compared to higher eucaryotes, could imply an alternative gene regulation mechanism which could be of parasitological relevance [reviewed in Hecker, 95].
Furthermore, beside these investigations on chromatin structure in trypanosomes, evidence accumulated that gene regulation in these organisms must involve mechanisms not directly linked to promoter oriented events. The polycistronic transcription mode appears to be

driven only by few promoters dispersed over the genome. Pre-mRNAs are going to be trans-spliced by the 5' addition of a 39 nucleotide long cap called spliced-leader (mini-exon) and a polyadenylated. Both events appear to be closely interdependent and to be controlled by the intergenic region (IR) found between the ORFs of a same transcription unit [Graham, 95; Vanhamme, 95]. The protein levels are regulated posttranscriptionally by controlling elongation, transcript processing or mRNA stability [Graham, 96]. Evidence for such mechanisms come from the observation that gene products coded by the same transcription unit can show great expression differences and can be differentially regulated during cell cycle and/or between life cycle stages [Gibson, 88; Bringaud, 93]. Beside this, even for VSG or procyclin genes for which promoters have been identified, the observed gene activity cannot be explained by transcription regulation since these promoters have been shown to remain active in both procyclic and bloodstream forms [Pays, 90], at least when placed out of their genomic context by transient translocation [Roditi, 96]. Indeed, all VSG expression sites are identical but only one is expressed at a time. It is proposed that the three-dimensional context or association with other nuclear structures may be involved [Biebinger, 96; Horn, 97; Ersfeld, 99].

In such a context the particular features of trypanosome chromatin gains new dimensions and a better understanding of the mechanisms by which the genetic information is organized may lead to the comprehension of cellular functions at more complex levels.

AIMS OF THE PRESENT THESIS

In 1994, a tandemly arranged gene family coding for heterogeneous histone H1 in *T.cruzi* has been published [Åslund, 94]. As expected, these genes were of astonishing small size and completely lacked the central globular domain. Expression analysis also showed that a fraction of H1 was synthesized constitutively outside the replication process and in differentiated resting trypomastigotes, no H1 production could be highlighted [Sabaj, 97].

However, to date, the question of the exact function of histone H1 in trypanosomes remains open. It is still unclear how this small histone, corresponding to the C-terminal region of higher eucaryote H1, is binding to chromatin and nothing is known about its implication in any other functions in this organism. In addition, the genes coding for H1 in African trypanosomes, one of the most studied organism, have never been analyzed. In

order to further investigate the function of histone H1 in African trypanosomes and to target new experimental approaches, gene sequences and organization have to be known.

The specific objectives were therefore defined as follows:

1) cDNA isolation and sequencing of histone H1 coding genes.

2) Work on different *T. brucei* subspecies and strains in order to gain a wider representation of how this gene family has evolved.

3) Comparison of H1 gene sequences of different trypanosomatid species in order to verify the usefulness of H1 in phylogenetic investigations.

4) Definition of the gene localization at the chromosome level as well as their ORF organization.

5) Comparaison of H1 transcription levels between the two stages of the parasitic life cycle in order to clarify the observed over-expression of one variant in bloodstream forms.

6) Recombinant H1 protein production and purification in a procaryotic expression system.

7) Immunization in mice in order to obtain specific anti-H1 antibodies.

8) Functional analysis at the cellular level by immunofluorescence and confocal microscopy.

II. MATERIAL AND METHODS

Parasitic material and libraries:

The lambda gt11 cDNA library of *Trypanosoma brucei gambiense* (1257) prepared by Barnes [Barnes, 89] was a kind gift of T. Seebeck and an additional cDNA library of *Trypanosoma brucei brucei* (427) was provided by I. Roditi, both from the University of Bern, Microbiology.

T. b. gambiense (STIB755) and *T. b. brucei* (Treu 927/4) were kindly obtained from the Swiss Tropical Institute (Basel).

Procyclic forms were cultured in SM medium supplemented with 10% heat inactivated fetal calf serum [Cunningham 1977]. Cells were grown to a density of 10^7/ml. The cultures were centrifuged for 20 minutes at RT, 1800g. Elimination of the culture medium was carried out by two washes in PBS. The cell pellet was then immediately processed or frozen in liquid nitrogen and stored at -80°C.

Bloodstream forms were propagated in CD1 mice (Charles River, Germany) by intraperitoneal inoculation with 10^5 trypanosomes. After 7 days, blood was harvested by heart puncture and trypanosomes were separated from blood cells by anion exchange chromatography through DEAE cellulose [Lanham 1970]. Shipment was made on dry ice in PSG containing 10% glycerol. The cells were thawed just before use in a 37°C water bath followed by two washing steps (see above) in order to eliminate glycerol.

cDNA library screening:

The λgt11 cDNA grown in *E. coli* strain Y1088 was screened by standard plaque lift procedures (after Stratagene Protocols, 1993). Briefly, the libraries were amplified at 40°C for 6-8 hours on LB plates in *E. coli* Y1088 previously grown in LB medium supplemented with 0,2% maltose / 10mM $MgSO_4$. The titre of the library was determined by serial dilution procedure and 30'000 plaque forming units (pfu) were then plated onto three 150 mm LB plates. After growth, nylon membrane lifts were prepared by depositing them 2 min. onto the plate. Afterwards, the membranes were treated by deposing them 2 min. on a 3MM Whatman soaked with 0,4M Tris-Cl pH 7,6 / 1,5M NaCl for denaturation. The

membranes were then neutralized and rinsed by successive passage onto 3MM Whatman papers soaked respectively with 0,4M Tris-Cl pH 7,6 / 2 X SSC and 2XSSC for 2 min. Finally, the membranes were baked for 2 hours at 80°C.

Hybridization was performed with three biotinylated oligonucleotides (Table 1) corresponding to parts of previously sequenced peptide fragments [Burri 95] by the use of a chemiluminescent (Biotin-Streptavidine) (TROPIX®) detection kit. Washes were performed as recommended in the kit protocol, 2 X 5 min. in 2XSSC / 1%SDS; 2 X 15 min. in 1XSSC / 1%SDS at hybridization temperature and 2 X 5 min. in 1XSSC. Several strongly hybridizing clones were selected and purified by at least 3 rounds of plating/hybridization, until all plaques were positive. Since all 3 oligonucleotides were tested successively on each starting membrane, care was also taken to select clones that were differentially hybridizing with the different probes. The selected and purified clones were then amplified by PCR using standard forward and reverse lambda gt11 primers (Table 1) and subcloned into Bluescript plasmide SK- for sequencing.

Subcloning and sequencing:

cDNA inserts and all other PCR amplified fragments were subcloned into Bluescript SK- plasmid by the "Plasmid + T" method [Marchuk 1991], except the PCR products obtained with primer 5'-MAKTT/1 (see below) which were cloned into TOPO TA cloning® kit (Invitrogen) according to manufacturers instructions.

Plasmid + T was obtained by digesting 10 µg of Bluescript plasmid with EcoRV. Rather than to perform directly the T addition, the digestion product was loaded onto agarose gel and the band corresponding to digested plasmid cut out and purified with Qiaquick® columns (Qiagen) in order to limit traces of undigested palsmid and to improve blue-white selection. The digested DNA was then incubated for 3 hours at 70°C with dTTP and Taq polymerase followed by standard phenol/chloroform purification.

Ligation into Bluescript + T vector and transformation into *E. coli* was performed by standard heat-shock procedure and blue-white selection. Inserts were controlled either by standard miniprep procedure or by direct PCR with M13 universal and reverse primers in which the PCR mix was inoculated with bacteria of white colonies. Positive clones were then grown and the DNA purified by Nucleobond® AX columns (Macherey-Nagel) or QIAGEN Tip-20® (QIAGEN).

Sequencing was carried out on LICOR 4000 automated sequencer (MWG) using the Thermosequenase sequencing kit (Amersham) and both M13 universal or reverse dye primers (IRD800). Sequence reactions were set up with 500 ng of purified plasmid DNA and PCR amplification according to MWG manual.
Sequence analysis was performed with “DNA StriderTM 1.0.1” and alignments were made with ClustalW [Thompson, 94]. BLAST analysis were made in different Blast servers, mainly at NCBI, EMBL and the EBI Parasite Genome blast server (subsection: Kinetoplastid databases).

DNA and RNA manipulations:

Genomic DNA was extracted by the standard phenol procedure [Sambrook, 89] with a prolonged proteinase K digestion step overnight. Specific isolation of H1 genes was done using 4 primers encompassing the different ORFs. Primer 5'-MNNTT, 5'-MAKAS and 5’-MAKTT are located upwards the start codon and H1orf3’ is situated after the stop codon (Table 1). PCR amplification was performed with the following cycling conditions: 94°C x 2min, 30 times: 94°C x 30 sec, Tm-3°C x 30 sec, 72°C x 1 min and a final elongation step at 72°C x 5 min.
RNA was extracted using TRIZOL® Reagent (GIBCO BRL®, Life technologies) following the manufacturer's instructions.
Theoretical secondary structures, especially for the tandemly organized genes were highlighted with GCG program "Stem-loop" [Devereux, 84].
Selective RT-PCR was performed according to standard procedures using an oligo-dT primer for reverse transcription (42°C). The second strand was amplified using primer TrypSL corresponding to part of the spliced-leader and the H1 specific primer H1orf3' (Table 1).
Northern blots were performed after standard procedures [Ausubel, 1995]. RNA was separated through formaldehyde containing agarose gels, transferred onto positively charged Nylon membranes by downwards capillary transfer [Ausubel, 1995] and hybridized against specific biotinylated DNA probes or oligonucleotides corresponding to the different ORFs. Calibration was made against Tubulin alpha (kindly obtained from T. Seebeck, University of Bern, Microbiology).

Protein extraction:

Histone H1 proteins were extracted from procyclic culture forms by adapting a combination of cavitation [Shapiro and Doxsey, 1982] and perchloric acid extraction [Sanders, 1977]. The cells were lysed as described by Schlimme [Schlimme, 93] and the obtained cell lysate suspension was then processed as described by Noll [Noll, 1997]. Briefly, the cells were centrifuged 10 min. at 1800g at 4°C, resuspended in Lysis buffer (1mM PIPES; 2,5mM $CaCl_2$; 500mM hexylen glycol; 1mM PMSF; pH 7,4), centrifuged and resuspended in lysis buffer and incubated in a cavitation chambre under 23bar nitrogen for 3 hours at 4°C. After careful drop by drop recovery, the cell debris mixture was vortexed for a few minutes and centrifuged 5 min. at 7000 rpm (Eppendorf) at RT. The pellet was resuspended in 5% PCA and left to agitate gently on a rotor for 1 hour at RT. The mixture was then centrifuged for 15 min, at 13000rpm (Eppendorf) at RT and the H1 containing supernatant recentrifuged 5 min as above. H1 proteins were then precipitated overnight at –20°C with 8 vol of 100% ethanol and centrifuged at 13000rpm (Eppendorf) at RT. The H1 pellet was then washed in 70% ethanol and resuspended in an appropriate volume of PBS.

Treatment by alkaline phosphatase:

Different amounts of perchloric acid extracted H1 fractions resuspended in PBS (pH 8) or Tris-HCL (pH 8) were incubated with 5µl alkaline phosphatase (stock: 10mg/ml) (Boehringer, calf intestine alkaline phosphatase Grade II) supplemented with 1mM phenylmethanesulfonylfluoride (PMSF). Incubation was carried out at 37°C for 0, 1, 3 and 9 hours and samples were frozen at –20°C until gel analysis. The different samples were then tested by western blot.

Protein production and immunization:

Two clones corresponding to an 80 amino acids and a 76 amino acids H1 variant of *T. b. brucei* TREU 927/4 were amplified on verified Bluescript inserts with primers 5'-PET80 or 5'-PET76 combined with primer 3'-PET (Table 1). They were then ligated into expression vector pET 9a (Novagen) between NdeI and BamHI and transformed into *E. coli* Bl21. Induction was made according to standard procedures [Ausubel, 1995]. Briefly,

a freshly propagated bacterial colony was grown overnight in 3 ml LB (50µg kanamycine/ml). Then, 50 ml LB kanamycine were inoculated with 500 µl of the overnight preculture and grown at 37°C until OD_{600} reached 0,8 –1,0. The culture was then induced with 0,2 mM IPTG and grown for an additional hour. After a 15 min. centrifugation step at 3000g at 4°C, the pellet was resuspended in 3 ml PBS and left 15 min on ice. The cells were then sonicated at low amplitude 5 times for 1 min. and the suspension centrifuged 5 min. at 10'000g. The supernatant was then loaded onto standard 15% SDS-PAGE minigels for control. For purification, 200 µl of supernatant was loaded onto 1,5 mm 15% SDS PAGE minigels. The induced H1 protein was cut out of the gel and purified by the Biotrap® system (Schleicher & Schuell) by leaving only one space between BT1 and BT2 membranes. This allows recovery of purified proteins in a volume of approximately 500 µl. BalbC mice were immunized with 75 µl of the purified protein obtained this way and boosted 5 times in 2 weeks intervals. After bleeding of the mice, blood was left to coagulate for a few hours and then centrifuged 15 min at 3000 rpm (Eppendorf) at RT. Serum was aliquoted and stored at –20°C until use.

A synthetic peptide corresponding to the 12 first amino acids of the *T. b. brucei* 80-variant which differs from the other variants in this region was kindly obtained from Dr. G. Corradin (ISREC, Lausanne). The peptide was linked to the helper epitope P30. BalbC mice were immunized by the subcutaneous injection of 50 or 150 µg of peptide. The mice were boosted 7 times in 2 weeks intervals, until response could be monitored on control western blots.

Western blots:

Transfers were performed on a Trans-blot semi dry transfer cell® (Biorad) according to manufacturers instructions. Blots were treated by the TBS-Tween method [Ausubel, 1995] and antigen-antibody binding was revealed by the use of a peroxydase labeled anti-mouse IgG secondary antibody. Briefly, the nitrocellulose sheets were equilibrated in TBS/0,05% Tween20 (TBST) for 1 hour. Serum was diluted 1:100 in TBST and the blots were incubated 3 hours. After 3 washes of 5 min in TBST, the secondary antibody diluted 1:1000 in TBST was applied for 1 hour. After 3 washed as above and 2 additional washes in TBS, antibody binding was revealed by incubation with the substrate (4-chloro-1-naphtol). Alternatively, a milder method in which the membrane was blocked with dry milk was also used. Nitrocellulose membranes were saturated 3 X 15 min in PBS (pH:

7,35) / 5% dry milk. The first antibody was diluted as above in PBS/1% dry milk and applied to the membranes overnight at RT. After 3 washes of 5 min in PBS/5% dry milk, the second antibody diluted as above was applied and membranes left to incubate for 2 hours. Following 3 washes of 5 min in PBS/5% dry milk, 2 washes of 5 min in PBS and 1 wash in TB, antibody binding was revealed as in the TBST method.

Cell preparation for confocal microscopy:

Cells were fixed and permeabilized as described in [Ferguson, 1992]. Briefly, 10^7 to 10^8 cells were centrifuged 5 min at 4000 rpm (Eppendorf) at RT, washed in PBS and immediately fixed in PBS 3,5% paraformaldehyde for 10 min under constant and gentle inversion of the tube at RT. After 10 min, 0,2% TritonX100 and 0,1M Glycine (final conc.) were added and the cells incubated for a further 10 min under the same conditions. Cells were then washed 3 times in PBS and stored in PBS at a concentration of approximately 10^6 cells per ml at 4°C for up to 3 weeks.

Immunological reaction for confocal microscopy:

The immunological reaction was made in suspension. The cells were first blocked with Gold buffer (10mM Tris base; 155 mM NaCl; 2mM EGTA; 2 mM $MgCl_2$; pH 7,2)/1% BSA/5% normal goat serum (Sigma; G6767) for 1 hour, followed by 1 hour incubation with the first antibody (immune serum) diluted 1:100 in Gold buffer/1%BSA. After 3 washing steps in PBS, the cells were incubated 1 hour with the secondary antibody labelled with CY3 (red signal) diluted [1:1000] and picogreen (green signal) (Molecular Probes Inc.) diluted [1:100] in Gold buffer/1%BSA. After 3 washing steps in PBS, the cells were spread on poly-L-lysine slides (Sigma) and left to adhere for 1 hour, then the slides were washed 3 times in PBS, mounted with Vectashield® mounting medium (Vector Labs) and sealed. All incubation steps were at room temperature.

Oligonucleotides (5'-biotinylated)**:**		**T_{hyb}:**
TH1S1	5’-AAG GCT GTC GCT AAG-3’	41°C
TH1S2	5’-GCT AAG AAG GCT GCT-3’	41°C
TH1S3	5’-GCA TCT GCT GCT CCC-3’	45°C
OligoRNA-80	5’-GGT GGC CTT CAC AGT AGC GGT TGT G-3’	61°C

Primers: (specific)		**T_m:**
H1orf3'	5’-CAC TCA GGT AAA CCT CAT AGC-3’	57,9°C
5’-MNNTT	5’-CTT TAT CGA CTC CCC ACA AG-3’	57,3°C
5’-MAKAS	5’-CAA TCT TAT CAA CAC TCG GAA G-3’	56,5°C
5’-MAKTT/1	5’-GAA CAT TTC ATA GGA AAG TAG AAA GG-3’	58,5°C
5’-MAKTT/2	5’-GAA AGT AGA AAG GAA AAT AAA ATA TGG-3’	55,8°C
Walk5'-MNNTT	5’-GCC TTC ACA GTA GCG GTT G-3’	58,8°C
Tryp-SL	5’-CGC TAT TAT TAG AAC AGT TTC TG-3’	52,4°C
Tub-α5’	5’-ATG CGT GAG GCT ATC TGC ATC C-3’	62°C
Tub-α3’	5’-CTT CGT AGT CCT TCT CAA GTG C-3’	60°C
5’-PET80	5’-CCC ACA **CAT ATG** AAC AAC AC-3’	55,3°C
5’-PET76	5’-CTC GGA **CAT ATG** GCG AAG G-3’	58,8°C
3’-PET	5’-ACT CAG G**GG ATC C**TC ATA GC-3’	59,4°C

Primers: (universal)		**T_m:**
λgt11-Forward	5’-GAC TCC TGG AGC CCG-3’	56°C
λgt11-Reverse	5’-GGT AGC GAC CGG CGC-3’	58,8°C
M13-universal	5’-TGT AAA ACG ACG GCC AGT-3’	54°C
M13-reverse	5’-CAG GAA ACA GCT ATG ACC-3’	54°C

Table 1: Listing of all used oligonucleotides and primers. **T_{hyb}**: hybridization temperature; **T_m**: melting point (PCR reactions were usually performed at T_m -3°C for PCR and T_m+3°C for sequencing). Bold sequences correspond to introduced restriction sites for subcloning into expression vector: NdeI in 5’ primers and BamHI in the 3’ primer.

III. RESULTS

TRYPANOSOMA BRUCEI GAMBIENSE H1 CDNA ISOLATION

Preliminary considerations:

Histones are known to be among the best conserved proteins in evolution, however, this appears not to be true when analyzing the situation in lower eucaryotes. All available data to this date clearly indicate that trypanosomatids and also other protozoans do possess sets of histones that are divergent from their higher eucaryote counterparts. In addition, differences are also seen among lower eucaryotes. Therefore, cDNA isolation of trypanosome H1 coding sequences becomes more difficult and care must be taken to use appropriate probes.
Combination of acid extraction and separation on Triton Acid-Urea gels permitted to demonstrate the presence of histone H1 in African trypanosomes that were previously hidden because of their co-migration with the set of core histones on normal SDS-PAGE gels [Schlimme, 93]. Four variants and or post-translational modifications have been described for both procyclic insect forms [Schlimme, 93] and bloodstream forms [Burri, 94] of *Trypanosoma brucei brucei* (strain 345). Several peptide fragments were sequenced after H1 separation by reversed phase HPLC and subsequent enzymatic digestion [Burri, 95]. An additional source of information upon histone H1 in African trypanosomes comes from a cDNA expressed sequence tags (EST) study of *Trypanosoma brucei rhodesiense* which contains 3 sequences presenting high score pairing with histone H1 [El-Sayed, 95].

Results:

The peptide sequences obtained by Burri [Burri, 95] were back-translated and aligned to an expressed sequence tag (EST - T223) obtained from *T. b. rhodesiense* [El-Sayed, 95]. The best aligned regions have been retained and three oligonucleotides TH1S1, TH1S2 and TH1S3 (see material and methods: Table 1) corresponding respectively to the amino acid motifs KAVAK, AKKAA and ASAAP (**Fig. 16**) were used for the screening of a λgt11 cDNA library of *Trypanosoma brucei gambiense* (1257).

After screening of the cDNA library, seven clones that were recognized by all three, or only two of the used oligonucleotide probes were selected (**Table 2**) and purified by several rounds of plating/hybridization. They were then amplified and lambda phage DNA was extracted. Control of the obtained DNA was made by southern blot hybridization. After digestion with EcoRI, the inserts were separated through an agarose gel and transferred onto nylon membranes. The obtained southern blots were then hybridized successively with the three oligonucleotide probes (**Fig. 17**). The different inserts were very heterogeneous in size, ranging from 200 to 1000 bp. Some clones were only partially digested and two of them (clone 3 and 6) were not digested at all. Even though, hybridization could be observed on the undigested phage DNA at the top of the gel. The expected hybridization pattern already observed during screening was found. Probes TH1S1 and TH1S2 show the same picture while probe TH1S3, situated near the start codon (**Fig. 16**) gives differential information since only clones c1.10, c2.4 and c8.1 hybridize with this probe. In addition, clone c2.4 appears to contain an EcoRI restriction site since 2 bands of 200 and 600 bp are obtained after digestion (not shown). The smallest of these bands does not hybridize with any of the three oligonucleotides. On the southern blot, a faint band in the 800 bp region shows undigested DNA of clone c2.4 (Fig. 17, lane 5).

Sequencing and comparison of the seven cDNAs showed that some of the isolates were not complete, lacking either the 5' region (clones c9.1), the 3'-region (clone c1.10) or even both (clone c7.10), probably explaining the absence of hybridization with the probe TH1S3 (**Fig. 18**). Clone c1.9 revealed no real open reading frame (ORF) and alignment with the TH1S1 and TH1S2 probe sequences showed that hybridization was unspecific. Among the 7 cDNA clones, only clones c1.1, c2.4 and c8.1 contained a complete ORF with clone c2.4 presenting an additional artifact, the insert being formed of 2 cDNA fragments which probably ligated together before introduction into the lambda vector. Four clones contained a short poly-A tail (clones c1.1, c2.4, c8.1 and c9.1) but none of them carried a clear 5' spliced-leader sequence. Finally, clone c1.1 is not recognized by probe TH1S3 although it contains a complete ORF (see **Fig. 18** and **Table 3** for summary).

Multiple alignments of the 3'-untranslated regions (UTR) of the four polyadenylated clones allowed to highlight the similarity of their nucleotide sequences in the region

downstream the stop codon which was always conserved throughout several tenths of bases (**Fig. 19**). When going further downstream, it appears that the cDNAs can be grouped into two types. The 3'-UTR of clone c2.4 being identical to the one of clone c8.1, while clone c9.1 is rather similar to clone c1.1 (**Fig. 19**).

Primer H1orf3' (**Fig. 19** and **Table 1**) located on the conserved portion of the 3'-UTR of all clones was then designed in order to allow further H1 coding clone isolation from the cDNA library. Since the library was not oriented, H1orf3' was associated either with λgt11-forward or -reverse primer and multiple PCR reactions were made on positive plaques from the first screen or directly on purified DNA from the whole library. This approach gave no satisfying results and only clone c7.9 gave additional information in containing a shorter ORF than the already isolated clones (**Fig. 18** and **Table 3**).

Complete nucleotide sequences alignment of the different H1 coding cDNA clones that were obtained shows the high sequence similarity but also several substitutions (**Fig. 20**). Clone c1.1 clearly denotes from the other clones in its 5' non-coding and 5' coding region. This clone also contains some specific substitutions in the region of good sequence conservation if compared to the other clones (**Fig. 20**). The three oligonucleotides that were used for screening are all present except TH1S3 on clone c1.1 (**Fig. 20**).

Alignment of the deduced amino acid sequences of the obtained cDNA clones all showed the same global amino acids pattern with some differences in size (**Fig. 21**). The three amino acid motifs used for the oligonucleotide design were present and the peptide fragments described earlier [Burri, 95] could also be well aligned. Clone c1.1 is presenting a very different 11 amino acid long N-terminal region.

At this point, the 3 complete cDNA clones (c1.1, c8.1, c2.4) and an additional direct PCR amplified clone (c7.9) gave deduced gene products of 3 different sizes, respectively 85, 81 and 76 amino acids.

A

Sequenced peptides and selected fragments for probe design:

	TH1S2 **AKKAA**	*TH1S1* **KAVAK**	*TH1S3* **ASAAP**
>H1.1	AAKAAPKKTV**AKKAA**PPXAV		
>H1.2A	KVAPK		
>H1.2B		**KAVAK**	
>H1.3			AK**ASAAP**PXAVAKSAAPP
>H1.4			P**ASAAP**KSA

B

Tbr-t223 GANGTCGCAATCTTATCANCACTCGGNAGTATGGCGAAGGCATCTCCTGCnCCCAAAAAG
GCATCTGCTGCTCCC
A S A A P

Tbr-t223 GCTGTGGCTAAGANGGCAGCCCCCAAAAAGNCTGTCGCTAAGAAGGCTGCTNCAAAGnnG
GCTAAGAAGGCTGCT *AAG*
A K K A A K

Tbr-t223 GCTGTCGCTAAGAAGGNTNCTNCCAAAAAGGCTNTTGCCAAGAAGGTTNCTCCCAAAAAG
GCTGTCGCTAAG
A V A K

Tbr-t223 GGTTNTTGNCAAAANGGTTTGCCCCAAGGAGGTTTCCGGCAANNAGGCCNCCNCTAAGAA
Tbr-t223 GGCNTTAGNCATTCCCCTTNNTNCCCNNTTTTTGGCCANGNTTTTGGGGTTTTCCTNNGT
Tbr-t223 TTGGGNGNNNGCTTTTCANANGNTTTTAGGNNGGTCCTNGTNNNGNTCCCCTTCCNGGGC
Tbr-t223 GGGNNTTTTGGNTTTTCCNCCAANTNTTTTTTTCNTTTAGTTT

Figure 16: Probe design for screening. **A:** alignment of the amino acid motifs corresponding to the three designed probes TH1S1, TH1S2 and TH1S3 with the peptide fragments (H1.1 – H1.4) sequenced by Burri [Burri, 95]. Black boxes show portion of peptide fragments that were selected. **B:** Backtranslation of the three amino acid motifs and best alignment to the EST T223 isolated by El-Sayed [El-Sayed, 95] in order to avoid degenerate oligonucleotide probes. Underlines show matching nucleotides.

N°	TH1S1	TH1S2	TH1S3
1	+	+	-
2	+	+	-
3	+	+	-
4	+	+	+
5	+	+	+
6	+	+	-
7	+	+	+

Table 2: Hybridization pattern of the seven selected *T. b. gambiense* cDNA clones with probes TH1S1, TH1S2 and TH1S3. + indicates positive hybridization; - indicates absence of hybridization.

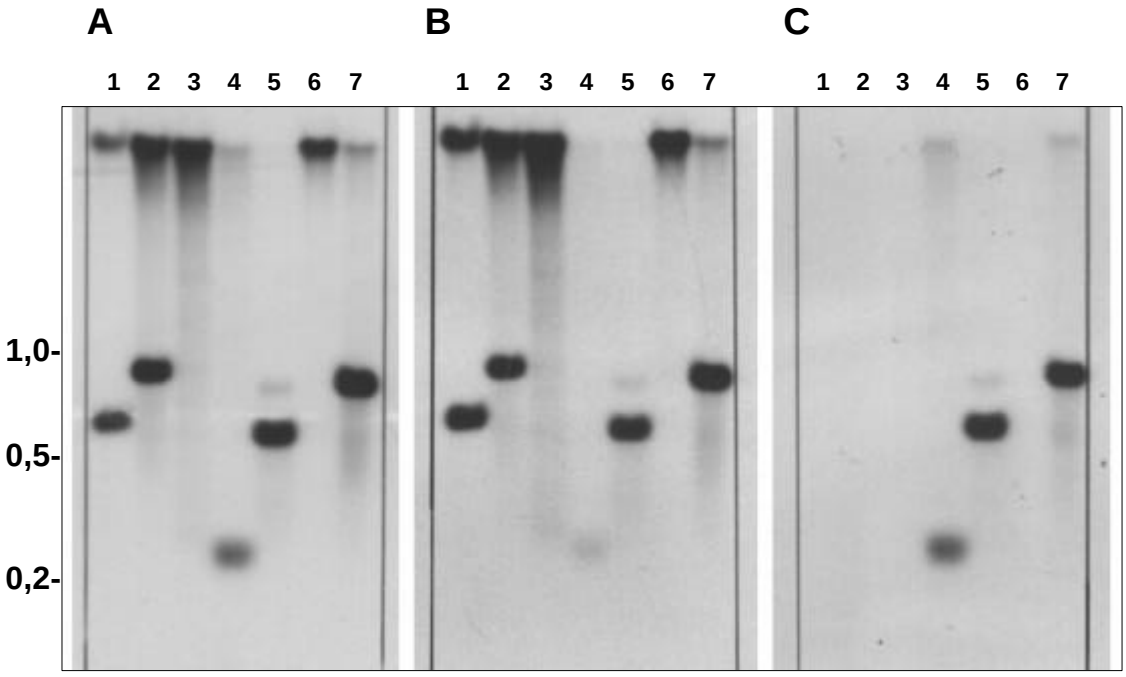


Figure 17: Southern blot hybridization of the 7 cDNA clones of *T. b. gambiense* that were selected by screening. **A**, **B** and **C**, hybridized respectively with probe TH1S1, TH1S2 and TH1S3. 1: clone c1.9; 2: clone c1.1; 3: clone c9.1; 4: clone c1.10; 5; clone c2.4; 6: clone c7.10; 7: clone c8.1.

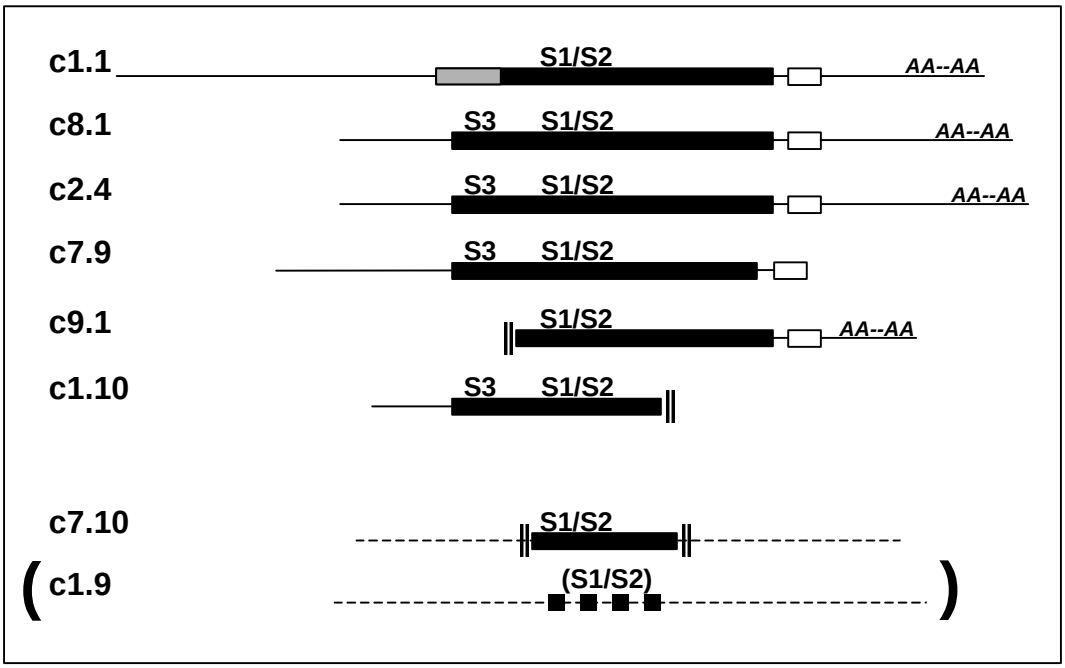


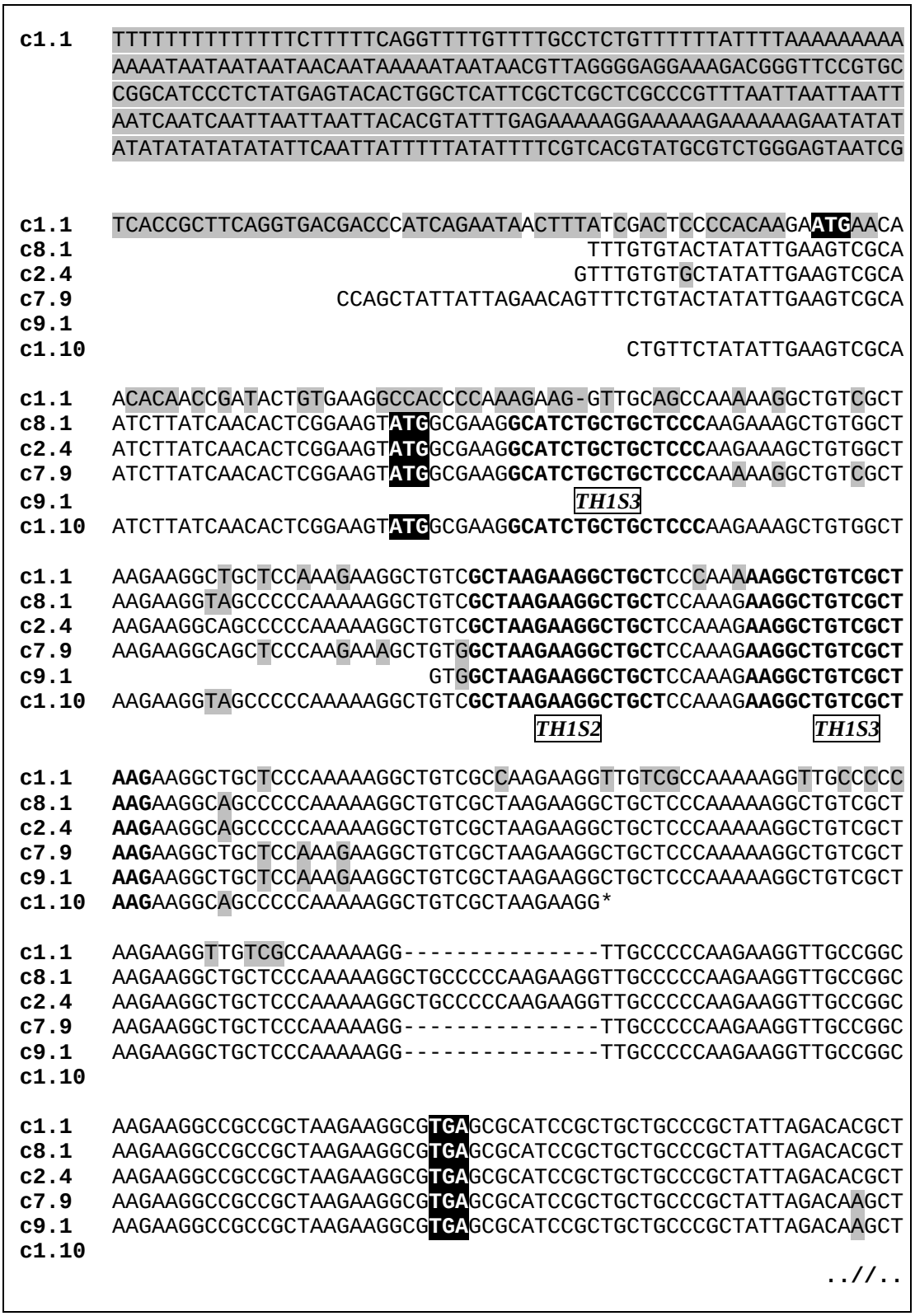
Figure 18: Schematic representation of the different *T. b. gambiense* cDNA clones. The lengths of the clones are represented by the lengths of the lines. Black boxes: ORFs; grey box: different 5'-coding region; open boxes: 3'-primer H1orf3' on the conserved region downstream the stop codon; double vertical bars: ORF truncation; ***AA-AA***: poly-A tail location; **S1**, **S2** and **S3**: oligonucleotide probes TH1S1, -2, -3; dotted lines: unrelated sequences; (S1/S2): unspecific probe alignment. Clone c7.9: additional clone obtained using the specific H1orf3' primer.

N°	cDNA clone	Insert size (bp)	ORF	Poly-A tail
1	c1.9	700	No	No
2	c1.1	900	Yes	Yes (18-A)
3	c9.1	450	No; only 3' region	Yes (9-A)
4	c1.10	250	No; only 5' region	No
5	c2.4	800 (600+200)	Yes	Yes (13-A)
6	c7.10	600	No; only 3' region	No
7	c8.1	800	Yes	Yes (16-A)
	c7.9	350	Yes	No

Table 3: Recapitulation of the features presented by the 7 selected cDNA clones. **1**: clone c1.9; **2**: clone c1.1; **3**: clone c9.1; **4**: clone c1.10; **5**; clone c2.4; **6**: clone c7.10; **7**: clone c8.1.

```
c2.4   TGAGCGCATCCGCTGCTGCCCGCTATTAGACACGCTATGAGGTTTACCTGAGTGCGGGAG
c8.1   TGAGCGCATCCGCTGCTGCCCGCTATTAGACACGCTATGAGGTTTACCTGAGTGTGGGAG
c1.1   TGAGCGCATCCGCTGCTGCCCGCTATTAGACACGCTATGAGGTTTACCTGAGTGTGGGAG
c9.1   TGAGCGCATCCGCTGCTGCCCGCTATTAGACAAGCTATGAGGTTTACCTGAGTGTGGGAG
       ******************************** ********************* *****

c2.4   AGAGCTGTCACACGTTTCAGGACGTCCTCGTGCGTCCCTCCAGGACGGAGTTAGATTTTT
c8.1   AGAGCTGTCACACGTTTCAGGACGTCCTCGTGCGTCCCTCCAGGACGGAGTTAGATTTTT
c1.1   AGAGCTGTCACACGTTTCAGGACGTCCTCGTGCGTCCCTCCAGGACGGAGTTAGATTTTT
c9.1   AGAGCTGTCACACGTTTCAGGACGTCCTCGTGCGTCCCTCCAGGACGGAGTTAGATTTTT
       ************************************************************

c2.4   CTTATCTTACTTTTTTATTTCTTTTTCACTGATTTTATTGGATATGTTTCGTTTGAGGGT
c8.1   CTTATCTTACTTTTTTATTTCTTTTTCACTGATTTTATTGGATATGTTTCGTTTGAGGGT
c1.1   CCTATCTTACTTGTTTAGTTCCCTTCTACCGTTTGTATTGGATATGTTTCGTTTGTGGGT
c9.1   CCTATCTTTTTTGTTTAGTTCCCTTCTACCGTTTGTATTGGAAAAAAAAAA----------
       * ******  ** **** ***  **  ** * ** *******

c2.4   TGCGTCTTATGTACCGCCATGCGGTGTTGGTGTCGTAGCGTTACAAAGAGCATATCATCC
c8.1   TGCGTCTTATGTACCGCCATGCGGTGTTGGTGTCGTAGCGTTACAAAGAGCATATCATCC
c1.1   TGCGTCTTATGTACGCCATGCGGTGTTGGTGTCGTAGCGTTACAAAGAGCATATCATCCT
c9.1   ------------------------------------------------------------

c2.4   TGATGTGTGGCTATTTTAACTGCCTGTGTATGGTTGTGTTCCAATAAAAAAAAAAAAAA
c8.1   TGATGTGTGGCTATTTTAACTGCCTGTGTATGGTTGTGTTCCAAAAAAAAAAAA-----
c1.1   GATGTGTGGCTATTTTAACTGCCTGTGTAAAAAAAAAAAAAAAAAA------------
c9.1   ------------------------------------------------------------
```

Figure 19: Alignment of the 3' non-coding sequences of the 4 cDNA clone of *T. b. gambiense* presenting a poly-A^+ tail. **TGA**: stop codon; black box show primer H1orf3'; poly-A tails are bold and italicized; *: matching positions between the 4 clones; best alignments are seen between clones c2.4 and c8.1 on the one hand and between c1.1 and c9.1 on the other hand; grey boxes indicate differences inside these two subdivisions. Note the differences in polyadenylation sites.


c1.1 TTTTTTTTTTTTTTCTTTTTCAGGTTTTGTTTTGCCTCTGTTTTTTATTTTAAAAAAAAA
AAAATAATAATAATAACAATAAAAATAATAACGTTAGGGGAGGAAAGACGGGTTCCGTGC
CGGCATCCCTCTATGAGTACACTGGCTCATTCGCTCGCTCGCCCGTTTAATTAATTAATT
AATCAATCAATTAATTAATTACACGTATTTGAGAAAAAGGAAAAAGAAAAAAGAATATAT
ATATATATATATATTCAATTATTTTTATATTTTCGTCACGTATGCGTCTGGGAGTAATCG
c1.1 TCACCGCTTCAGGTGACGACCCATCAGAATAACTTTATCGACTCCCCACAAGAAATGAACA
c8.1 TTTGTGTACTATATTGAAGTCGCA
c2.4 GTTTGTGTGCTATATTGAAGTCGCA
c7.9 CCAGCTATTATTAGAACAGTTTCTGTACTATATTGAAGTCGCA
c9.1
c1.10 CTGTTCTATATTGAAGTCGCA
c1.1 ACACAACCGATACTGTGAAGGCCACCCCAAAGAAG-GTTGCAGCCAAAAAGGCTGTCGCT
c8.1 ATCTTATCAACACTCGGAAGTATGGCGAAGGCATCTGCTGCTCCCAAGAAAGCTGTGGCT
c2.4 ATCTTATCAACACTCGGAAGTATGGCGAAGGCATCTGCTGCTCCCAAGAAAGCTGTGGCT
c7.9 ATCTTATCAACACTCGGAAGTATGGCGAAGGCATCTGCTGCTCCCAAAAAGGCTGTCGCT
c9.1 TH1S3
c1.10 ATCTTATCAACACTCGGAAGTATGGCGAAGGCATCTGCTGCTCCCAAGAAAGCTGTGGCT
c1.1 AAGAAGGCTGCTCCAAAGAAGGCTGTCGCTAAGAAGGCTGCTCCCAAAAAGGCTGTCGCT
c8.1 AAGAAGGTAGCCCCCAAAAAGGCTGTCGCTAAGAAGGCTGCTCCCAAAAGAAGGCTGTCGCT
c2.4 AAGAAGGCAGCCCCCAAAAAGGCTGTCGCTAAGAAGGCTGCTCCAAAGAAGGCTGTCGCT
c7.9 AAGAAGGCAGCTCCCAAGAAAGCTGTGGCTAAGAAGGCTGCTCCAAAGAAGGCTGTCGCT
c9.1 GTGGCTAAGAAGGCTGCTCCAAAGAAGGCTGTCGCT
c1.10 AAGAAGGTAGCCCCCAAAAAGGCTGTCGCTAAGAAGGCTGCTCCAAAGAAGGCTGTCGCT
TH1S2 TH1S3
c1.1 AAGAAGGCTGCTCCCAAAAAGGCTGTCGCCAAGAAGGTTGTCGCCAAAAAGGTTGCCCC
c8.1 AAGAAGGCAGCCCCCAAAAAGGCTGTCGCTAAGAAGGCTGCTCCCAAAAAGGCTGTCGCT
c2.4 AAGAAGGCAGCCCCCAAAAAGGCTGTCGCTAAGAAGGCTGCTCCCAAAAAGGCTGTCGCT
c7.9 AAGAAGGCTGCTCCAAAGAAGGCTGTCGCTAAGAAGGCTGCTCCCAAAAAGGCTGTCGCT
c9.1 AAGAAGGCTGCTCCAAAGAAGGCTGTCGCTAAGAAGGCTGCTCCCAAAAAGGCTGTCGCT
c1.10 AAGAAGGCAGCCCCCAAAAAGGCTGTCGCTAAGAAGG*
c1.1 AAGAAGGTTGTCGCCAAAAAGG--------------TTGCCCCCAAGAAGGTTGCCGGC
c8.1 AAGAAGGCTGCTCCCAAAAAGGCTGCCCCCAAGAAGGTTGCCCCCAAGAAGGTTGCCGGC
c2.4 AAGAAGGCTGCTCCCAAAAAGGCTGCCCCCAAGAAGGTTGCCCCCAAGAAGGTTGCCGGC
c7.9 AAGAAGGCTGCTCCCAAAAAGG--------------TTGCCCCCAAGAAGGTTGCCGGC
c9.1 AAGAAGGCTGCTCCCAAAAAGG--------------TTGCCCCCAAGAAGGTTGCCGGC
c1.10
c1.1 AAGAAGGCCGCCGCTAAGAAGGCGTGAGCGCATCCGCTGCTGCCCGCTATTAGACACGCT
c8.1 AAGAAGGCCGCCGCTAAGAAGGCGTGAGCGCATCCGCTGCTGCCCGCTATTAGACACGCT
c2.4 AAGAAGGCCGCCGCTAAGAAGGCGTGAGCGCATCCGCTGCTGCCCGCTATTAGACACGCT
c7.9 AAGAAGGCCGCCGCTAAGAAGGCGTGAGCGCATCCGCTGCTGCCCGCTATTAGACAAGCT
c9.1 AAGAAGGCCGCCGCTAAGAAGGCGTGAGCGCATCCGCTGCTGCCCGCTATTAGACAAGCT
c1.10
..//..

```
c1.1    ATGAGGTTTACCTGAGTGTGGGAGAGAGCTGTCACACGTTTCAGGACGTCCTCGTGCGTC
c8.1    ATGAGGTTTACCTGAGTGTGGGAGAGAGCTGTCACACGTTTCAGGACGTCCTCGTGCGTC
c2.4    ATGAGGTTTACCTGAGTGCGGGAGAGAGCTGTCACACGTTTCAGGACGTCCTCGTGCGTC
c7.9    ATGAGGTTTACCTGAGTG*
c9.1    ATGAGGTTTACCTGAGTGTGGGAGAGAGCTGTCACACGTTTCAGGACGTCCTCGTGCGTC
c1.10

c1.1    CCTCCAGGACGGAGTTAGATTTTTCCTATCTTACTTGTTTAGTTCCCTTCTACCGTTTGT
c8.1    CCTCCAGGACGGAGTTAGATTTTTCTTATCTTACTTTTTTATTTCTTTTTCACTGATTTT
c2.4    CCTCCAGGACGGAGTTAGATTTTTCTTATCTTACTTTTTTATTTCTTTTTCACTGATTTT
c7.9
c9.1    CCTCCAGGACGGAGTTAGATTTTTCCTATCTTTTTTGTTTAGTTCCCTTCTACCGTTTGT
c1.10

c1.1    ATTGGATATGTTTCGTTTGTGGGTTGCGTCTTATGTAC-GCCATGCGGTGTTGGTGTCGT
c8.1    ATTGGATATGTTTCGTTTGAGGGTTGCGTCTTATGTACCGCCATGCGGTGTTGGTGTCGT
c2.4    ATTGGATATGTTTCGTTTGAGGGTTGCGTCTTATGTACCGCCATGCGGTGTTGGTGTCGT
c7.9
c9.1    ATTGGAAAAAAAAAA*
c1.10

c1.1    AGCGTTACAAAGAGCATATCATCCTGATGTGTGGCTATTTTAACTGCCTGTGTAAAAAAA
c8.1    AGCGTTACAAAGAGCATATCATCCTGATGTGTGGCTATTTTAACTGCCTGTGTATGGTTG
c2.4    AGCGTTACAAAGAGCATATCATCCTGATGTGTGGCTATTTTAACTGCCTGTGTATGGTTG
c7.9
c9.1
c1.10

c1.1    AAAAAAAAAAA*
c8.1    TGTTCCAAAAAAAAAAA*
c2.4    TGTTCCAATAAAAAAAAAAAAA*
c7.9
c9.1
c1.10
```

Figure 20: Complete alignment of the nucleotide sequences of the six main *T. b. gambiense* cDNA clones c1.1, c8.1, c2.4, c7.9, c9.1 and c1.10. Black boxes: start and stop codons; grey boxes: varying positions; ***AAA***: poly-A tails; *: end of sequences. Regions in bold letters show the location of the three oligonucleotide probes TH1S1, -2, -3.

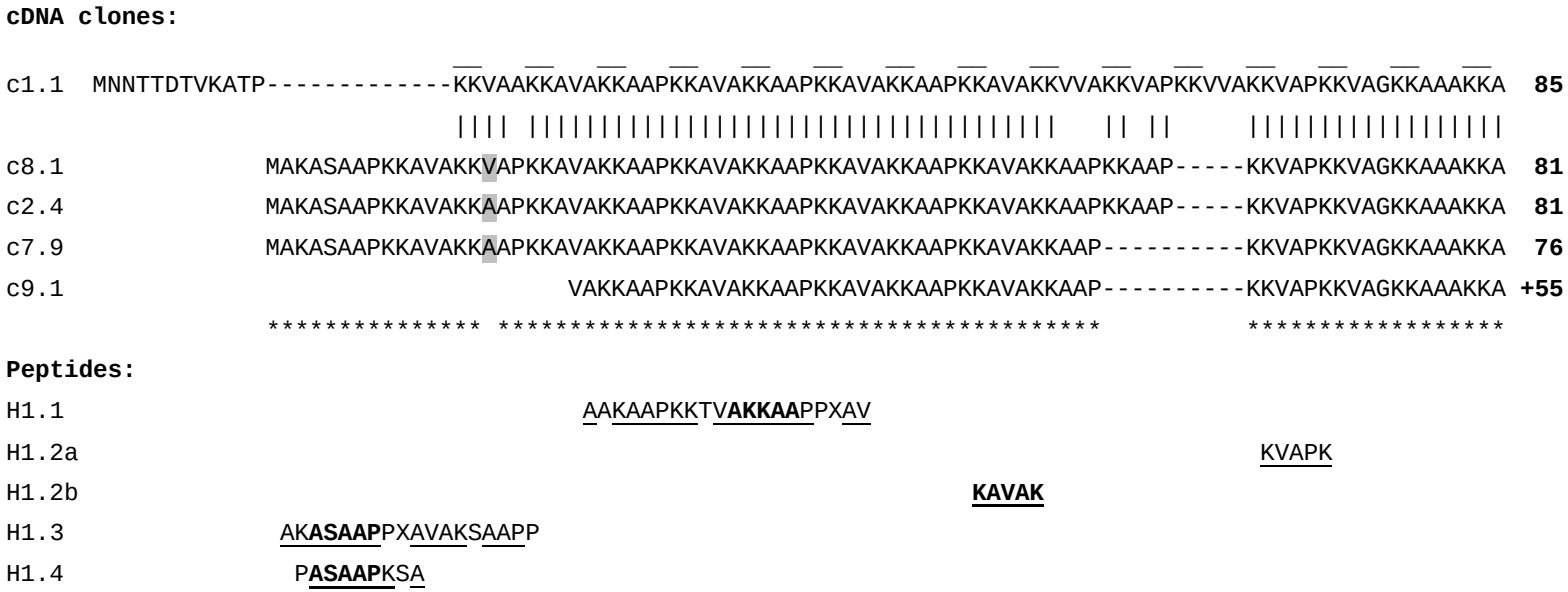


Figure 21: Protein alignment of the deduced amino acids sequences of different cDNA clones of *T. b. gambiense*. |: matching positions between clone c1.1 and the others; *: matching positions between all clones except c1.1; - indicate introduced gaps; —: shows perfect separation of lysine pairs; grey box shows varying position; numbers indicate variant size in amino acids. Below: alignment with the previously sequenced peptide fragments [Burri, 1995]. Matching positions are underlined; amino acid motifs corresponding to our probes are in bold letters.

GENOMIC H1 GENE ISOLATION

Preliminary considerations:

The screening of a cDNA library of *T. b. gambiense* allowed the isolation of 3 H1 variants of 85, 81 and 76 amino acids. This indicates that the 4 bands obtained by Schlimme and Burri on 2-dimensional gels were not due to posttranslational modifications but represented genetic variants. The approach taking advantage of 3 different oligonucleotide probes for screening allowed to select 2 variant types greatly differing in their 5' coding sequence. This is in good agreement with the heterogeneous H1 gene family described in *T. cruzi* [Åslund, 94] in which 2 different N-terminated proteins were obtained. In this organism, at least 6 variants that greatly resemble the isolated *T. b. gambiense* H1 cDNA clones when considering their length, global primary structure and amino acid composition were found. However, since four variants were described in *T. brucei* [Schlimme, 93; Burri, 93; 94], the cDNA clones obtained by screening of the *T. b. gambiense* library do probably represent only a partial image of the H1 gene family found in African trypanosomes. The additional fact that two equally sized H1 variants that differ by one amino acid were isolated lets suppose a still more complex situation.

Results:

Alignment of the 5'-sequence immediately upstream the start codon of the different cDNA clones obtained from the *T. b. gambiense* library shows a perfect match among all clones with the exception of c1.1 which also exhibited a different 5' coding sequence (**Fig. 22A**). Two additional H1-specific primers were designed on this region, primer 5'-MNNTT corresponding to the cDNA clone c1.1 and 5'-MAKAS corresponding to the other clones (**Fig. 22B** and **Table 1**).

In order to obtain a more complete image of the *T. brucei* H1 gene family, these two primers were used in association with primer H1orf3' to allow the gene isolation by PCR from genomic DNA.

Two PCR reactions associating primers 5'-MNNTT / H1orf3' or 5'-MAKAS / H1orf3' were performed on gDNA of *T. b. gambiense* (STIB755) and of *T. b. brucei* (TREU927/4). In both subspecies, both PCR reactions amplified fragments of about 300-350bp (**Fig. 23**).

The bands obtained in *T. b. gambiense* appear more smeary indicating that they probably contain a mixture of closely sized fragments (**Fig 23; lanes b and e**). Bands ranging around 900 bp and faint shadows beside the 1,5 kbp marker were also observed in all amplifications (**Fig. 23**).
The banding patterns are quite similar in both subspecies but the band obtained with primer 5'-MAKAS is slightly smaller in *T. b. brucei* if compared to the ones obtained in *T. b. gambiense* (**Fig. 23; lanes e and f**). In addition, unlike in *T. b. gambiense*, in *T. b. brucei* the 5'-MAKAS band of 350 bp is smaller than the 5'-MNNTT band, indicating that the histone H1 gene family could display some intraspecific heterogeneity.

Subcloning and sequencing of the two 300-350 bp bands of *T. b. gambiense* resulted in a surprising assortment of 7 H1 variants ranging from 91 to 61 amino acids. Two intermediate variants of 85 and 75 amino acids were bearing the different 11 amino acids long N-terminal region (**Fig. 24**). Complete nucleotide sequence alignments of *T. b. gambiense* H1 genes are presented in **Annex B**.
Sequencing of the two 300-350 bands obtained in *T. b. brucei* allowed to group similar but different deduced ORFs into 4 different size classes coding for proteins of 80, 76, 71 and 61 amino acids (**Fig. 25**). Among these four variants, only the longest one carries the different N-terminal sequence and the three smaller ones that are obtained with primer 5'-MAKAS correspond to variants differing from each other only by some deletions in their simple and quite repetitive sequence. The peptide sequences published earlier [Burri, 95] could also be well aligned (**Fig. 25**). Complete nucleotide sequence alignments of *T. b. brucei* H1 genes are presented in **Annex C**.

With the exception of the variety of differently sized variants, the H1 gene products of the 2 subspecies show quite the same global organization and only some minor sequence differences are seen. The primary structure of African trypanosome H1 do present a really monotone motif throughout the whole length consisting of lysine pairs always separated by three non-polar and hydrophobic amino acids (alanine, valine, proline and leucine) with the exception of the glycine positions (**Fig. 24 and 25**). Only 2 variants of *T. b. gambiense* display a short region where the lysine pairs rule is disturbed (**Fig. 24**). Inside the 11 amino acids extra N-terminal sequence found on the 80 amino acid variant in *T. b. brucei* and in the 75 and 85 amino acid variants in *T. b. gambiense*, the two subspecies do only differ by 1 amino acid, the alanine at position 6 seen in *T. b. brucei* being substituted by an aspartate

in *T. b. gambiense* (**Fig. 26A**). The heterogeneity seen inside the rest of these variants or among the other size variants is only due to different arrangement of the 3 amino acids separating the lysines. The last 18 C-terminal amino acids are conserved among all variants and in both subspecies (**Fig. 26A and B**).

The amino acid composition of H1 of *T. b. brucei* and *T. b. gambiense* is very similar and only 7 to 9 different residues are used (**Table 4A**). Histone H1 is usually characterized by a high lysine to arginine ratio (around 20), however, because of the absence of arginine in *T. brucei* H1 variants, this value cannot be calculated. The four major amino acids lysine (36-38%), alanine (30-40%), valine (8-15%) and proline (7-10%) constitute 87 to 96% of the total amino acid content. Beside this, one or two glycines or leucines are introduced inside the C-terminal region and the other additional amino acids methionine, threonine and serine are found only near the start codon where most differences are seen between variants. Two asparagines are exclusively found on the H1 variants that bear the extra N-terminal region in both subspecies, while one aspartate is found only in *T. b. gambiense*.

When comparing the *T. b. brucei* H1 amino acid composition obtained from the deduced sequences with the data which was previously obtained by biochemical procedures [Burri, 93], the proteins appear to display much more simple features than would have been expected and does not reflect previously obtained data (**Table 4B**).

When submitting the H1 clones to a BLAST search, the sequences presenting the highest similarity are, indeed, the different *T. cruzi* H1 variants. Unlike in *T. b. brucei* H1, the lysine pair organization is less regular in *T. cruzi* (**Fig. 27A**), but beside this, size and general primary structure appear similar. However, with exclusion of the extra N-terminal end, the variety of used amino acids is higher in *T. cruzi* (**Table 5**) which introduces arginine and histidine as well as some additional threonines but excludes glycine.

Interestingly, there were three other proteins showing yet a better alignment pattern. The first is a recently isolated H1 of *Leishmania brasiliensis* (unpublished, Genbank Acc. Number AF131892) that exhibits a C-terminal region very similar to the one seen in African trypanosomes (**Fig 27B**). The second and probably the most interesting one is a histone H1-homologue from the prokaryote *Bordetella pertussis* [Scarlato, 95]. The central region of this latter shows the same repetitive pattern of two lysines followed by three non-polar and hydrophobic amino-acids (A, V, P or L) (**Fig 27B**).

The third is referred to as a *T. b. brucei* microtubule associated protein and matches the isolated H1 clones along the 3' coding region but is incomplete, missing the 5' end (unpublished, Genbank Acc. Number L41654) (**Fig 27C**).

The comparison of global amino acid content against other lower eucaryotes (**Table 5**) clearly shows that the H1 histones of *T. b. brucei* are very simple proteins. If compared to American trypanosomes, other kinetoplastids, other protozoa or even a H1-homologue of a procaryote, the diversity of the used amino acids appears to be the most limited in *T. brucei*. It has already been reported that H1s of kinetoplastids only correspond to the C-terminal tail of their higher eucaryotic counterparts [Galanti, 98]. Indeed, the listing of the amino acid content of a classic H1 represented by the sea urchin *Lytechinus pictus* in table 5 shows a much higher complexity but when limiting this latter to its C-terminal region, the composition strongly resembles the one of lower eucaryote H1 histones.

A

```
c1.1     301  TCACCGCTTCAGGTGACGACCCATCAGAATAACTTTATCGACTCCCCACAAGAATG
                       |  |    | |      ||    | | |    |   | |  ||| |||
c2.4          -------GTTTGTGTGCTATATTGAAGTCGCAATCTTATCAACACTCGGAAGTATG
c8.1     238  CTGGGTCTTTTGTGTACTATATTGAAGTCGCAATCTTATCAACACTCGGAAGTATG
c1.10         -----------CTGTTCTATATTGAAGTCGCAATCTTATCAACACTCGGAAGTATG
                         *** ************************************
```

B

```
c1.1  301  TCACCGCTTCAGGTGACGACCCATCAGAATAACTTTATCGACTCCCCACAAGAATGAACA
                 |||  |    | |      ||    | | |    |   | |  ||| |||   |
c8.1  238  CTGGGTCTTTTGTGTACTATATTGAAGTCGCAATCTTATCAACACTCGGAAGTATGGCGA

c1.1       ACACAACCGATACTGTGAAG---GCCACCCCAAAGAAGGTTGCAGCCAAAAAGGCTGTCG
           |  || | | | ||   |||   ||     | |||||||| ||  |||||||||||||||
c8.1       AGGCATCTGCTGCTCCCAAGAAAGCTGTGGCTAAGAAGGTAGCCCCCAAAAAGGCTGTCG

c1.1       CTAAGAAGGCTGCTCCAAAGAAGGCTGTCGCTAAGAAGGCTGCTCCCAAAAAGGCTGTCG
           |||||||||||||||||||||||||||||||||||||||| || ||||||||||||||||
c8.1       CTAAGAAGGCTGCTCCAAAGAAGGCTGTCGCTAAGAAGGCAGCCCCCAAAAAGGCTGTCG

c1.1       CTAAGAAGGCTGCTCCCAAAAAGGCTGTCGCCAAGAAGGTTGTCGCCAAAAAGGTTGCCC
           |||||||||||||||||||||||||||||| ||||||| ||   ||||||||| |||||
c8.1       CTAAGAAGGCTGCTCCCAAAAAGGCTGTCGCTAAGAAGGCTGCTCCCAAAAAGGCTGCCC

c1.1       CCAAGAAGGTTGTCGCCAAAAAGGTTGCCCCCAAGAAGGTTGCCGGCAAGAAGGCCGCCG
           |||||||||               ||||||||||||||||||||||||||||||||||||
c8.1       CCAAGAAGG--------------TTGCCCCCAAGAAGGTTGCCGGCAAGAAGGCCGCCG

c1.1       CTAAGAAGGCGTGAGCGCATCCGCTGCTGCCCGCTATTAGACACGCTATGAGGTTTACCT
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
c8.1       CTAAGAAGGCGTGAGCGCATCCGCTGCTGCCCGCTATTAGACACGCTATGAGGTTTACCT

c1.1       GAGTGTGGGAGAGAGCTGTCACACGTTTCAGGACG   694
           |||||||||||||||||||||||||||||||||
c8.1       GAGTGTGGGAGAGAGCTGTCACACGTTTCAGGACG   618
```

Figure 22: A: Alignment of the sequences immediately preceding the start codon of 4 cDNA clones from *T.b.gambiense* (1257). |: matching positions between clone c1.1 and the others; *: matching positions between all clones but c1.1; **ATG**: start codon. **B**: Alignment of two representative ORFs. |: matching positions; **ATG** and **TGA**: start and stop codons; ORFs are in italics; single line: primer 5'-MNNTT; double line: primer 5'-MAKAS; dotted line: primer H1orf3'; numbers indicate position in the cDNA clone.

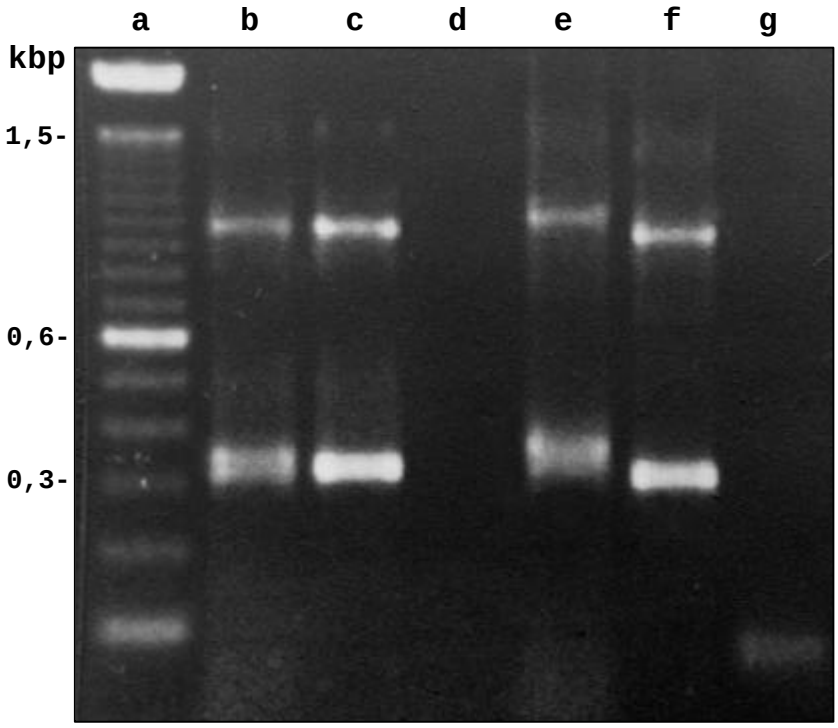


Figure 23: PCR amplification on genomic DNA with the different 5'-primers in association with the conserved 3'-primer H1orf3'. Lane a: 100bp DNA ladder. Lanes b and e: amplification on *T. b. gambiense* gDNA with primer 5'-MNNTT (lane b) and with primer 5'-MAKAS (lane e). Lanes c and f: amplification on *T. b. brucei* gDNA with primer 5'-MNNTT (lane c) and with primer 5'-MAKAS (lane f). Lanes d and g: respective negative controls.

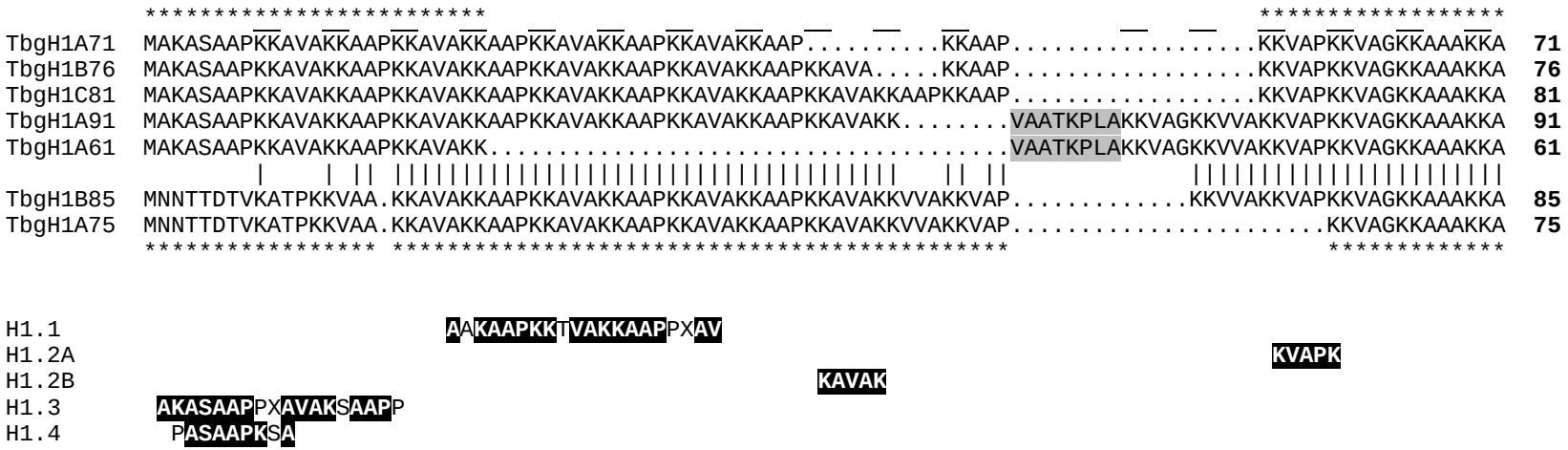


Figure 24: Alignment of the different H1 variants obtained by PCR amplification on genomic DNA in *T. b. gambiense* (STIB755). Upper asterisks: matching positions between the variants obtained by primer 5'-MAKAS; Asterisks below; matching positions between the variants obtained by primer 5'-MNNTT; |: matching positions between the two differently 5'-ended types of H1 variants. —: regular lysine pairs organization; grey box: unique position where the regular spacing of lysine pairs is abolished; numbers indicate variant size in amino acids. Below: alignments of the five peptide fragments sequenced by Burri [Burri; 95]; black boxes: matching positions. Names of variants correspond to Genbank submission (see Annex A).

Figure 25: Alignment of the different H1 variants obtained by PCR amplification on genomic DNA in *T. b. brucei* (TREU927/4). |: matching positions between the two differently 5'-ended types of H1 variants; *: conserved positions inside the variants obtained with primer 5'-MAKAS; ;grey boxes: substitutions; —: regular spacing of the lysine pairs; numbers indicate variant sizes in amino acids. Below: alignments of the five peptide fragments sequenced by Burri [Burri; 1995]; black boxes: matching positions. Names of variants correspond to genbank submission (see Annex A).

A

```
..........++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
TbgH1B85  MNNTTDTVKATPKKVAAKKAVAKKAAPKKAVAKKAAPKKAVAKKAAPKKAVAKKVVAKKVAPKKVVAKKVAPKKVAGKKAAAKKA  85
TbgH1A85  MNNTTDTVKATPKKVAAKKAVAKKAAPKKAVAKKAAPKKAVAKKAAPKKAVAKKVVAKKVAPKKVVAKKVAPKKVAGKKAAAKKA  85
TbbH1A80  MNNTTATVKATPKKVAAKK-----AAPKKTVAKKAAPKKAVAKKPLAKKVVAKKAAPKKAVAKKAAPKKVAPKKVAGKKAAAKKA  80
          ***** **************     ***** **************   ** ****   **   **   ******************
```

B

```
TbgH1B76  MAKASAAPKKAVAKKAAPKKAVAKKAAPKKAVAKKAAPKKAVAKKAAPKKAVAKKAAPKKVAPKKVAGKKAAAKKA  76
TbbH1A76  MAKASAAPKKAVAKKAAPKKAAPKKAVAKKAAPKKAVAKKAAPKKAVAKKPLAKKAAPKKVAPKKVAGKKAAAKKA  76
          ********************   ***  ***  ***  ***  ***  **  ************************
```

Figure 26: Alignments comparing between *T.b.brucei* and *T.b.gambiense* H1 variants. **A**: Alignment of three variants obtained by primer 5'-MNNTT. TbgH1B85: *T.b.gambiense* (strain STIB755); TbgH1A85: *T.b.gambiense* (strain 1257); TbbH1A80: *T.b.brucei* (strain TREU927/4). +: indicate conserved positions between both *T.b.gambiense* variants; *: matching positions between both subspecies; grey box: substitution inside the special N-terminal end between both subspecies. **B**: Alignment of two 76 amino acid long variants obtained by primer 5'-MAKAS. TbgH1B76: *T.b.gambiense*; TbbH1A76: *T.b.brucei*. *: matching positions; numbers indicate variant size in amino acids. Names correspond to genbank submission (see Annex A).

A aa	*Tbb* H1-80	*Tbg* H1-85	*Tbg* H1-75	*Tbb* H1-76	*Tbb* H1-71	*Tbb* H1-61	*Tbg* H1-91	*Tbg* H1-81	*Tbg* H1-76	*Tbg* H1-71	*Tbg* H1-61
ala	31,2	30,6	32,0	37,1	39,0	37,7	36,3	39,5	39,5	39,4	34,4
cys											
asp		1,2	1,3								
glu											
phe											
gly	1,2	1,2	1,3	1,8	1,9	3,3	2,2	1,2	1,3	1,4	3,3
his											
ile											
lys	36,2	36,5	36,0	38,2	38,0	37,2	37,4	38,3	38,2	38,0	36,1
leu	1,2			1,3			1,1				1,6
met	1,2	1,2	1,3	1,3	1,4	1,6	1,1	1,2	1,3	1,4	1,6
asn	2,5	2,4	2,7								
pro	8,8	7,1	6,7	9,9	9,9	9,8	7,7	9,9	9,2	9,9	6,6
gln											
arg											
ser				1,3	1,4	1,6	1,1	1,2	1,3	1,4	1,6
thr	6,2	4,7	5,3				1,1				1,6
val	11,2	15,3	13,3	9,2	8,5	8,2	12,1	8,6	9,2	8,5	13,1
trp											
tyr											

B aa	**e1**	**e2**	**e3**	**e4**	***Tbb*H1-80**	***Tbb*H1-76**	***Tbb*H1-71**	***Tbb*H1-61**
ala	36,5	36,3	31,2	35,8	31,2	37,1	39,0	37,7
cys	-	-	-	-				
asp	2,7	3,9	3,4	3,1				
glu	2,3	3,0	2,0	2,2				
phe	0,7	0,7	0,3	0,3				
gly	3,9	3,2	4,2	2,4	1,2	1,8	1,9	3,3
his	0,0	0,0	0,0	0,0				
ile	0,7	0,7	0,3	0,8				
lys	20,6	22,4	27,8	20,2	36,2	38,2	38,0	37,2
leu	1,7	0,8	0,7	1,0	1,2	1,3		
met	0,0	0,0	0,0	0,0	1,2	1,3	1,4	1,6
asn	-	-	-	-	2,5			
pro	7,5	9,6	11,6	11,2	8,8	9,9	9,9	9,8
gln	-	-	-	-				
arg	1,5	1,4	1,4	1,2				
ser	3,0	3,6	3,8	3,7		1,3	1,4	1,6
thr	3,5	2,5	2,9	3,6	6,2			
val	14,0	10,2	10,1	13,9	11,2	9,2	8,5	8,2
trp	-	-	-	-				
tyr	1,6	1,7	0,5	0,6				

Table 4: Amino acid compositions of the different H1 variants in mol%. **4A**: Representation of the deduced amino acid composition of the different H1 size classes from both subspecies; *Tbb*: *T.b.brucei* (highlighted in grey) and *Tbg*: *T.b.gambiense*. Variants presenting similar N-terminal ends are listed side by side. **4B**: Comparison between the previously described amino acid composition of the 4 *T.b.brucei* H1 variants and/or posttranslational modifications (e1–e4) [Burri, 93] and the deduced amino acid composition obtained by gene isolation (TbbH1-80, -76, -71 and -61).

A

```
76B        MAKASAAPKKAVAKKAAPKKAAPKKAVAKKAAPKKAVAKKAAPKKAVAKKPLAKKVVAKK
           |  |   ||||  |||| ||| |||| | | | ||   | |  | |  |    ||  | |
Tcr(c8)    MSDAAVPPKKASPKKAAAKKASPKKAAARKTAAKKTAKKPAVRKPAAKKRAAPKKKPAAK

76B        VAPKKVAGKKAAAKKA*      76
            ||||   |||  ||
Tcr(c8)    KAPKKAV-KKAPKKK*       74
                                  Total size
```

B

```
Bp     61..VAKKAVAKKAVAKKAVAKKAVAKKAVAKKAVAKKAVAKKAPAKKAVAKKAVAKKAVAKKA
                        |  ||||||||  |||  ||||||||  ||||||||  |||||||
76B          ---------MAKASAAPKKAVAKKAAPKKAAPKKAVAKKAAPKKAVAKKAAPKKAVAKKP
                        || ||| ||||| ||||||||  | |||| |  ||||||||  ||
Lbr    14..SNSPQRSPRPSPKKAAVKKAAAKKAAAKKAAPKKAAPKRAAPKRAAPKKAAPKKAAAKRA

Bp         VAKKAVAKKAVPKKAPAKKAAPKKPATPPSTAAAPGAKTALN..163    182
            ||| ||||  |||   |||| ||
76B        LAKKVVAKKVAPKKVAGKKAAAKKA*                        76
            |||   ||   | |   | | |||
Lbr        -AKKSAPKKAVKKAVKAAKKAVKKAAKKAT*                   102
                                                            Total size
```

C

```
MAP        -ARATAVPKKAVAKKAAPKKTVAKKAAPKKAVAKKVAPKKAVAKKVVAKKAVAKKVVAKK
            | | | ||||||||||||     |||||||||||| |||||||||   |||||||  ||
76B        MAKASAAPKKAVAKKAAP-----KKAAPKKAVAKKAAPKKAVAKKAAPKKAVAKKPLAK-

MAP        VAPKKVVAKKVAPKKVAGKKAAAKKA
               ||||||||||||||||||||||
76B        ----KVVAKKVAPKKVAGKKAAAKKA
```

Figure 27: Alignments of a 76 amino acid H1 variant of *T. b. brucei* (TbbH1B76) against other organisms. *: end of sequence; |: matching positions; numbers indicate protein size in amino acids or position. **A**: Alignment with the closest sized H1 variant described in *T. cruzi*: Tcr(c8) [Åslund, 94]; Underline: indicates less regular spacing of lysine pairs in *T. cruzi*; grey boxes: substitution of lysines by arginines. **B**: Alignment with H1-homologue of *Bordetella pertussis* (BpH1) [Scarlato, 95] and H1 of *Leishmania brasiliensis* (Lbr) [unpublished]. **C**: Alignment with a sequence referred to as a microtubule associated protein (MAP) [unpublished].

aa	*Tbb* H1-80	*Tc* H1.C2	*Tbb* H1-71	*Tc* H1.C8	*Lbr*	*Lmj*	*Cf*	*Vc*	*Bp*	*Lp*	*Lp* C-term
ala	31,2	23,7	39,0	31,1	36,6	20,0	31,8	37,8	34,6	29,5	38,3
cys		2,1									
asp		1,0		1,4					1,1	0,5	
glu							0,7	2,8		2,4	2,8
phe		2,1			0,9	1,0	0,7		0,5	1,0	
gly	1,2		1,9			1,9	4,6	0,7	0,5	4,3	
his		1,0					1,3			0,5	
ile		2,1					2,6			2,4	
lys	36,2	33,0	38,0	37,8	31,2	35,2	18,5	39,9	32,4	31,9	45,8
leu	1,2	4,1					2,0		1,1	2,9	1,9
met	1,2	1,0	1,4	1,4	0,9	1,0	2,0		0,5	1,0	
asn	2,5	1,0			1,8	1,0			0,5	1,9	
pro	8,8	10,3	9,9	13,5	8,9	4,8	9,9	14,0	8,8	5,7	7,5
gln		1,0			0,9	1,0	1,3			2,9	0,9
arg		2,1		4,1	5,4	1,9	7,3		1,1	2,4	0,9
ser		4,1	1,4	4,1	6,2	12,4	10,6	1,4	2,2	2,9	
thr	6,2	4,1		2,7	2,7	4,8	1,3	2,8	3,3	3,3	0,9
val	11,2	7,2	8,5	4,1	4,5	15,2	5,3	0,7	12,6	3,8	0,9
trp									0,5		
tyr										1,0	

Table 5: Comparison of mol% amino acid contents of two H1 variants of *T. b. brucei* with several other organisms. The four trypanosome H1 proteins are grouped regarding their N-terminal ends. ***Tbb***: *Trypanosoma brucei brucei* (highlighted in grey); ***Tc***: *Trypanosoma cruzi*; ***Lbr***: *Leishmania brasiliensis*; ***Lmj***: *Leishmania major*; ***Cf***: *Crithidia fasciculata*; ***Vc***: *Volvox carteri*; ***Bp***: *Bordetella pertussis*; ***Lp***: *Lytechinus pictus*; ***Lp C-term***: *Lytechinus pictus* restricted to the C-terminal region.

GENOMIC ORGANISATION OF H1 GENES IN *T. B. BRUCEI*

Preliminary considerations:

The isolation of histone H1 genes from genomic DNA allowed to obtain several variants differing in size and structure. The expected 4 size variants of *T. b. brucei* that were obtained proved that the four bands obtained on TAU gels in *T. b. brucei* [Schlimme, 93; Burri, 93; 94] were not due to posttranslational modifications but to true genetic variants. However, the heterogeneity inside the four different size classes shows that the genetic organization is very complex with a multiplicity of genes that differ by some substitutions. In *T. cruzi*, the H1 genes have been described to be tandemly arranged and an array of at least 20 copies must be present in a head to tail fashion [Åslund, 94]. In this organism, the H1 genes have also been shown to be located on a chromosome of 2,2 Mbp while the core histones are located on different chromosomes [Bontempi, 94]. This is very different from what is found in higher eucaryotes where histone genes are usually organized in quintets containing one copy of each histone. The good general similarity between H1 of American and African trypanosomes lets expect the same kind of genomic organization.

Results:

In order to analyze the genomic organization of the isolated H1 coding genes, a PCR reaction was performed on genomic DNA of *T. b. brucei* (TREU927/4) with double quantity of template DNA and by increasing the number of cycles by 5 if compared to previous amplifications that were performed (see Genomic H1 gene isolation) (**Fig. 28A**). This adaptation was made in order to enhance amplification and to see if the shadows seen around 1,5 kbp were real bands. At least the reaction integrating primer 5'-MNNTT allowed to obtain a clear banding of this size (**Fig. 28A; lane c**).

The additional bands I, II and III obtained by this new PCR amplification (**Fig. 28A**) were subcloned for sequencing (mainly the 0,9 kbp bands). After plasmid DNA purification and PCR amplification of the inserts with universal primers (M13universal and reverse), several clones showing faint size variations were selected (**Fig. 28B**). Sequencing of three clones corresponding to band I that was obtained by primer 5'-MAKAS (**Fig 28B; lanes b, c and h**) as well as of three clone corresponding to band II that was obtained by primer 5'-

MNNTT (**Fig 28B; lanes o, p and q**) allowed to obtain several different tandems consisting in head to tail arranged ORFs coding for histone H1. All three tandems issued from band II presented the same composition with an 80 amino acid coding ORF being followed by a 76 amino acid variant coding ORF. Band I, on the other hand, gave rise to different tandems. The two largest clones comprised respectively a 76-76 and a 76-71 amino acid variant coding ORF pair (**Fig. 29A**).

In all cases, the first ORF was longer or equal in size if compared with the following one and ORFs coding for 80 amino acid variants were always in head position if present. A tandem composed of two 80 variant genes following each others was not found. A control PCR associating primer 5'-MNNTT and primer Walk5'-MNNTT (**Table 1**) located on the N-terminal coding region of the 80 variants and oriented upstream gave no amplification, indicating that 80 variant genes are not situated near from each other (**Fig. 29B**). In order to ascertain the head position of the 80 variant coding gene in the cluster, a PCR reaction using primers 5'-MAKAS with primer Walk5'-MNNTT was performed. No amplification could be obtained and thus clearly locates this gene in front of the other genes (**Fig. 29B**).

The different tandems that were obtained and the additional information concerning respective positioning of the ORFs allows to construct a partial and theoretical image of the H1 gene cluster in African trypanosomes (**Fig. 29A; bottom**).

The third clone issued from band I was containing a 71 amino acid variant ORF followed by a peculiar ORF coding for a protein of only 36 amino acids (**Fig. 29C**). This variant virtually corresponds only to the N-terminal and C-terminal regions always conserved in all the other variants, and completely lacks the median zone in which substitutions and deletions are seen in the other variants (**Fig. 30**). Complete nucleotide sequence of this tandem is presented in **Annex D**.

Only one clone of 1,5 kbp issued from band III (**Fig. 28A**) was analyzed. The composition of this array of H1 coding genes added to the complexity of the theoretical organization presented in figure 29A. This clone presented 3 ORFs coding for variants of 80, 76 and 71 amino acids following each other (**Fig. 29D**). Though, since a tandem composed of two 76 variant genes was obtained, the decreasing size arrangement appears to not to be strict. Complete nucleotide sequence of this clone is presented in **Annex E**.

The ORFs are all separated by an extremely well conserved intergenic region (IR) of about 430 bases (**Fig. 31A**). Only a few substitutions are seen, especially for the IR separating the ORFs coding for 80 and 76 amino acid variants (between positions 389 and 412 on Figure 31A). A perfectly conserved 95 base long polypyrimidine tract is present in all IRs that were obtained (**Fig. 31A**). This kind of motifs have been shown to be implicated in both trans-splicing and polyadenylation. However, the substitutions mentioned above are not located within this important region.

When submitting the different IRs to computer analysis for theoretical stem-loop formation, the longest and most stable formation implicated the region of high variability. The stem-loop is only seen in the IR separating 80 and 76 variant ORFs and is formed of 15 stem-bases, the loop consisting quite exclusively of uracile (**Fig. 31A and B**).

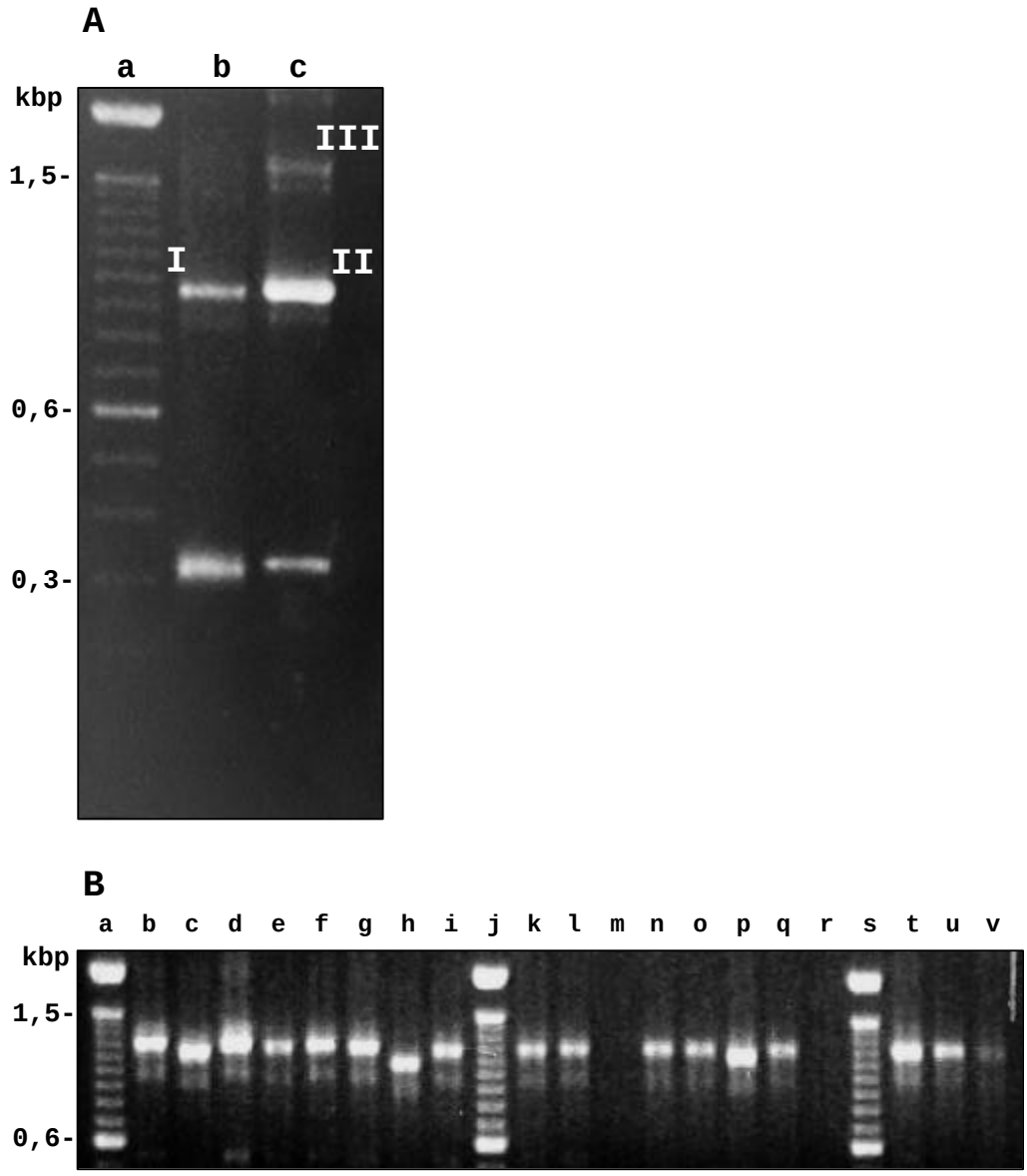


Figure 28: **A**: PCR amplification on T. b. brucei genomic DNA with the different 5'-primers in association with the conserved 3'-primer H1orf3'. lane a: 100bp DNA ladder; lanes b: PCR amplification with primer 5'-MAKAS; lane c: with primer 5'-MNNTT. **I**, **II** and **III**: Purified bands for subcloning and sequencing. **B**: Control of insert size after subcloning of bands **I** and **II** (see above). Lanes a, j and s: 100bp DNA ladder; lanes b-l: subcloned products of band **I**; bands m-v: subcloned products of band **II**. Clones from lanes b, c, h, o, p and q were selected for sequencing.

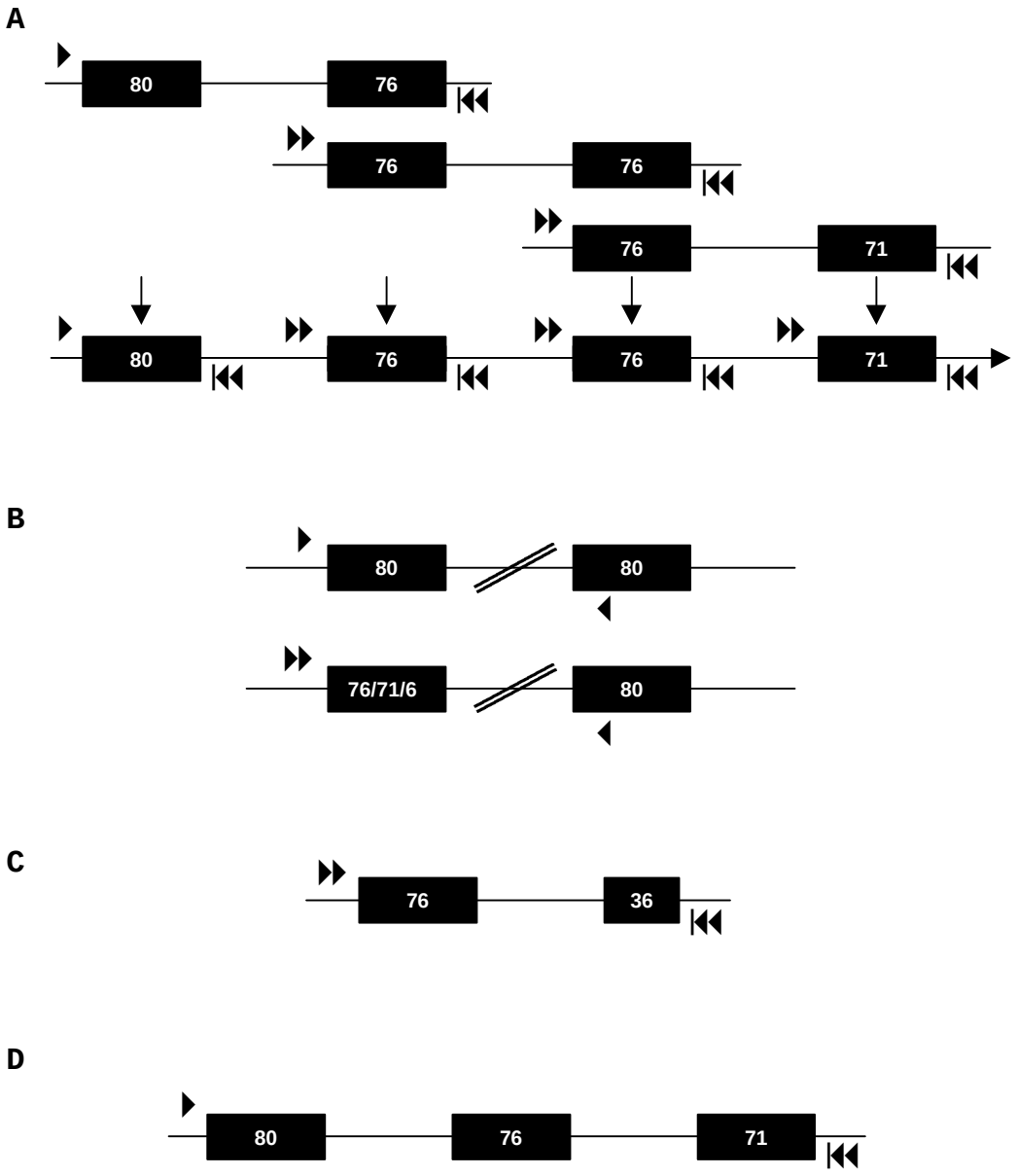


Figure 29: **A**: Schematic representation of three different genomic DNA tandems that were sequenced and hypothetical partial reconstruction of the H1 gene cluster. **B**: Amplifications that were never obtained. **C**: Isolation of a very short H1 gene. **D**: Representation of a triplet that was obtained. Black boxes represent ORFs; Number indicate length of coded variant in amino acids; ⏵: primer 5'-MNNTT; ⏩primer 5'-MAKAS; ⏮: primer H1orf3'; ⏴ : primer Walk5'-MNNTT; double bars: impossible amplification.

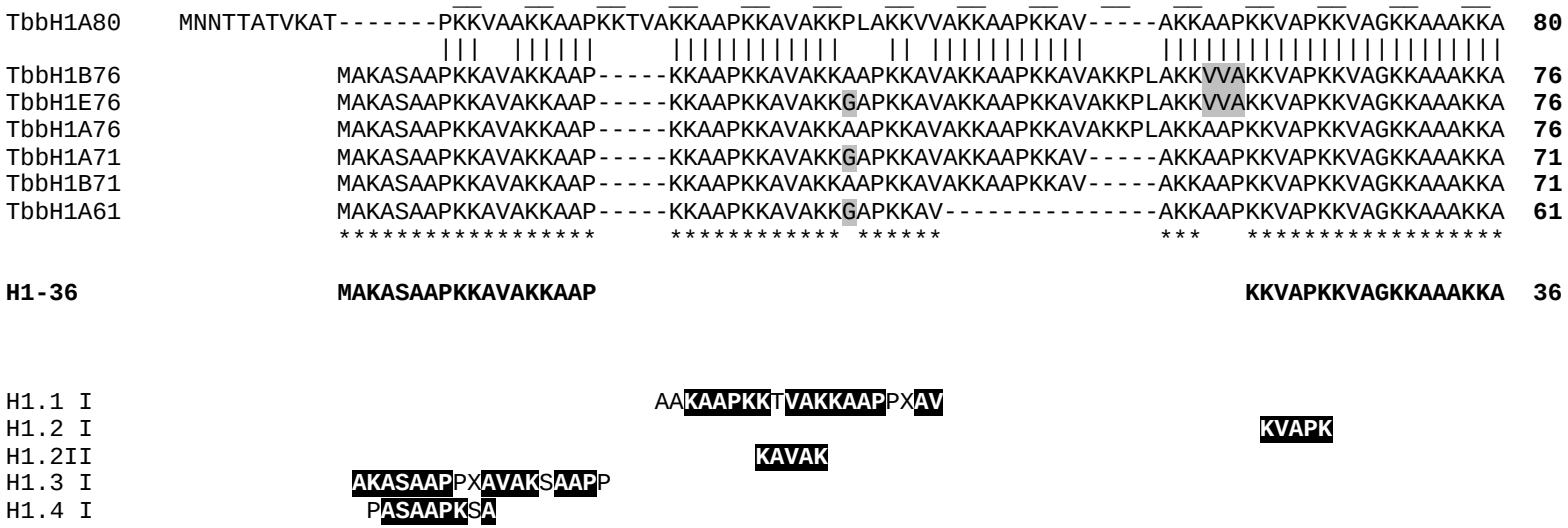


Figure 30: Recapitulative alignement of the different H1 variant obtained by *T. b. brucei* gDNA amplification with gene **H1-36** corresponding to the conserved amino acids found in all the variants amplified by primer 5'-MAKAS. |: matching positions between the two differently 5'-ended types of H1 variants; *: conserved positions inside the variant group obtained with primer 5'-MAKAS; ;grey boxes: substitutions; —: regular spacing of the lysine pairs; numbers indicate variant size in amino acids. Below: alignments of the five peptide fragments sequenced by Burri [Burri; 1995]; black boxes: matching positions.

A

```
ctttatcgactccccacaagaATGAACAACACAACCGCTACTGTGAAGGCCACCCCAAAG   60
                     M  N  N  T  T  A  T  V  K  A  T  P  K
AAGGTTGCAGCCAAAAAGGCCGCTCCAAAGAAAACTGTGGCCAAAAAGGCTGCTCCAAAG  120
 K  V  A  A  K  K  A  A  P  K  K  T  V  A  K  K  A  A  P  K
AAGGCTGTTGCCAAGAAACCTTTAGCTAAGAAGGTTGTCGCTAAGAAGGCTGCTCCCAAG  180
 K  A  V  A  K  K  P  L  A  K  K  V  V  A  K  K  A  A  P  K
AAAGCTGTCGCCAAGAAGGCCGCTCCCAAAAAGGTTGCCCCGAAGAAGGTTGCCGGCAAG  240
 K  A  V  A  K  K  A  A  P  K  K  V  A  P  K  K  V  A  G  K
AAGGCCGCCGCTAAGAAGGCGtgagcgcatccgctgctgcccgctattagacacgctatg  300
 K  A  A  A  K  K  A

aggtttacctgagtgtgggagaaagctgtcacacgtttcaggacgtcctcgtgcgtctct  360

                                ac  t       tt  t  tt-
ccaggacggagttagatttttcctatctttttgtttagttcccttctaccgtttttatt  420
                                ac  t       t  t  t catt

ggatatgtttcgtttgtgggttgcgtcttatgtaccgccatgcggtgttggtgtcgtagc  480

                                      ac
gttgcaaagagcatatcatcctgatgtgtggctattttaactgcctgtgtatggttgtgg  540
                                      ac
            a
tccaatagtattgctgtaggtctttttcctttccgcttatttcctccaaactcattgtat  600
            a             c

tctcatgttctttcacattcattttatttgctttcatttttctctttctccctgtggaa  660

gaagtcgcaatcttatcaacactcggaagtATGGCGAAGGCATCTGCTGCTCCCAAGAAA  720
                                M  A  K  A  S  A  A  P  K  K
GCTGTGGCCAAGAAGGCAGCCCCCAAAAAGGCTGCTCCAAAGAAGGCTGTCGCTAAGAAG  780
 A  V  A  K  K  A  A  P  K  K  A  A  P  K  K  A  V  A  K  K
GCTGCTCCAAAGAAGGCTGTCGCTAAGAAGGCTGCTCCCAAAAAGGCTGTTGCCAAGAAA  840
 A  A  P  K  K  A  V  A  K  K  A  A  P  K  K  A  V  A  K  K
CCTTTAGCCAAGAAGGCCGCTCCCAAAAAGGTTGCCCCCAAGAAGGTTGCCGGCAAGAAG  900
 P  L  A  K  K  A  A  P  K  K  V  A  P  K  K  V  A  G  K  K
GCCGCCGCTAAGAAGGCGtgagcgcatccgctgctgcccgctattagacacgctatgagg  960
 A  A  A  K  K  A
tttacctgagtg
```

B

```
                  UUUCC
                U       U
AGGACGGAGUUAGAU           A
||||  ||||||||||          |
UCUUCCCUUGAUUUG           U
                U       C
                  UUUUU
```

Figure 31: A: Sequence representation of a tandem displaying a 80-variant followed by a 76-variant. Capital lettres show the two ORFs ; bold capital letters show deduced amino acid sequences; bold lettres in the intergenic region locate the predicted stem-loop structure ; black box : polypirimidine tract ; grey boxes ; substitutions if compared to the intergenic region separating other variants: above for 76-71 variants and below for 76-76 variants ; singly underlined : primer 5'-MNNTT ; double underlined : primer 5'-MAKAS ; dotted underlined : primer H1orf3'. **B**: Representation of the theoretical stem-loop structure found between ORFs coding for 80 and 76 amino acid variant.

TRANSCRIPTION

Preliminary considerations:

The gene organization of histone H1 shows that, like numerous other gene systems described to this date in trypanosomes, the transcription is of polycistronic nature. It is then clear that different regulation mechanisms than those observed in higher eucaryotes are used. In *T. b. brucei*, two-dimensional separation of histone H1 from the 2 life stages of the parasite showed that one variant is over expressed in bloodstream forms while in procyclic culture forms, all 4 variants are present in equivalent amounts [Burri, 94]. However, from the obtained genomic data, it was not possible to discriminate between the four size variants. The only gene that exhibits some specific features is the one coding for the 80 amino acid variant. This gene is located in head position in the cluster, it is preceded by a different non-coding sequence, it bears a different N-terminal amino acid end and it is followed by an IR that contains a region that could form a potential stem-loop structure. All other variants are too similar in front, inside or even behind their ORFs so that it appears difficult to imagine that they can be differentially regulated. Additional information on H1 gene transcription and mRNA processing was then to be investigated.

Results:

In order to verify trans-splicing that coud neither be identified on *T. b. gambiense* cDNA clones, nor in *T. cruzi* by cDNA screening [Åslund, 94] and to target the used site, the conserved sequence patterns observed in the H1 3'-UTR were used to design a "selective RT-PCR" transcript isolation. Using primer Tryp-SL (**Table 1**) located on the spliced leader in association with the H1-specific primer H1orf3' it was possible to selectively amplify transcripts coding for histone H1. The splicing site was expected to be located

close to the start codon since the sequence separating the ORFs from the polypyrimidine tract is very short (see Fig. 31). The PCR amplified band obtained with primer Tryp-SL actually appeared to be shifted upwards by less than 50bp if compared to the PCR amplifications obtained with primer 5'-MNNTT / 5'-MAKAS and H1orf3' (**Fig. 32; lane 5**). Splicing is known to occur on GA dinucleotides situated downwards the polypyrimidine tract [Huang, 91]. Three of such positions are found on all the IRs (**Fig. 33**).

After subcloning and sequencing of the PCR product obtained with primer Tryp-SL, all the obtained sequences exhibited the same splicing site only about thirty bases upstream the ATG codon (27 bases for 80 amino acid coding transcripts and 29 bases for the other variants) (**Fig. 33**). However, alternative splicing cannot be excluded since only a few clones were analyzed.

Several cDNA clones obtained from the *T. b. gambiense* cDNA library screening were polyadenylated and in *T. cruzi*, northern hybridization experiments showed that all histone H1 mRNAs were in the poly-A^+ fraction [Åslund, 94].

Alignment of the 3'-UTR of four *T. b. gambiense* polyadenylated cDNA clones with the IR found in *T. b. brucei* shows several different but close polyadenylation sites. The used sites are all located about 10 to 30 bases upstream the polypyrimidine tract, with the exception of cDNA clone c9.1 in which the site is located 138 bp upstream the polypyrimidine tract (**Fig. 34**).

Transcription was analyzed by northern blot. After having tested the integrity of total RNA by separation on agarose gels (**Fig. 35A; lane 1**), four probes containing the four size variant ORFs and one probe corresponding to alpha-tubulin for calibration were prepared and tested separately on blotted total RNA. The obtained signal was sharp and no cross hybridization was seen, allowing the simultaneous use of both H1 and tubulin probes (**Fig. 35; lanes 3-6**). RNA samples were free of genomic DNA contaminants as shown by digestion with RNase (**Fig. 35B**).

Using a mixture of the four biotinylated probes corresponding to the four size variant ORFs, northern blots comparing the global transcription rate of histone H1 genes between the two life stages of the parasite showed no significant difference (**Fig. 36**).

In order to analyze if differential transcription of the 4 size variants between the two life stages of the parasite can be demonstrated, each H1 probe was tested separately against

blots of four plasmid DNA containing the four size variant ORFs (**Fig. 37**). Unfortunately, the obtained hybridization patterns show several cross reactions. The probe corresponding to 80-variants also hybridizes the ORF of 76-variants and the probe of this latter cross reacts onto the ORF coding for 80-variants (**Fig. 37A** and **B**). The probes corresponding to the two shorter variants also cross hybridize to each other but not onto the two longer ones (**Fig. 37C** and **D**).

A biotinilated oligonucleotide (oligoRNA-80) (**Table 1**) corresponding to the different 5'-coding sequence of the 80-variant was then tested and showed good specificity for the 80-variant ORF (**Fig. 37E**). Because of the weak signal obtained with this kind of probes, carrying only one biotine, the hybridized membranes had to be exposed for a prolonged time which results in a strong background. The more or less diffuse H1 transcript spots that could be obtained clearly showed that 80-variants are either equally transcribed in both stages of the parasite or eventually less transcribed in bloodstream forms (**Fig. 38A**).
Finally, the attempt was made to prepare a probe by using the PCR amplified fragment obtained with primer Tryp-SL and Walk-80 (**Table 1**). However, the use of this probe gave no signal and only the tubulin could be seen on autoradiographs. By exposing the membrane for a prolonged period of time, a smear of aspecific hybridization appears which is due either to the tubulin probe or to the H1 probe, but no additional marked spot was obtained (**Fig. 38B**). This is probably resulting from the very short fragment used as template for probe preparation (approximately 70 bp). The use of a random primer procedure for labeling appears not to be indicated here, this kind of labeling being normally used for probes that are a few hundreds base pairs long.

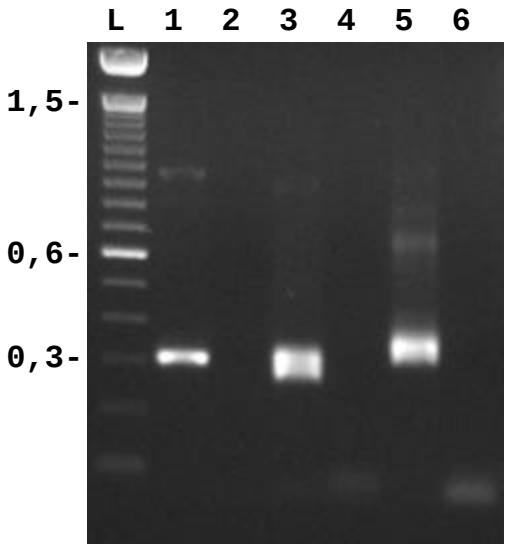


Figure 32: Evidence of near splicing site. gDNA of *T. b. brucei* was amplified with the conserved primer H1orf3' in combination with primer 5'-MNNTT (lane 1), with primer 5'-MAKAS (lane 3) and with primer Tryp-SL (lane 5) corresponding to part of the spliced leader. L: 100bp DNA ladder; lanes 2, 4 and 6, respective negative controls.

```
5'-80                                          GaaagctctttatcgactccccacaagaATG

                                        (-27)
          ************************************************************
80-2R     CGCTATTATTAGAACAGTTTCTGTACTATATTGaaagctctttatcgactccccacaagaATG
80-11U    CGCTATTATTAGAACAGTTTCTGTACTATATTGaaagctctttatcgactccccacaagaATG
80-15U    CGCTATTATTAGAACAGTTTCTGTACTATATTGaaagctctttatcgactccccacaagaATG
80-14U    CGCTATTATTAGAACAGTTTCTGTACTATATTGaaagctctttatcgactccccacaagaATG
          ||||||||||||||||||||||||||||||||
SL      AACGCTATTATTAGAACAGTTTCTGTACTATATTG
          ||||||||||||||||||||||||||||||||
76-5U     CGCTATTATTAGAACAGTTTCTGTACTATATTGaagtcgcaatcttatcaacactcggaagtATG
76-10U    CGCTATTATTAGAACAGTTTCTGTACTATATTGaagtcgcaatcttatcaacactcggaagtATG
76-4U     ----------AGAACAGTTTCTGTACTATATTGaagtcgcaatcttatcaacactcggaagtATG
71-1R     CGCTATTATTAGAACAGTTTCTGTACTATATTGaagtcgcaatcttatcaacactcggaagtATG
71-6R     CGCTATTATTAGAACAGTTTCTGTACTATATTGaagtcgcaatcttatcaacactcggaagtATG
          ************************************************************
                                        (-29)

5'-other  ctttcatttttctctttctccctgtggaagaagtcgcaatcttatcaacactcggaagtATG
```

Figure 33: Alignment of the 5'-ends of several RT-PCR clones showing the conserved splicing site among H1 variants of *T. b. brucei* bearing the same 5'-ends. SL: spliced leader with primer Tryp-SL in bold letters; |: spliced leader presence on all cDNA clones; *: sequence conservation among clones of the same type; (5'-80): sequence preceding 80-variants; (5'-other): sequence preceding all other variants; grey boxes: start codons; white boxes: potential splicing sites; (-27) and (-29): positions of observed splicing sites; black box: partial polypyrimidine tract.

```
Tbb80-76    TGAGCGCATCCGCTGCTGCCCGCTATTAGACACGCTATGAGGTTTACCTGAGTGTGGGAG
C9.1        TGAGCGCATCCGCTGCTGCCCGCTATTAGACAAGCTATGAGGTTTACCTGAGTGTGGGAG
c1.1        TGAGCGCATCCGCTGCTGCCCGCTATTAGACACGCTATGAGGTTTACCTGAGTGTGGGAG
c2.4        TGAGCGCATCCGCTGCTGCCCGCTATTAGACACGCTATGAGGTTTACCTGAGTGCGGGAG
c8.1        TGAGCGCATCCGCTGCTGCCCGCTATTAGACACGCTATGAGGTTTACCTGAGTGTGGGAG
            ******************************** ******************** *****

Tbb80-76    AAAGCTGTCACACGTTTCAGGACGTCCTCGTGCGTCTCTCCAGGACGGAGTTAGATTTTT
c9.1        AGAGCTGTCACACGTTTCAGGACGTCCTCGTGCGTCCCTCCAGGACGGAGTTAGATTTTT
c1.1        AGAGCTGTCACACGTTTCAGGACGTCCTCGTGCGTCCCTCCAGGACGGAGTTAGATTTTT
c2.4        AGAGCTGTCACACGTTTCAGGACGTCCTCGTGCGTCCCTCCAGGACGGAGTTAGATTTTT
c8.1        AGAGCTGTCACACGTTTCAGGACGTCCTCGTGCGTCCCTCCAGGACGGAGTTAGATTTTT
            * ********************************** **********************

Tbb80-76    CCTATCTTTTTTGTTTAGTTCCCTTCTACCGTTTTTATTGGATATGTTTCGTTTGTGGGT
c9.1        CCTATCTTTTTTGTTTAGTTCCCTTCTACCGTTTGTATTGGAAAAAAAAAA----------
c1.1        CCTATCTTACTTGTTTAGTTCCCTTCTACCGTTTGTATTGGATATGTTTCGTTTGTGGGT
c2.4        CTTATCTTACTTTTTTATTTCTTTTTCACTGATTTTATTGGATATGTTTCGTTTGAGGGT
c8.1        CTTATCTTACTTTTTTATTTCTTTTTCACTGATTTTATTGGATATGTTTCGTTTGAGGGT
            * ******  ** **** ***  **  ** * ** ******* *

Tbb80-76    TGCGTCTTATGTACCGCCATGCGGTGTTGGTGTCGTAGCGTTGCAAAGAGCATATCATCC
c9.1        ------------------------------------------------------------
c1.1        TGCGTCTTATGTAC-GCCATGCGGTGTTGGTGTCGTAGCGTTACAAAGAGCATATCATCC
c2.4        TGCGTCTTATGTACCGCCATGCGGTGTTGGTGTCGTAGCGTTACAAAGAGCATATCATCC
c8.1        TGCGTCTTATGTACCGCCATGCGGTGTTGGTGTCGTAGCGTTACAAAGAGCATATCATCC

Tbb80-76    TGATGTGTGGCTATTTTAACTGCCTGTGTATGGTTGTGGTCCAATAGTATTGCTGTAGGT
c9.1        ------------------------------------------------------------
c1.1        TGATGTGTGGCTATTTTAACTGCCTGTGTAAAAAAAAAAAAAAAAAA-------------
c2.4        TGATGTGTGGCTATTTTAACTGCCTGTGTATGGTTGTGTTCCAATAAAAAAAAAAAAA--
c8.1        TGATGTGTGGCTATTTTAACTGCCTGTGTATGGTTGTGTTCCAAAAAAAAAAA-------

Tbb80-76    CTTTTTCCTTTCCGCTTATTTCCTCCAAACTCATTGTATTCTCATGTTCTTTCACATTCA

Tbb80-76    TTTTATTTTGCTTTCATTTTTCTCTTTCTCCCTGTGGAAGAAGTCGCAATCTTATCAACA

Tbb80-76    CTCGGAAGTATG
```

Figure 34: Alignment of Intergenic region separating TbbH1A80 and TbbH1A76 in *T. b. brucei* against the 3' UTRs found on the different cDNA clones obtained from *T. b. gambiense*. **TGA**: stop codon of upstream ORF; **ATG**: start codon of downstream ORF; ***AA..AA***: poly-A tails; grey boxes: differences between the two subspecies; black box: polypyrimidine tract; *: matching positions.

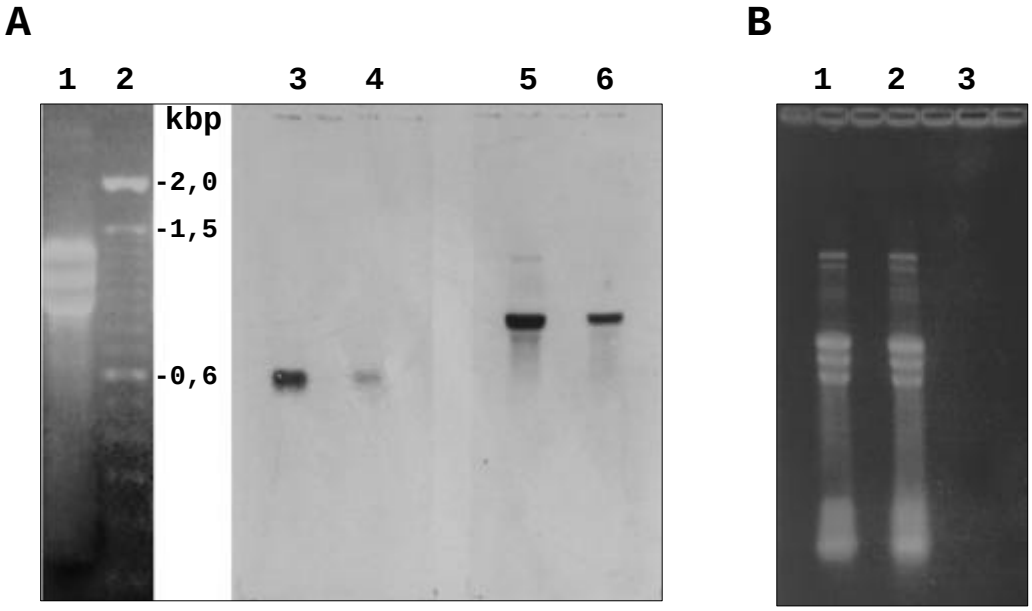


Figure 35: Specificity of H1 and alpha-tubulin probes on northern blots. **A**: Lane 1: total RNA separated on agarose gel; lane 2: 100bp DNA ladder; lane 3 to 6: northern blots of total RNA from *T. b. brucei* procyclic forms (lanes 3 and 5) and bloodstream forms (lanes 4 and 6); lanes 3 and 4: hybridization with a mixture of H1 probes corresponding to the different size-classes; lanes 5 and 6, hybridization against alpha-tubulin probe. **B**: Control of RNA purity on agarose gel by RNase digestion; lane 1: total RNA conserved at –80°C; lane 2: total RNA incubated at 37°C for 1 hour; lane 3: total RNA incubated at 37°C for 1 hour with RNase.

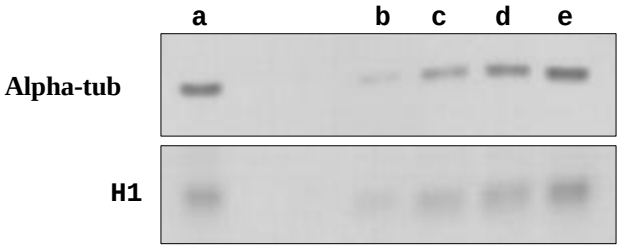


Figure 36: Analysis of global H1 transcription between the two parasitic life stages by total RNA northern blot. a: bloodstream forms; b-e: serial dilutions of procyclic forms RNA. H1: hybridization with a mixture of 4 DNA probes corresponding to the 4 H1 size classes; Alpha-tub: calibration against an alpha-tubulin specific DNA probe.

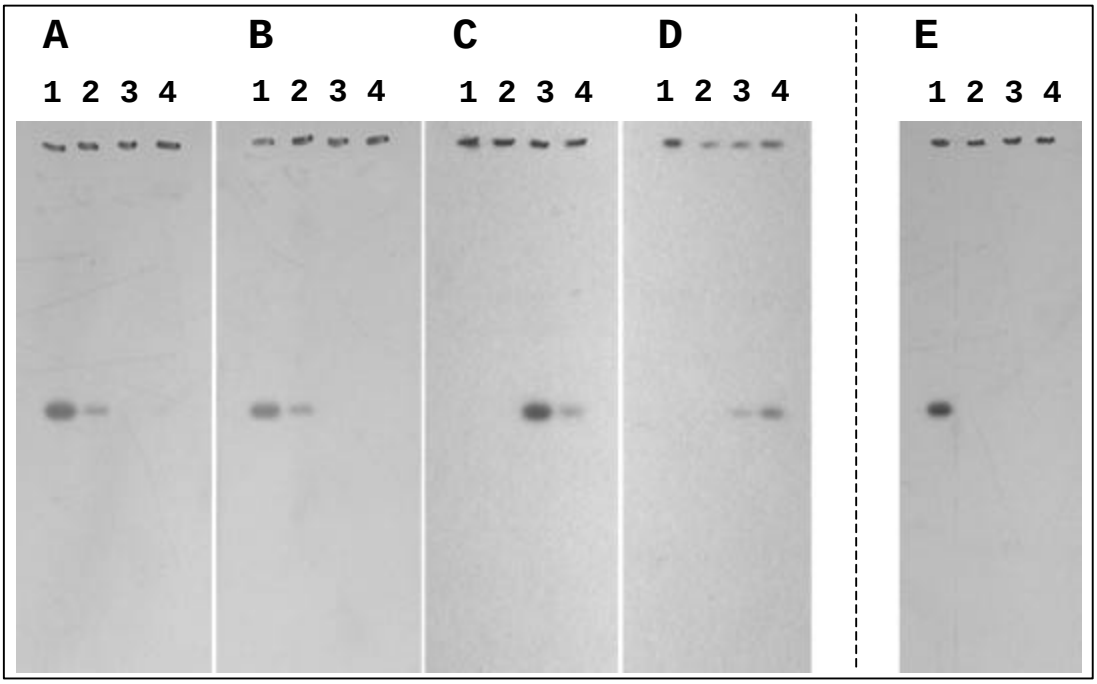


Figure 37: A, B, C and D: Respective hybridization patterns obtained by the 4 different DNA probes corresponding to the 4 H1 size classes. 1, 2, 3 and 4: southern blot of plasmids containing respectively an 80, 76, 71 and 61 variant coding insert. Testing of DNA probe corresponding to an 80-variant (**A**), to a 76-variant (**B**), to a 71-variant (**C**) and to a 61-variant (**D**). **E**: Testing of biotinylated oligonucleotide (oligoRNA-80) corresponding to the different 5' coding sequence of 80-variants.

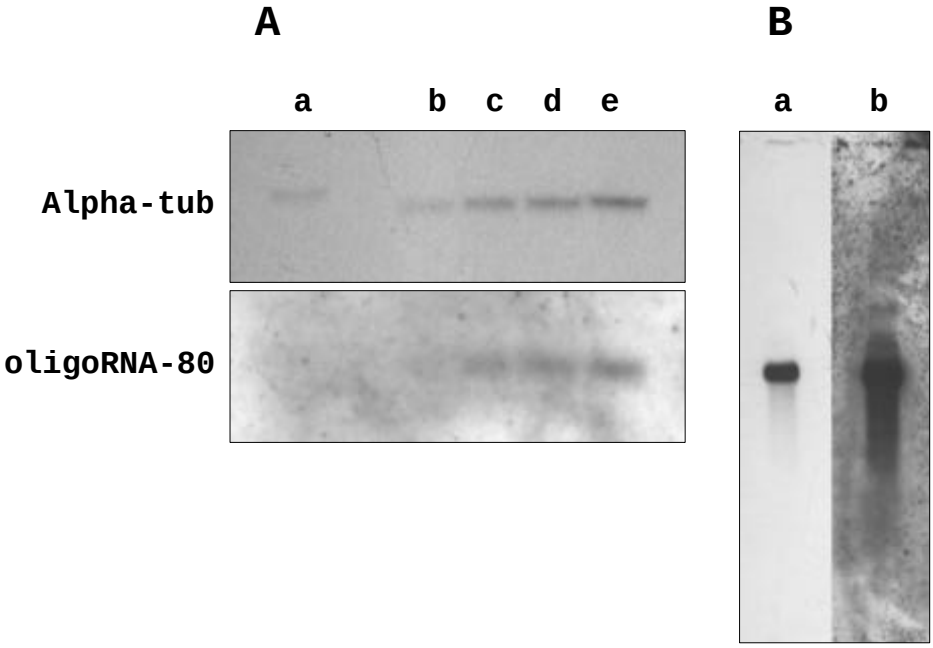


Figure 38: Analysis of 80-variant gene transcription between the two parasitic life stages by total RNA northern blot. **A**: Use of biotinylated oligonucleotide (oligoRNA-80); a: bloodstream forms; b-e: serial dilutions of procyclic forms RNA; Alpha-tub: calibration against an alpha-tubulin specific DNA probe. **B**: Concomitant hybridization with an alpha-tubulin specific DNA probe and an H1 probe prepared with PCR fragment obtained by primers Tryp-SL and Walk5'-MNNTT; lane a: short exposure; lane b: long exposure.

A NEWLY IDENTIFIED H1 GENE

Preliminary considerations:

The observed over-expression of one variant in bloodstream forms could not be pointed out on the level of transcription of the different variants constituting the H1 cluster. The strong sequence similarity between the different size variants does not allow differential probe design that can be used for transcription analysis by classical techniques. Even if the 80-variant coding gene exhibits some differential features regarding its cluster location, coding sequence or 3'-UTR, it does not appear to be differentially transcribed. Regulation probably takes place at the translational level, or results from other elements that were not yet known. Sheared DNA sequences from another *T. b. brucei* strain (GUTat10.1) recently added to the Parasite Genome Blast Server are showing strong homology to *T. b. brucei* H1 histones, with some of them presenting new characteristics. If these sequences are also present in the strain used as model in the present investigation, they may give new insights into the comprehension of H1 expression.

Results:

On the amino acid level, two sequences found on the Parasite Genome Blast Server match the 76 amino acid H1 proteins isolated from strain TREU 927/4 very well but they do also present two special features. First, they possess a pair of threonines at position 4 and 5 and second, they are preceded by a different 5' sequence, the 3' sequence being conserved (**Fig. 39**). In order to verify if these genes are also present in strain TREU 927/4, a new PCR reaction was performed with primer 5'-MAKTT/1 (**Table 1**) corresponding to this new 5'-end and the conserved H1orf3' primer. This amplification gave only rise to two sharp bands of approximately 300 and 900 bp and no multiple banding could be obtained as seen with the other primers 5'-MNNTT and 5'-MAKAS (**Fig 40A**). Southern blot analysis clearly shows that no hybridization signal was present above the 900 bp band and denotes the limited number of H1 genes in this array (**Fig 40B**).

Sequencing of the 300 bp band gave only 76-variant coding ORFs, and the new pattern of two threonines was always present. These new H1 variants do not increase the number of size classes seen in *T. b. brucei* but the yet complex H1 gene family appears still more difficult to analyze, with at least two additional variants differing from each others by 3 amino acids (**Fig. 41**). Analysis of the 900 bp band always revealed the same organization, with a 76-variant ORF coding for the two threonines followed by a 71-variant ORF similar to previously isolated genes of this size. Complete nucleotide sequences of two such tandems are presented in **Annex F**.

In order to investigate a potential association of these genes with the other H1 coding genes previously described, PCR test reactions associating primer 5'-MAKTT/1 with Walk5'-MNNTT, 5'-MNNTT or 5'-MAKAS were performed, but no amplification was obtained. Interestingly, the IR present on MAKTT tandems are of the same type than the IR following the ORFs of 80-variants (**Fig. 42**). In the IR following these new H1 variants, the theoretical stem-loop structure is also present (**Fig. 42**).

For clarity, the isolated series of *T. b. brucei* H1 variants can be grouped into four size classes with the 76 amino acid class showing the greatest heterogeneity. However, despite of this heterogeneity between and inside the H1 gene size classes, they do also present strong similarities and all variants can be organized into 3 groups regarding their 5' coding or non-coding sequences (**Fig. 43A**). The absence of close topographic relation between the newly isolated gene and the previous ones suggests that the different genes could be separated into two clusters. First there is the MNNTT group that constitutes the head of the first H1 cluster isolated and which is only constituted of 80 amino acid proteins. The second group is the MAKTT group which is probably the head of a newly isolated H1 cluster. This group is only constituted of variants of 76 amino acids. Finally, the MAKAS group contains proteins of 76, 71 and 61 amino acids (with the exception of the 36 amino acid variant that was found on one tandem; see "genome organization"). Their coding sequences are downstream the ORFs of the two other groups (**Fig. 43B**).

In order to verify if the over-expressed variant in bloodstream forms was linked to the (in)activation of a cluster specifically containing the newly isolated MAKTT variant, the presence/absence of transcripts corresponding to the three groups of genes, considering their 5' surroundings were analyzed in both stages.

Both procyclic and bloodstream forms template RNA stocks were treated with DNaseI before further experimentation. After 1 hour incubation, DNaseI was heat inactivated and control of RNA purity was made by performing a PCR reaction associating primer H1orf3' and 5'-MNNTT. No amplification was obtained (**Fig. 44A; lanes b-d**).

Reverse transcription was carried out with an oligo-dT and the products of reverse transcription were used as templates for PCR amplification with the three different 5'-primers. All three groups of H1 genes were well amplified on RT-products of both stages of the parasite life-cycle (**Fig. 44A; lanes f-q**). Note that primer MAKTT/1 (**Fig. 44A; lanes l-n**) gave no amplification, probably because it is located over the splicing site. A second primer, MAKTT/2 (**Table 1**) was then constructed with its 5' end on the next GA dinucleotide (**Fig. 44B**). This primer gave a good amplification (**Fig. 44A, lanes o-q**) and sets the splicing site for this type of genes 23 bases upstream the start codon.

In order to make sure that no cross amplification was present between the three groups of genes with their respective primers, a control PCR on plasmid DNA containing the 3 groups of genes was made with each 5'-primer. Absolutely no cross amplification can be obtained (**Fig. 45**).

All variants are transcribed in both stages of the life cycle but the amount of transcription could vary and would need additional quantitative analysis.

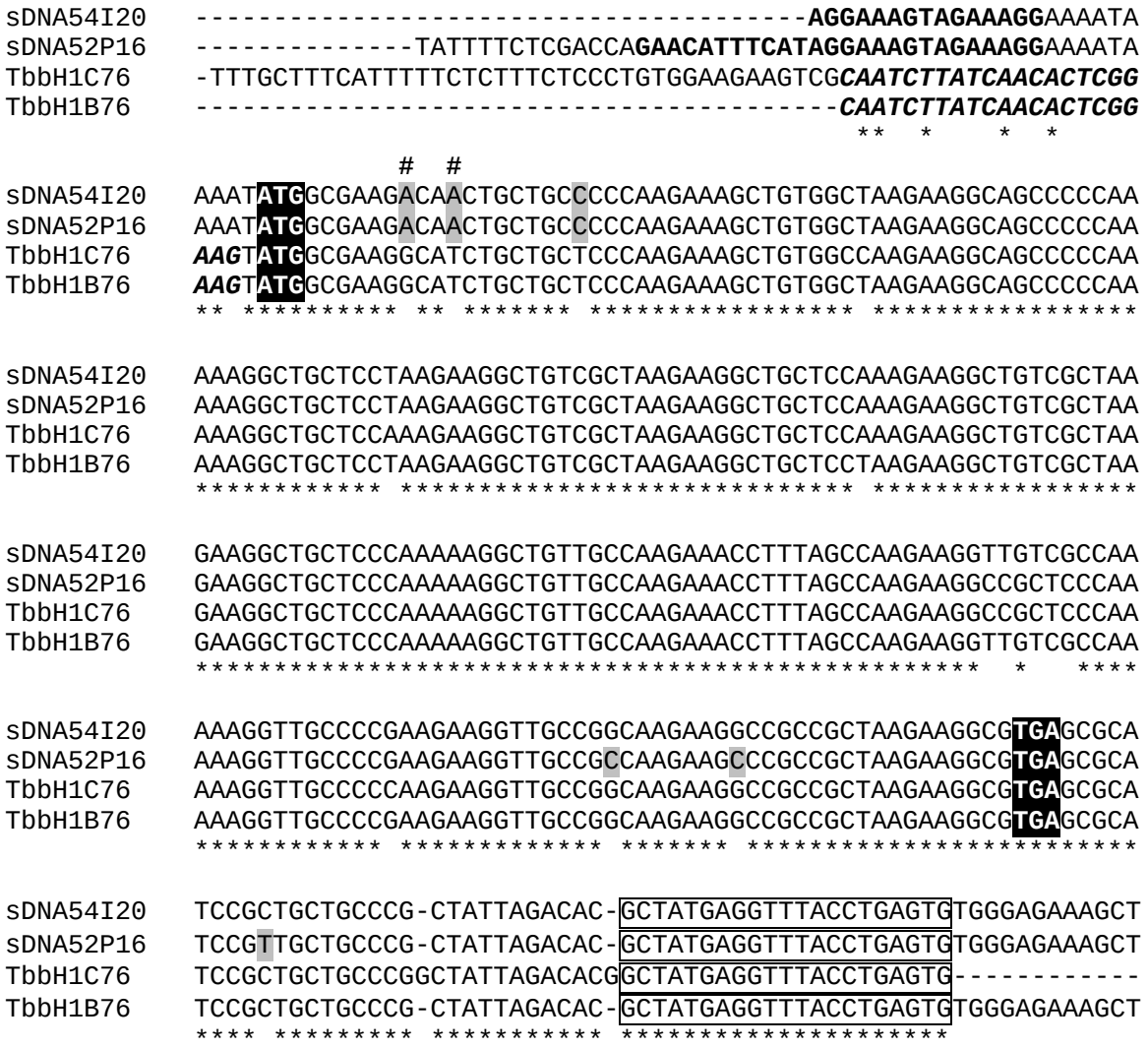


Figure 39: Alignment of two 76-variants ORFs and two recent Genbank entries corresponding to *T. brucei* GUTat 10.1 clones of a shared DNA library, (Genbank Acc. Numbers: sDNA54I20: AQ942626; sDNA52P16: AQ941417). Black boxes: start and stop codons; bold letters: new primer 5'-MAKTT/1; bold italics: primer 5'-MAKAS; open boxes: primer H1orf3'; grey boxes: positions from the 2 shared DNA ORFs differing from already known H1 genes; #: Substitutions responsible for the threonine pair; *: matching positions among all 4 sequences.

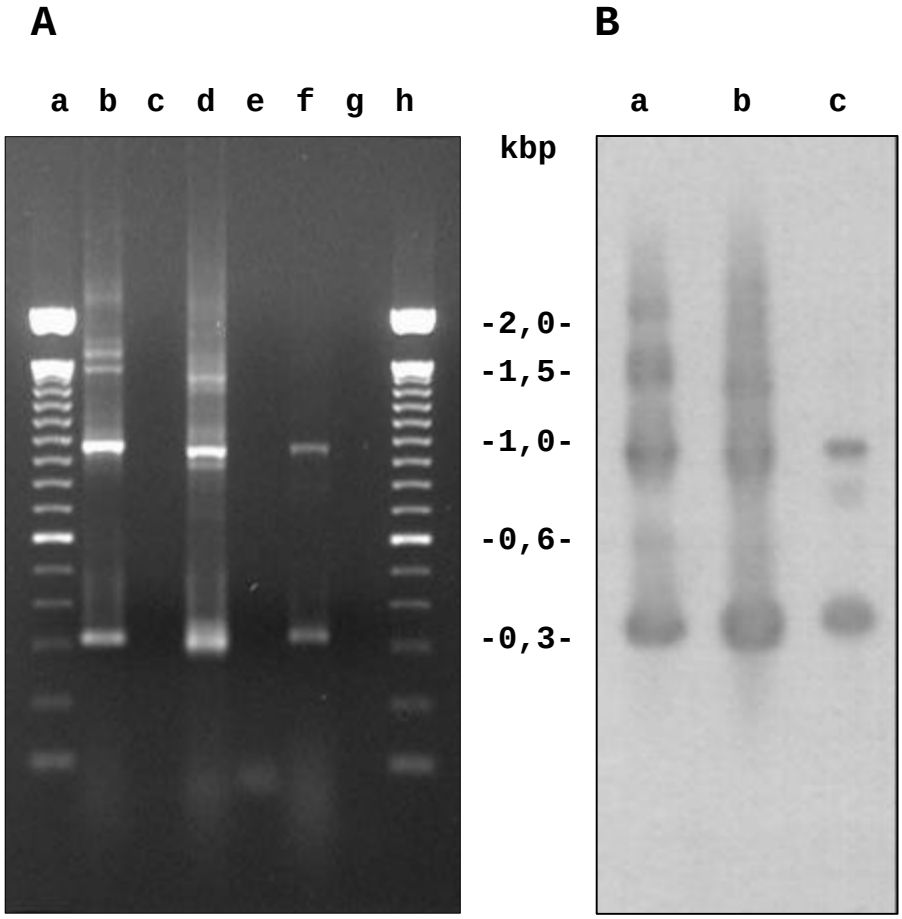


Figure 40: A: PCR amplification on genomic DNA of *T. b. brucei* with the different 5'-primers and primer H1orf3'. Lane a and h: 100bp DNA ladder; lane b: amplification with primer 5'-MNNTT; lane d: with primer 5'-MAKAS; lane f; with primer 5'-MAKTT/1; lanes c, e and g: respective negative controls. **B**: Southern blot hybridization of the same gel. Lane a: amplification with primer 5'-MNNTT; lane b: with primer 5'-MAKAS; lane c; with primer 5'-MAKTT/1.

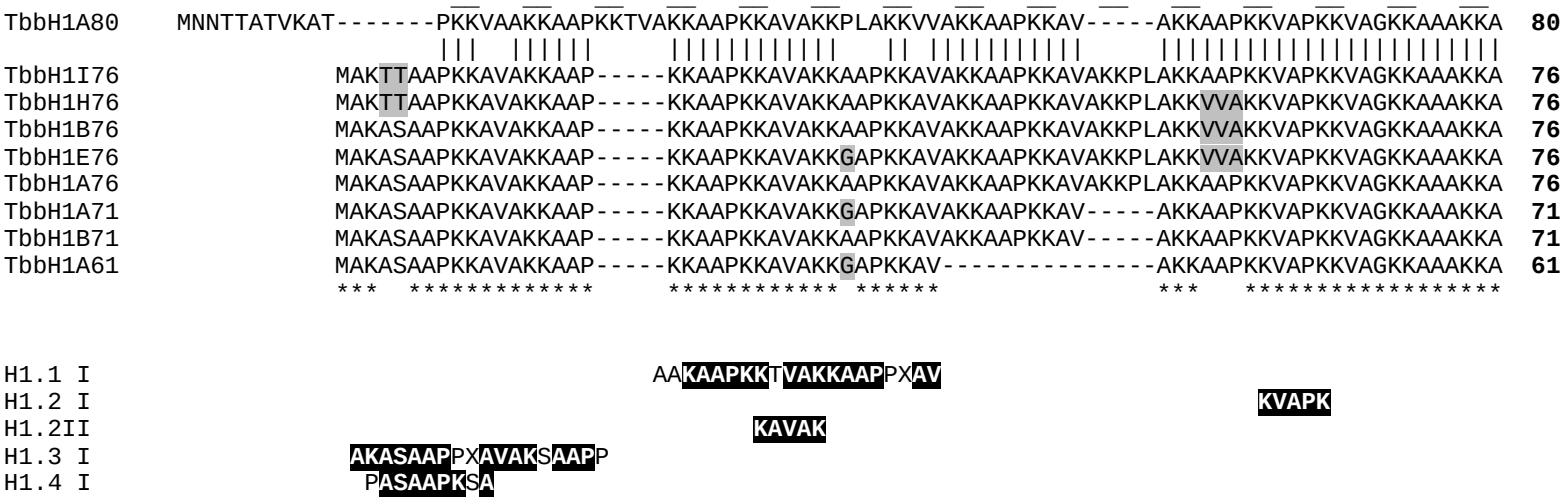


Figure 41: Recapitulative alignment of all isolated H1 variants with the two additional variants presenting the threonine pair. |: matching positions between the two differently 5'-ended types of H1 variants; *: conserved positions inside the variant group obtained with primer 5'-MAKAS and 5'-MAKTT/1; grey boxes: substitutions; —: regular spacing of the lysine pairs; numbers indicate variant size in amino acids. Below: alignments of the five peptide fragments sequenced by Burri [Burri; 95]; black boxes: matching positions.

```
Tbb80-76    TGAGCGCATCCGCTGCTGCCCGCTATTAGACACGCTATGAGGTTTACCTGAGTGTGGGAG
IR-MAKTT    TGAGCGCATCCGCTGCTGCCCGCTATTAGACACGCTATGAGGTTTACCTGAGTGTGGGAG
Tbb76-76    TGAGCGCATCCGCTGCTGCCCGCTATTAGACACGCTATGAGGTTTACCTGAGTGTGGGAG
            ************************************************************

Tbb80-76    AAAGCTGTCACACGTTTCAGGACGTCCTCGTGCGTCTCTCCAGGACGGAGTTAGATTTTT
IR-MAKTT    AAAGCTGTCACACGTTTCAGGACGTCCTCGTGCGTCTCTCCAGGACGGAGTTAGATTTTT
Tbb76-76    AAAGCTGTCACACGTTTCAGGACGTCCTCGTGCGTCTCTCCAGGACGGAGTTAGATTTTT
            ************************************************************

Tbb80-76    CCTATCTTTTTTGTTTAGTTCCCTTCTACCGTTTTTATTGGATATGTTTCGTTTGTGGGT
IR-MAKTT    CCTATCTTTTTTGTTTAGTTCCCTTCTACCGTTTTTATTGGATAT-TTTCATTTGTGGGT
Tbb76-76    CCTATCTTACTTTTTTATTTCTTTTTCATT-TTTTTATTGGATATGTTTCGTTTGTGGGT
            ********  ** **** ***  **  *   ************** **** *********

Tbb80-76    TGCGTCTTATGTACCGCCATGCGGTGTTGGTGTCGTAGCGTTGCAAAGAGCATATCATCC
IR-MAKTT    TGCGTCTTATGTACCGCCATGCGGTGTTGGTGTCGTAGCGTTGCAAAGAGCATATCATCC
Tbb76-76    TGCGTCTTATGTACCGCCATGCGGTGTTGGTGTCGTAGCGTTGCAAAGAGCATATCATCC
            ************************************************************

Tbb80-76    TGATGTGTGGCTATTTTAACTGCCTGTGTATGGTTGTGGTCCAATAGTATTGCTGTAGGT
IR-MAKTT    TGATGTGTGGCTATTTTAACTGCCTGTGTATGGTTGTGGTCCAATAGTATTGCTGTAGGT
Tbb76-76    TGATGTGTGGCTATACTAACTGCCTGTGTATGGTTGTGGTCCAATAGTATTACTGTAGGT
            **************  ********************************** ********

Tbb80-76    CTTTTTCCTTTCCGCTTATTTCCTCCAAACTCATTGTATTCTCATGTTCTTTCACATTCA
IR-MAKTT    CTTTTTCCTTTCCGCTTATTTCCTCCAAACTCATTGTATTCTCATGTTCTTTCACATTCA
Tbb76-76    CTTTTCCCTTTCCGCTTATTTCCTCCAAACTCATTGTATTCTCATGTTCTTTCACATTCA
            ***** ********************************************************

Tbb80-76    TTTTATTTTGCTTTCATTTTTCTCTTTCTCCCTGTGGAAGAAGTCGCAATCTTATCAACA
IR-MAKTT    TTTTATTTTGCTTTCATTTTTCTCTTTCTCCCTGTGGAAGAAGTCGCAATCTTATCAACA
Tbb76-76    TTTTATTTTGCTTTCATTTTTCTCTTTCTCCCTGTGGAAGAAGTCGCAATCTTATCA-CA
            ************************************************************ **

Tbb80-76    CTCGGAAGTATG
IR-MAKTT    CTCGGAAGTATG
Tbb76-76    CTCGGAAGTATG
            ************
```

Figure 42: Alignment of the intergenic region (IR-MAKTT) following the H1 gene amplified with primer 5'-MAKTT/1 on genomic DNA of *T.b.brucei* (TREU927/4) with the IR separating an 80-variant from a 76-variant (above) and the IR separating two 76-variants. Black boxes: stop and start codons; grey boxes: positions indicating stronger resemblance with the IR following the 80-variant; open boxes: stem of the theoretical stem loop structure described earlier; *: matching positions among the 3 sequences.

A

```
                ctttatcgactccccacaagAATGAACAACACAACCGCTACTGTG
                                    M  N  N  T  T

GTGGAAGAAGTCGcaatcttatcaacactcggaagTATGGCGAAGGCATCTGCTGCTCCC
                                    M  A  K  A  S

gaacatttcataggaaagtagaaaggAAAATAAAATATGGCGAAGACAACTGCTGCCCCC
                                    M  A  K  T  T
```

B

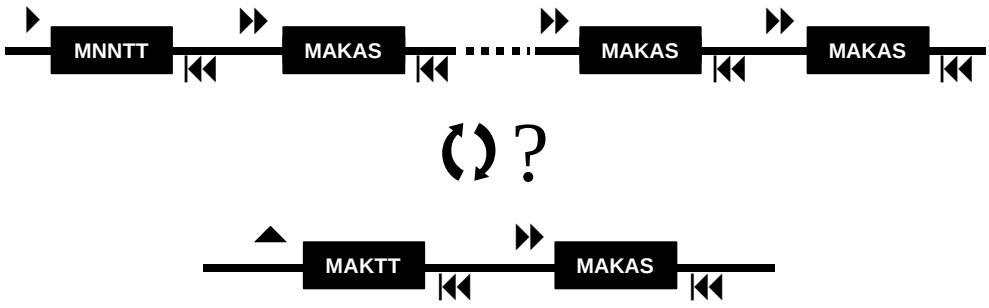
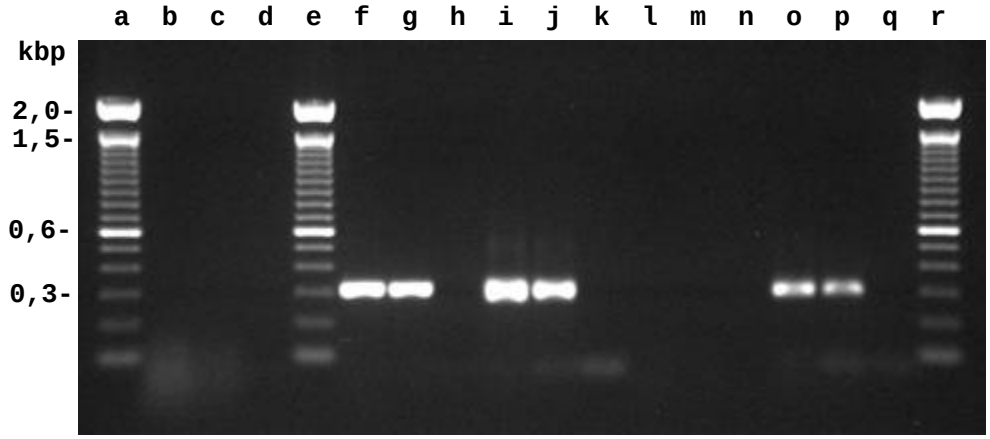


Figure 43: A: Representation of the 3 different 5' coding and non-coding H1 gene sequences. Black boxes: start codons; capital and bold letters: amino acid sequence of N-terminal end; single line: primer 5'MNNTT; double line: primer 5'MAKAS; open box: primer 5'-MAKTT/1. **B**: Hypothetical representation of the 2 H1 gene clusters. ▸: primer 5'-MNNTT; ▸▸primer 5'-MAKAS; ⏮: primer H1orf3'; ▲: primer 5'-MAKTT/1. Relation between these two sets of genes is unknown.

A



B

Figure 44: A: Control of transcription of the three groups of genes in both stages of the life cycle. Lanes a, e and r: 100bp DNA ladder; lanes b to d: control PCR with primer 5'-MNNTT and H1orf3' after DNase treatment of total RNA of procyclic forms (lane b) and bloodstream forms (lane c) extracted by Trizol; PCR negative control (lane d). Lanes f, i, l, o: amplification on procyclic RT-PCR product; lanes g, j, m, p: amplification on bloodstream forms RT-PCR products; lanes h, k, n, q: respective negative controls. Sets of primers are: 5'-MNNTT / H1orf3' (lanes f-h); 5'-MAKAS / H1orf3' (lanes i-k); 5'-MAKTT/1 / H1orf3' (lanes l-n); 5'-MAKTT/2 / H1orf3' (lanes o-q). **B**: Presence of the potential splicing sites in front of the ORF coding for the MAKTT variant. Black box: start codon; grey boxes, potential splicing sites; open box: primer 5'-MAKTT/1; *: primer 5'-MAKTT/2.

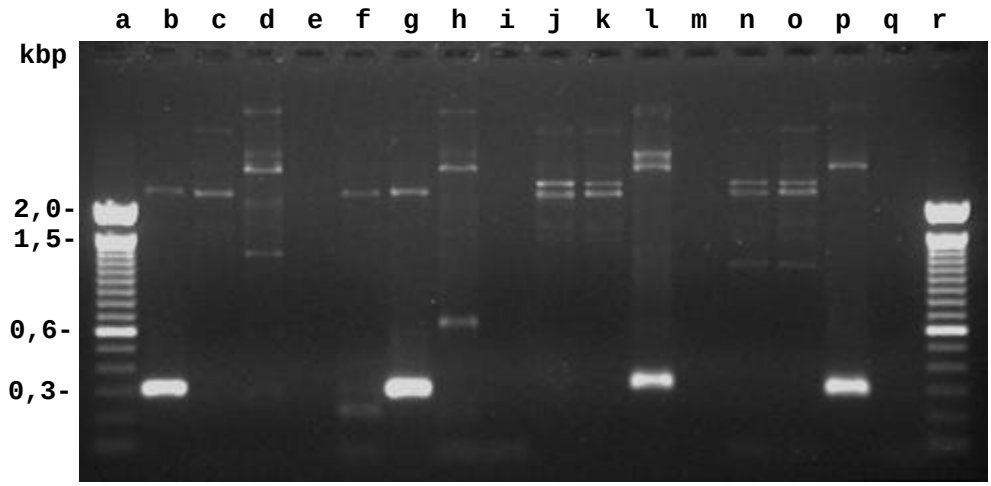


Figure 45: Control of potential cross amplification of the different H1 ORFs by the use of the different 5'-primer in association with primer H1orf3'. Lanes a and r: 100bp DNA ladder; lanes b, f, j, n: bluescript plasmid containing an "MNNTT" insert; lanes c, g, k, o: bluescript plasmid containing a "MAKAS" insert; lanes d, h, l, p: TOPO-TA cloning vector containing a "MAKTT" insert; lanes e, i, m, q: respective negative control. PCR amplifications were as follow: lanes b to e: primer 5'-MNNTT; lanes f to i: primer 5'-MAKAS; lanes j to m: primer 5'-MAKTT/1 and lanes n to q: primer 5'-MAKTT/2.

ANTIBODIES AND IMMUNOLOCALISATION

Preliminary considerations:

In order to analyze histone H1 functionality in African trypanosomes, the need of an antibody appeared necessary. Since no real chromatin compaction can be seen in any stage of the parasitic life cycle or cell cycle, the question why trypanosomes spend energy to express H1 histones remains open. The idea that, beside the chromatin location and its ability to influence chromatin structure, H1 could display additional function(s) in the cell became worth to be investigated. Gene regulation could be, directly and/or indirectly influenced even by blocking access to regulatory elements, competing with other proteins implicated in activation or participating in three dimensional nuclear chromatin structure. The genbank sequence described as a microtubule associated protein which corresponds to H1 (see "Genomic H1 gene isolation") as well as studies tending to show affinity of H1 for microtubules [Multigner, 92] that are structurally such abundant and important in trypanosomes [Kohl, 98] could also let think that trypanosome H1 could play original function in other cellular mechanisms, perhaps outside the nucleus.

Results:

Histone H1 of *T. b. brucei* procyclic forms isolated by a combination of cavitation and perchloric acid extraction were first separated on normal 15% SDS-PAGE gels in order to monitor their migration pattern on these type of gels. Actually, the particular biochemical properties of these small and truncated H1 don't allow to expect them to migrate normally according to their theoretical molecular weight and special histone markers or histone references should be used. The deduced amino acid sequences of the four major H1 variants isolated on the gene level display theoretical molecular weights of respectively 8100, 7550, 7000 and 6000 kDa. When perchloric acid extracted protein fractions are separated on normal SDS-PAGE gels, only three strong bands can be visualized (**Fig. 46A**). Some very faint and quite invisible additional bands that are probably co-extracted with H1 by this procedure are also seen. The three bands show an apparent migration that greatly diverges from the theoretical molecular weights mentioned above since they locate in immediate proximity of the 14 kDa marker band. The fact that only three bands can be

seen probably comes from co-migration of the two intermediate variants. Supporting this explanation, it was shown that in procyclic forms the four variants were expressed in equal amounts and, in addition, co-migration was already described even on gels that allow much greater separation [Schlimme, 93].

For antibody production, a MNNTT-80 and a MAKAS-76 variants were selected. Preliminary experiments in which the two ORFs of the selected variants were subcloned into the PET9a expression system did not allow to obtain satisfying purification of the proteins [Hirschy, diploma work, 98].
The methodology was then adapted in order to obtain sufficient and pure enough H1 for immunization. The induction was enhanced by applying IPTG only after the culture reached an optical density of 1,0 at 600nm. The standard concentrations of IPTG was doubled and induction shortened to only one hour. After induction, cells were lysed by 5 to 7 smooth sonication pulses of 1 min at very low amplitude. This way, induction was very high and H1 was present in the supernatant after centrifugation of the cell lysate (**Fig. 46B**). The obtained soluble protein mixture was then loaded onto 15% SDS-PAGE gels and the induced band cut out for purification through Elutrap system (Schleicher & Schuell). The obtained native proteins should migrate to the same position as the perchloric acid extracted proteins. Parallel migration of the two recombinant variants and of the perchloric extracted H1 showed very good alignment, the 80 variant migrating beside the upper band and the 76 variant beside the middle band of the total H1 extract (**Fig. 46C**).

After separate immunization of the two variants in mice, serums were collected and tested on western blots of induced cultures expressing the two variants and also a culture of the used bacterial strain transformed with the plasmid without insert (**Fig. 47**). Unfortunately, both serums recognized both variants and therefore, no specific signal for one or the other H1 could be obtained (**Fig. 47B and C**). The faint bands around the H1 band can easily be explained by the fact that cutting of the H1 band from the gel cannot be exclusive, causing co-immunization of some bacterial proteins migrating in the same region.
Testing of the serums on perchloric extracted protein blots also revealed the same recognition pattern, the three H1 bands being highlighted in both case (**Fig. 48A**). Interestingly, the serum appeared to be very specific for H1 since absolutely no other signal is seen when testing them onto total trypanosome crude extracts (**Fig. 48B**). The

serum also recognizes calf thymus H1, and most probably the tails of this more complex and classic histone H1 (**Fig. 48C**).

In order to try to discriminate among the various H1, a synthetic peptide was used. The only region that appeared to be exploitable was the N-terminal region of the MNNTT-80 variants. It is the only amino acid region which clearly denotes from the rest of this gene family products and which should be long enough to allow immune response. A synthetic peptide corresponding to the 12 first amino acids of this variant was coupled to a helper epitope (P30). Immunization followed the same periodicity as for the whole proteins, with injections of 50 or 150 µg. The response was monitored from the fourth injection and then until signal was obtained, at each boost. The response became perceptible only after 7 boosts in the mice that were injected with 150 µg. As in the case of the serums described above, the peptide anti-serums (PEPserum) were tested against western blots of induced bacterial cultures and also against a western blot of the peptide itself (**Fig 49A**). The PEPserum clearly shows a band of 15 kDa (**Fig. 49B; lane b**) corresponding to the MNNTT-80 variant in the induced culture that doesn't appear in the MAKAS-76 variant induction (**Fig. 49B; lane d**). All other bands are visible in the cultures with or without induced H1 (**Fig. 49B; lanes d and f**) and are even highlighted with the naive serum from day zero (**Fig. 49B; lanes g-i**). However, no reaction could be detected neither on perchloric acid extracted H1, nor on total trypanosome extracts (data not shown).
In order to verify if phosphorylation plays a role in this absence of response, perchloric acid extracted H1 were treated with alkaline phosphatase for 1, 3 and 9 hours and then blotted onto cellulose membranes that were subsequently tested against the PEPserum. In parallel, the samples were tested against a mixture of the serums obtained with whole proteins as a positive control (**Fig. 50; lanes a, c, e, g)**). No signal was obtained with PEPserum even by the use of softer experimental procedure based on membrane blocking with dry milk in PBS (**Fig. 50; lanes b, d, f, h**), indicating that phosphorylation is not involved. A positive control was performed on the peptide itself (**Fig. 50; lane i**).

Localization of histone H1 was investigated by confocal microscopy. Several cell preparation procedures were tested, but only fixation with paraformaldehyde and good permeation by Triton X-100 gave a result. Both stages of the parasitic life-cycle where investigated, giving quite similar results (**Fig. 51 and 52**). No signal was seen outside the nucleus. H1 appears to be following the DNA repartition very well since superposition of

the DNA signal obtained with pico-green (green) and the H1 signal (red) always matched. DNA appears to be located in the periphery of the nucleus, leaving a dark hollow more or less in the middle of the nucleus which probably locates the large nucleolus. This observation is less evident when looking at bloodstream forms even if this structure is often also seen (**Fig. 52A**). Finally, histone H1 colocates with DNA along the whole cell-cycle and the entire division process can be followed (**Fig. 51A-E** and **52A-E**).

Control experiments with naive serum give absolutely no fluorescence (**Fig. 51F and 52F**). Control of overlapping light emission between the two filters used was made by cutting the green canal; no difference is notable in the red H1 signal (**Fig. 53; A and B**). An additional control was made by omitting the primary antibody, absolutely no signal remains in the red canal (**Fig 53: C and D**).

Observations of bloodstream forms first suggested a kinetoplast H1 signal, but the whole control procedure showed that the red spots sometimes obtained in this form of the parasite was resulting from the secondary antibody. Superposition of the red and green canals also shows that these spots always were juxtaposing the kinetoplast (**Fig. 53: E and F**). These spots are probably corresponding to paraflagellar structures and most probably to the poach. Since the bloodstream trypanosomes were raised in mice, it is possible that mouse blood remained in there and since the used secondary antibody was an anti-Mouse-IgG, this could explain this disturbing observation.

Finally, confocal analysis with PEPserum never gave any signal (not shown).

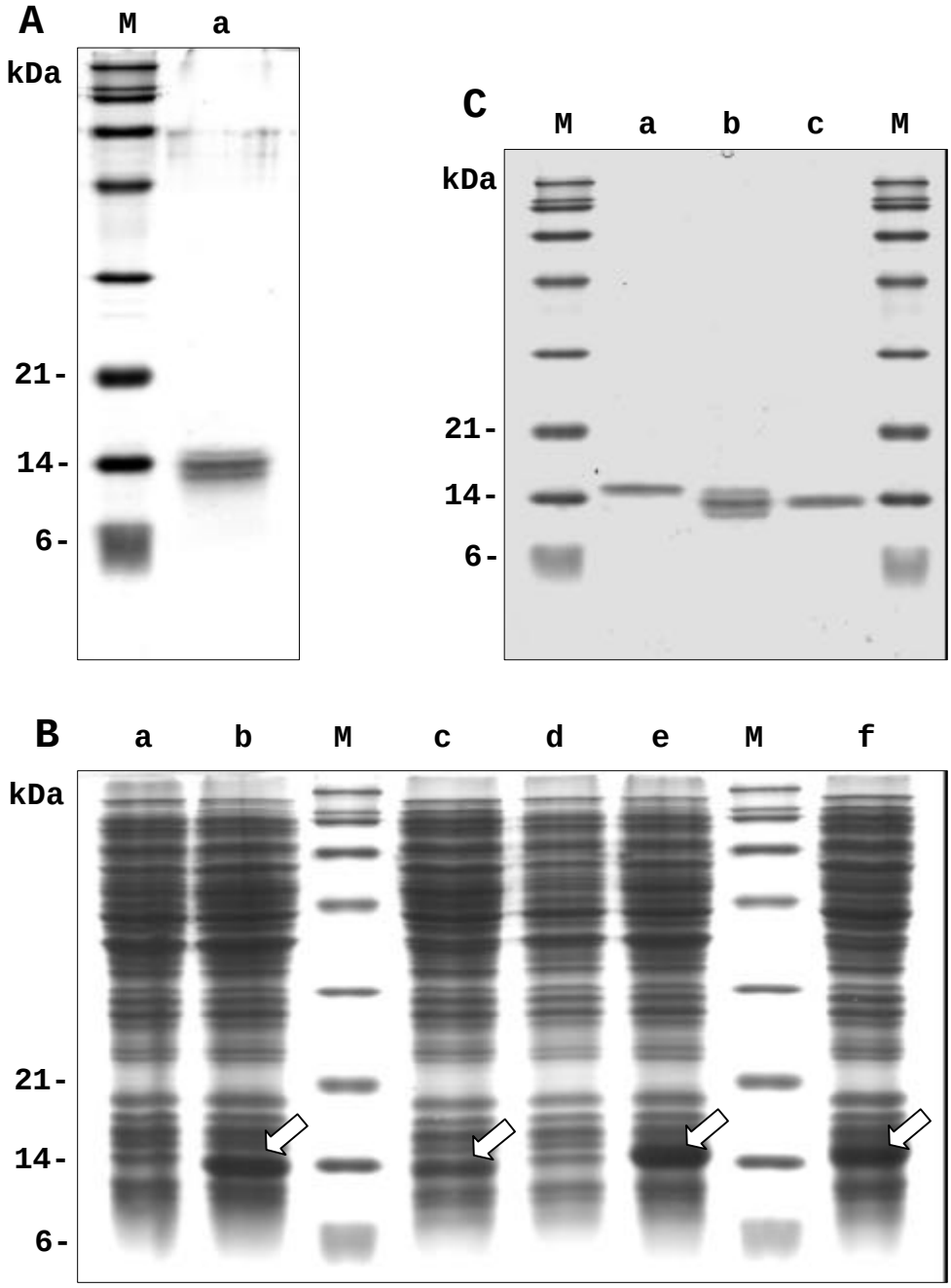


Figure 46: A: Perchloric acid extracted histone H1 of *T. b. brucei* TREU 927/4 on 15% SDS PAGE gel. M: size marker; a: H1 proteins of procyclic culture forms. **B**: Production of two H1 variants in *E. coli*. Lanes a-c: induction of a MAKAS gene of 76 amino acids; lanes d-f: induction of a MNNTT variant (80 amino acids); a and d: uninduced; b and e: induced; c and f: supernate after sonication. M: size marker; arrows: induced bands. **C**: SDS PAGE gel showing migration patterns of the purified H1 variants. M: size marker; lane a: elutrap purified MNNTT variant; lane b: perchloric acid extracted total H1 from procyclic culture forms of *T. b. brucei* TREU 927/4; lane c: elutrap purified MAKAS variant of 76 amino acids.

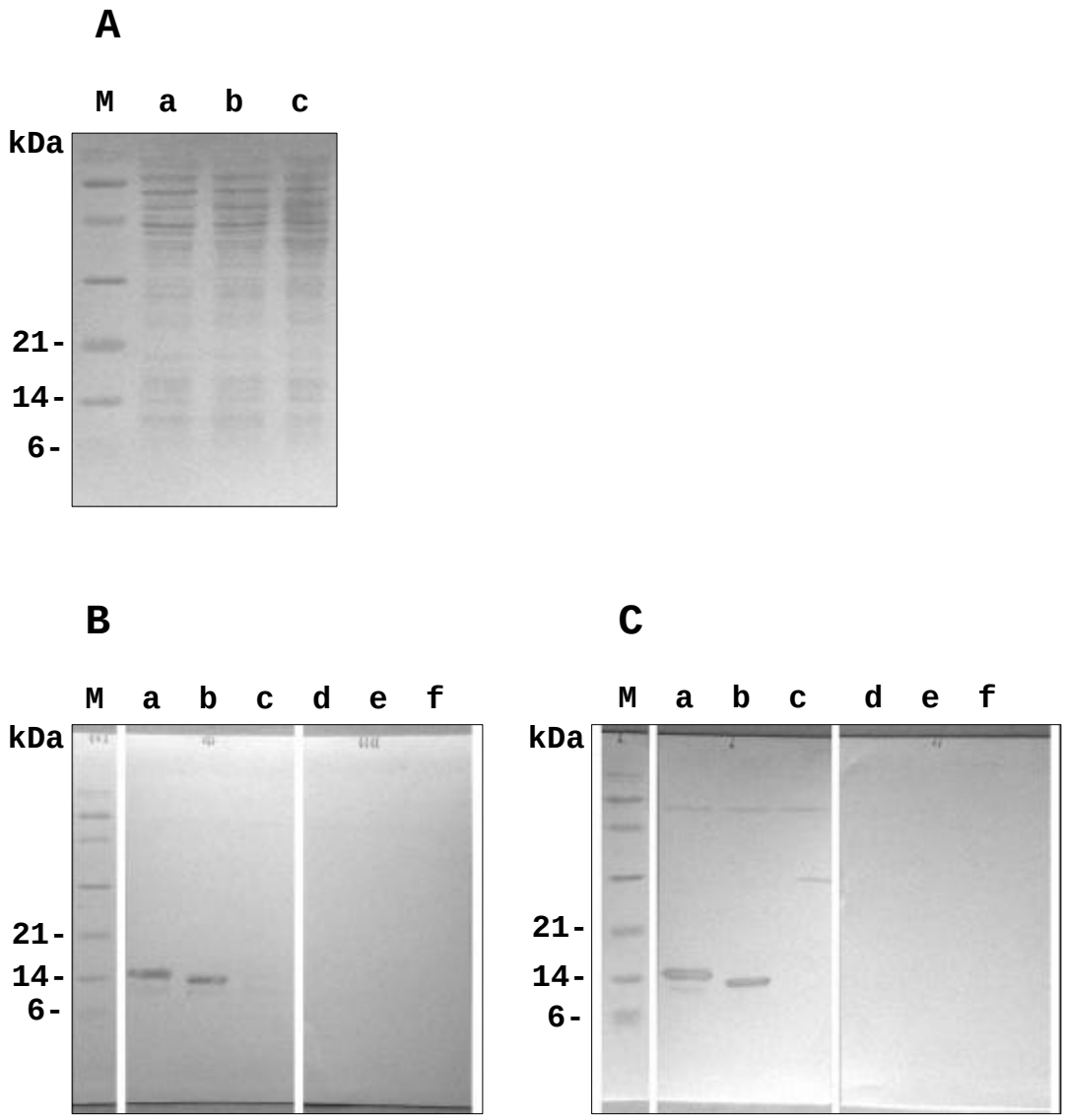


Figure 47: Testing of the serums obtained by immunization of BalbC mice with the two purified H1 variants produced in *E. coli*. **A**: Western blot stained with rouge ponceau which was subjected to immunoreaction with the mice serums. *E. coli* crude extracts containing induced MNNTT variant (lane a), MAKAS variant (lane b) and native bacterial crude extract (lane c); M: size marker. **B**: testing of the serum obtained by immunization with MNNTT variant. Lanes a, b, c: positive serum; lanes d, e, f: control with naive serum. **C**: testing of the serum obtained by immunization with MAKAS variant. Lanes a, b, c: positive serum; lanes d, e, f: control with naive serum.

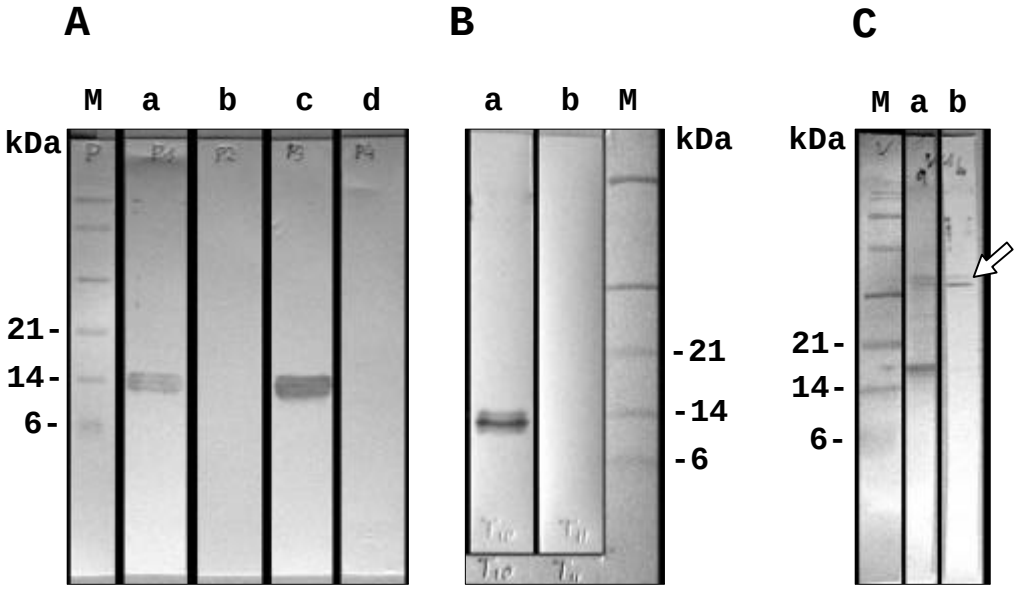


Figure 48: Testing of serums on different H1 proteins preparations and origins. **A**: Testing of both serums separately on perchloric acid extracted H1s of *T. b. brucei* TREU 927/4 procyclic culture forms. Lane a: test with MNNTT serum; lane c: test with MAKAS serum; lanes b and d: respective negative controls with naive serums; M: size marker. **B**: test with mixture of both serums on crude extracts of *T. b. brucei* TREU 927/4 procyclic culture forms. Lane a: positive mixture of both serums; lane b: negative control with naive serum; M: size marker. **C**: Testing of mixture of MNNTT and MAKAS serums on calf thymus H1 (lysine rich fraction). Lane a: membrane coloured with rouge Ponceau; lane b: incubation with serum mixture; M: size marker.

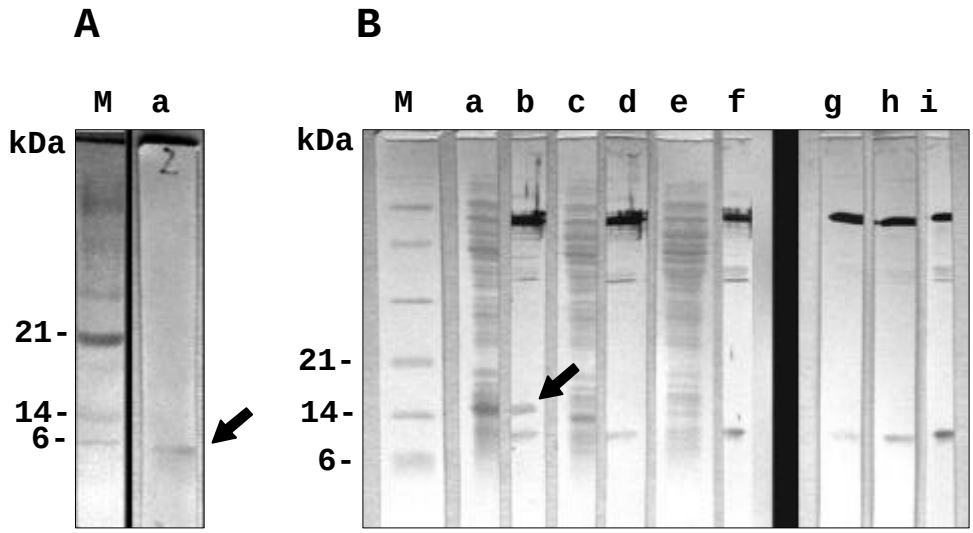


Figure 49: Testing of serum obtained by immunization of BalbC mice with synthetic peptide corresponding to the 12 N-terminal amino acids of MNNTT variant (PEP^+ serum). **A**: test on peptide separated on 15% SDS PAGE gel. Lane a: peptide; M: size marker; arrow: band corresponding to the synthetic peptide. **B**: Test on induced crude extracts of *E. coli*. Lanes a, b and g: induced extracts of MNNTT variant; lanes c, d and h: induced extracts of MAKAS variant; lanes e, f and i: extracts of native *E. coli* srtain. Lanes a, c, e: colored with Ponceau red; lanes b, d, f: incubated with PEP^+ serum; lanes g, h, i: incubated with naive serum. M: size marker; arrow: band corresponding to MNNTT variant.

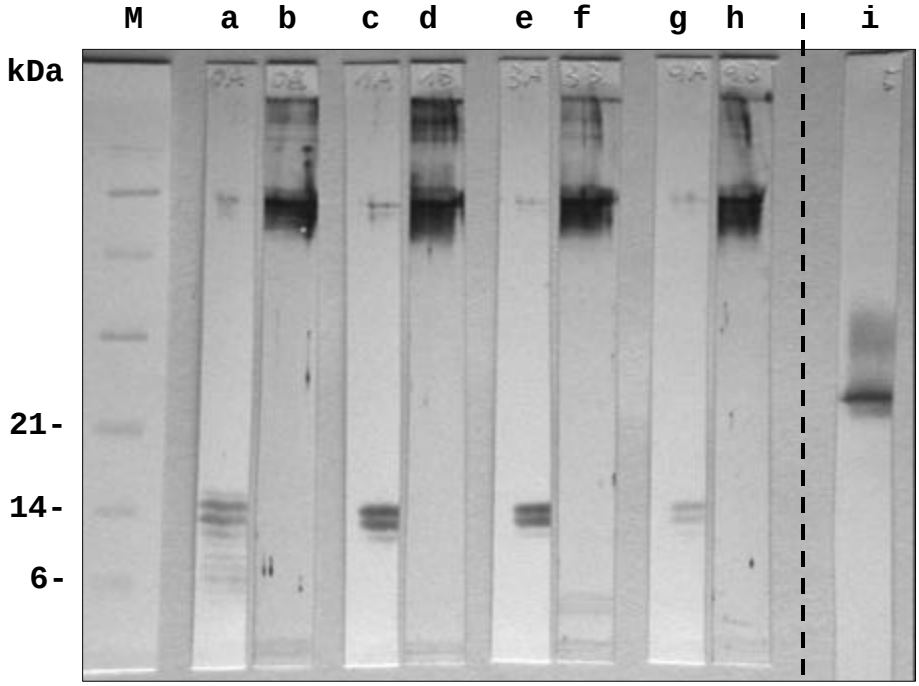


Figure 50: Testing of PEPserum on dephosphorylated perchloric acid extracted histone H1 of *T. b. brucei*. Lanes a and b: not incubated with alkaline phosphatase; lanes c and d: 1h incubation; lanes e and f: 3h incubation; lanes g and h: 9h incubation. Lanes a, c, e and g: tested with mixture of positive serum obtained with native proteins; b, d, f and h: tested with PEP^{+} serum. Lane i: positive control on peptide itself.

Figure 51: Confocal analysis of H1 distribution in procyclic culture forms of *T. b. brucei* TREU 927/4. **A** to **E**: cell at different division stages; **F**: negative control with naive serum; VIS: phase contrast; p-Green: staining of DNA with pico-green (green canal); Anti-H1: localisation of anti H1 antibodies (red canal); SUP: superposition of both previous wavelength.

Figure 52: Confocal analysis of H1 distribution in bloodstream forms of *T. b. brucei* TREU 927/4. **A** to **E**: cell at different division stages; **F**: negative control with naive serum; VIS: phase contrast; p-Green: staining of DNA with pico-green (green canal); Anti-H1: localisation of anti H1 antibodies (red canal); SUP: superposition of both previous wavelength.

Figure 53: A to D: additional controls. **A** and **B**: testing of overlapping signal between used canals; **A**: green and red open; **B**: green shut. **C** and **D**: test omitting first antibody; **C**: procyclic form; **D**: bloodstream form.
E to F: Artefact sometimes observed in bloodstream forms with secondary antibody. VIS: phase contrast; p-Green: staining of DNA with pico-green (green canal); Anti-H1: localisation of anti H1 antibodies (red canal); SUP: superposition of both previous wavelength.

Figure 51:

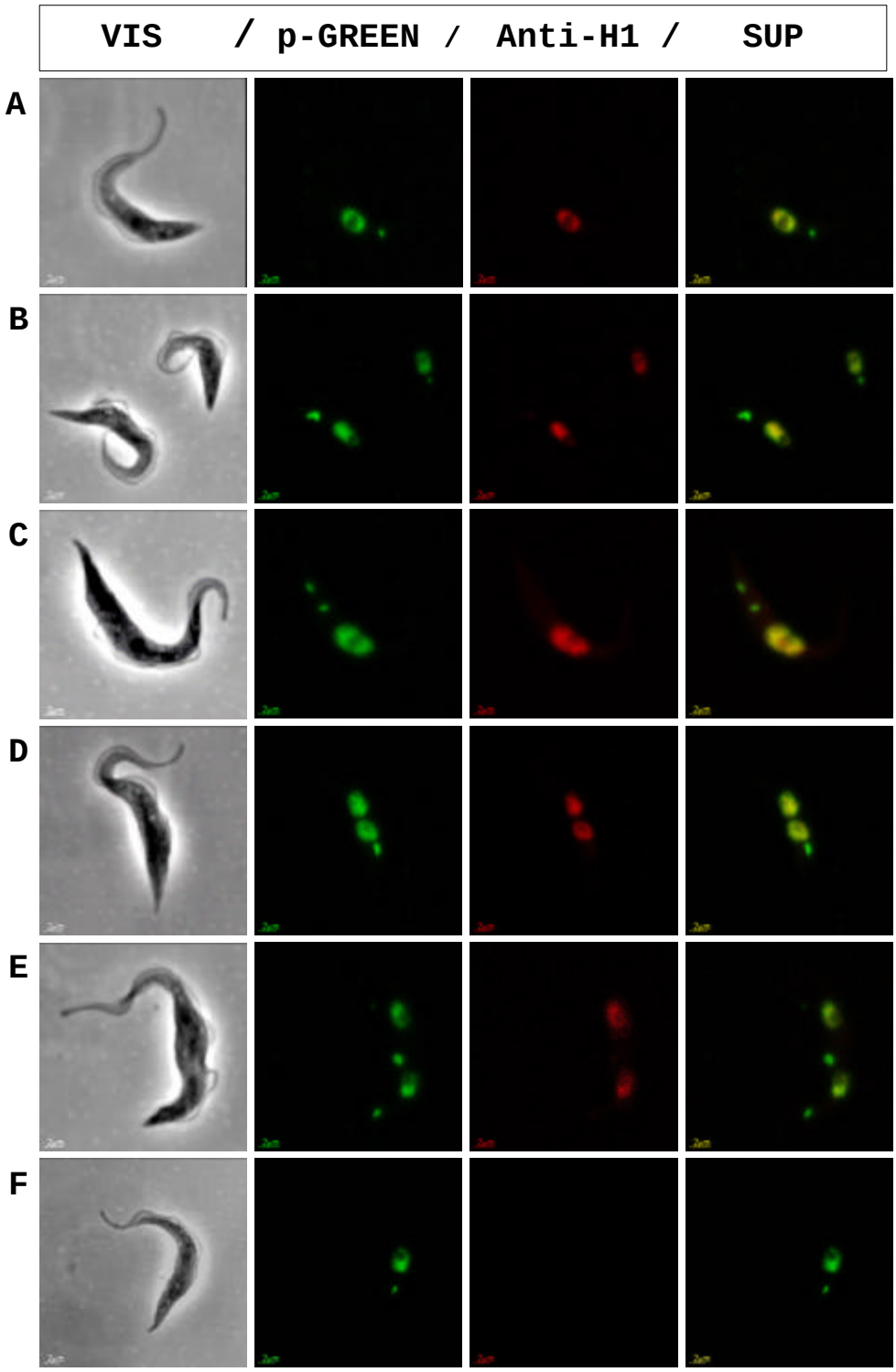


Figure 52:

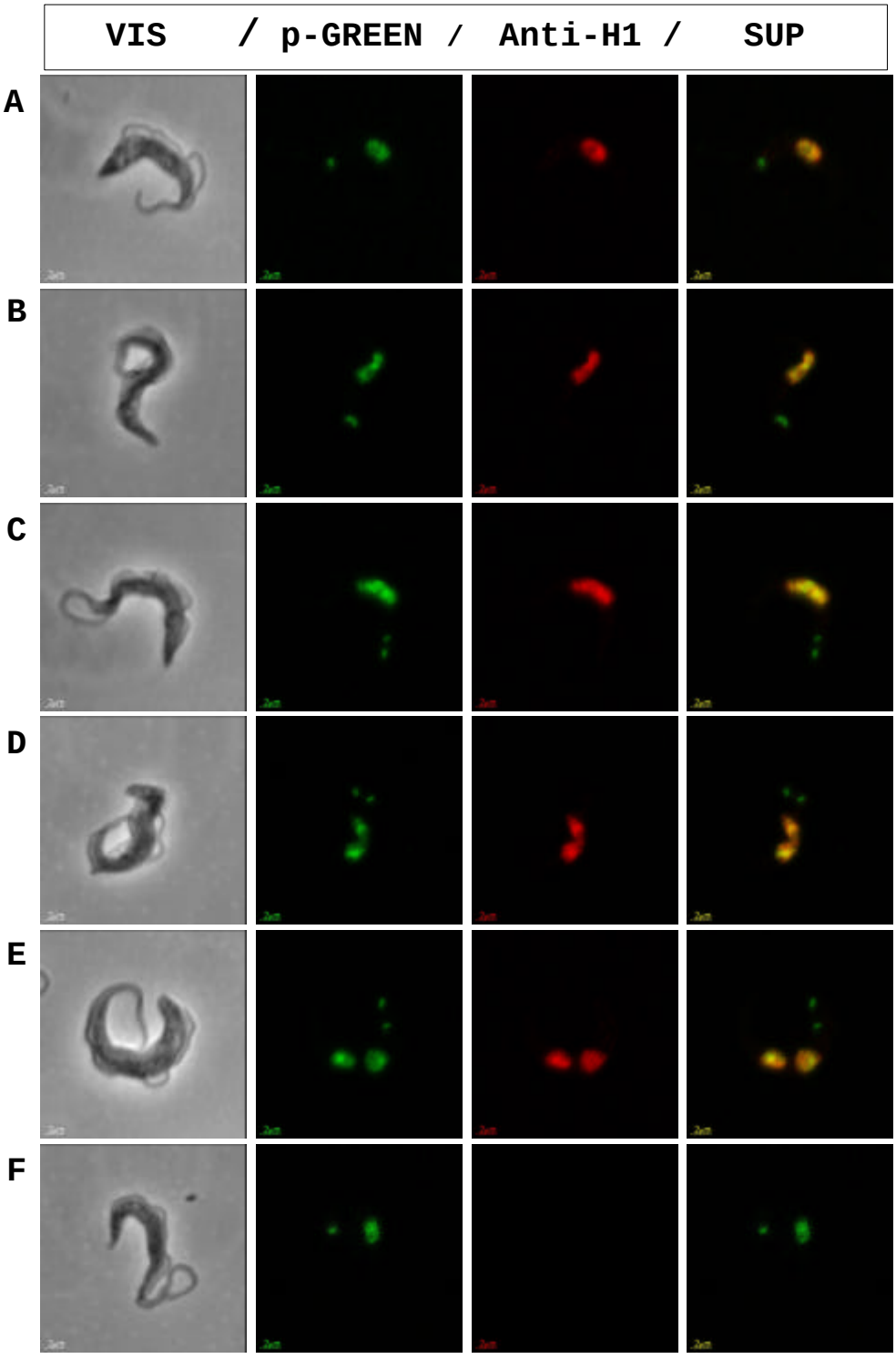
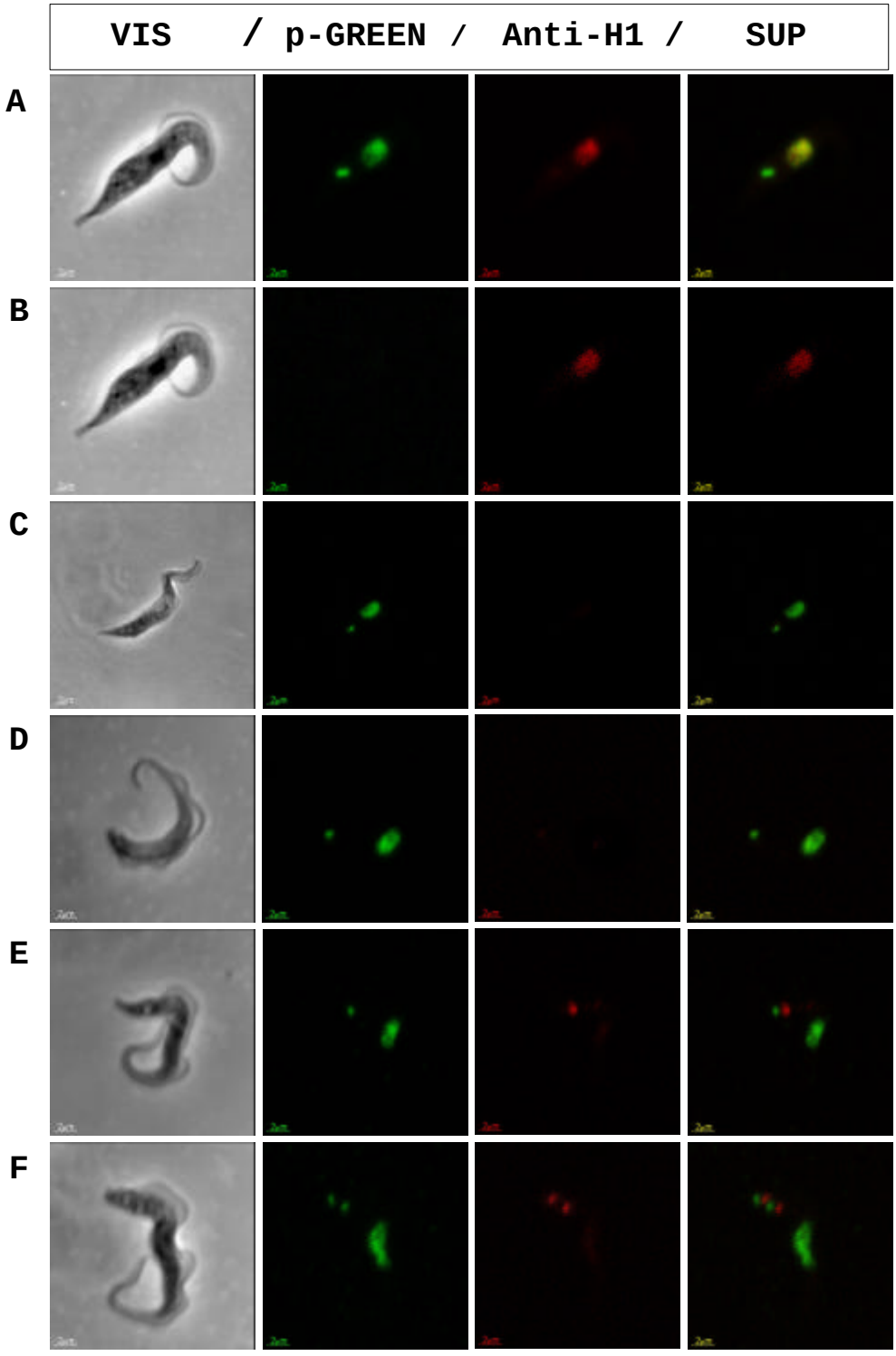


Figure 53:



IV. DISCUSSION

HISTONE H1 GENES AND DEDUCED AMINO ACID SEQUENCES

First of all, the obtained results are in good correlation with previous studies on *T. b. brucei* in which 4 H1 variants and/or posttranslational modifications were described by the use of biochemical procedures [Schlimme, 93; Burri, 93, 94, 95] as well as with the situation encountered in *Trypanosoma cruzi* where a tandemly arranged gene family coding for histone H1 has been described [Åslund, 94]. In *T. cruzi*, the H1 gene family is also formed by several very similar histones H1 of 74 to 97 amino acids from which the longest one is carrying a different 20 amino acid long N-terminal region. However, this extra N-terminal region is very different from the one found on the MNNTT variants in *T. brucei*. The beginning of the other shorter variants found in *T. cruzi* is also always the same among all isolated cDNA or gDNA clones, but it differs from the N-terminus of *T. brucei* MAKAS H1 variants. On the other hand, no clone isolated from *T. cruzi* was bearing a third type of N-terminus like the MAKTT variants found in *T. b. brucei*.
Apart from these short N-terminal regions, the H1 genes of *T. brucei* and *T. cruzi* are very similar if considered grossly and appear to correspond only to the C-terminal tails of higher eucaryote H1s. They completely lack the usual central globular domain and N-terminal tail and are therefore not really comparable to higher eucaryote H1 since the very conserved tertiary structure composed of three domains playing different roles in terms of interaction with DNA and/or (non) histone proteins is absent. This simpler structure and the small size appear to be a general feature of trypanosomatid H1 histones [Galanti, 98] and may be responsible or at least be implicated in the absence of real chromatin condensation.

The various H1 variants that were obtained do also demonstrate the heterogeneity of histone H1 among African trypanosomes. The isolation of 4 H1 size classes in *T. b. brucei* and 7 size classes in *T. b. gambiense* with each class presenting additional sequence heterogeneity, the presence of one type of genes that exhibit a short extra N-terminal motif as well as the stage regulated over-expression of one size class described by Burri [Burri, 94] indicate a very complex situation. Adding to this complexity, a cDNA clone coding for H1 from a different *T. b. brucei* strain (427) which displayed a new intermediate size of 66

amino acids was recently isolated [Hynek, certificate work, 99] (**Annex G**). This asks the question if the comparison between the 4 H1 size classes that were isolated from *T. b. brucei* strain TREU 927/4 and the 4 bands described earlier in *T. b. brucei* strain STIB 345 [Schlimme, 93; Burri, 93, 94] is really pertinent. In the same line of preoccupation, a high degree of histone H2A genes plasticity was observed in *T. cruzi* [Thomas, 00]. These authors demonstrated the existence of polymorphism in terms of relative H2A gene copy number, relative abundance of transcripts and chromosomal location between several strains.

Is the longest H1 variant always the one carrying the different N-terminal sequence in different *T. b. brucei* strains? The fact that the two variants displaying the different N-terminus in *T. b. gambiense* were of intermadiate size among the seven size classes that were found adds to the problem. Furthermore, if such kind of divergence is seen between strains, the whole question of which variant is over-expressed when the parasite is in its vertebrate host becomes really difficult to answer. Burri showed the over-expressed variant of *T. b. brucei* (strain 345) to be the second in size when examining the two-dimensional separations he made. Fortunately, the partial peptidic sequences that were analyzed clearly showed that the over-expressed variant named H1.3 is not related to the variant exhibiting the different N-terminus. Comparison with other *T. b. brucei* strains should be made with special care, especially when only the protein banding patterns are observed with no exact knowledge of the gene sequences.

Even if it cannot be excluded that the 66 amino acid variant was missed in TREU 927/4, this new variant and the additional fact that it exhibits some other nucleotide substitutions in its non-coding sequence tends to show heterogeneity even at the strain level. Species and strain histone H1 heterogeneity have also been reported in Leishmania [Belli, 99].

HISTONE H1 GENE ORGANIZATION

The isolation of gDNA PCR amplified tandems shows that like it was seen in *T. cruzi*, the general genomic head-to-tail organization of histone H1 genes is maintained in African trypanosomes. In *T. cruzi*, partial digestion of DNA with StuI, which restriction site is located just in front of the H1 gene sequences, indicated that there were at least 15 to 20 copies per haploid genome [Åslund, 94]. When making the addition of all variants presenting some substitutions that were isolated from *T. b. brucei*, 10 different ORFs from

which some can be repeated are obtained. In addition, some substitutions that were not influencing amino acid composition were not always further controlled and are not taken in count here. Thus, the number of H1 gene copies in *T. brucei* could well suit the number of 20 copies proposed in *T. cruzi*.

Among the 6 MNNTT-80 clones that were analyzed, the only substitution concerned the codon usage for the lysine at position 64 which was either “aaa” or “aag” with an equal distribution, indicating that this gene must be present at least two times in the genome. The existence of two MNNTT genes differing by the substitution of only one nucleotide which does not affect the amino acid level as well as the fact that MNNTT ORFs were always located in head position on the different tandems that were analyzed in *T. b. brucei* could suggest the existence of two transcription units. However, these variants could also represent the allelic forms of this variant.
Since this variant appears to constitute the head of the polycistronic transcription unit in *T. b. brucei*, it would be interesting to know if the two variants carrying the different N-terminus in *T. b. gambiense* or the one that was described in *T. cruzi* [Åslund, 94] also have the same location in the cluster.

The additional tandems containing two ORFs coding for H1 from which the first is presenting a pair of threonines at its 5’-end and which is preceded by a different 5’ non-coding sequence greatly added to the apparent complexity of histone H1 organization and expression found in *T. b. brucei*. However, the size of this new H1 corresponding to the size of the over-expressed variant described by Burri in bloodstream forms [Burri, 94] gave also new insights into the possible H1 expression mechanisms in the parasite.
Like the MNNTT variants, MAKTT variants are also present in two forms differing by more pronounced substitutions. These may also correspond to allelic forms in the diploid genome.
On the other hand, experimental evidence showing that MAKTT genes are not situated closely to the other H1 genes tend to show that two clusters of H1 genes may be present in the genome. Actually, no PCR amplification could be obtained by the association of a primer specific to MAKTT genes with any primer specific to the other H1 groups. In addition, this kind of variant was never present on any isolated tandem. Relation between these two sets of genes are not known but it appears unlikely that they are located in the same transcriptional unit. Similarly, in *T. cruzi*, genes coding for H2A were found to be

dispersed into two clusters that are located on a single chromosome. The first cluster was shown to contain at least 6 copies of the gene while the second comprises more than 18 units [Puerta, 1994].
It may be that the two H1 clusters are differentially regulated, but to this date, knowledge of the sequences preceding them is limited to a few tenths of bases.

The tandem arrangement of H1 genes in *T. brucei* does also correlate with other histone genes described to this date in other kinetoplastids. H2B has been shown to be tandemly repeated and distributed on 2 chromosomes as well in *Leishmania enrietti* as in *T. cruzi*. At least 4 H2A genes are found in *Leishmania donovani* probably separated into 2 transcriptional units and in *T. cruzi*, 2 H2A loci are present on a single chromosome. Furthermore, an estimation of 14 tandemly arranged H3 gene copies were described in *T. cruzi* [for review, see Galanti, 98]. Fewer data is available on trypanosomatid H4 genes. Up to seven H4 gene copies have been reported in *Leishmania infantum* [Soto, 97]. On the other hand, in *Leishmania major*, only two functional copies of histone H1 that do not appear to be tandemly repeated were reported [Belli, 99].

As it was shown for the two H2A clusters found in *T. cruzi* [Puerta, 94], H1 genes must also be located on one single chromosome in *T. brucei*. Actually, EST T223 showing homology to histone H1 [El-Sayed, 95] and which was tested in order to provide some chromosomal marker does only hybridize to chromosome XI [Website: African Trypanosome Genome Project, http://parsun1.path.cam.ac.uk/index.htm]. Some prospects in order to localize H1 genes in the genome were also made. Using the ORFs of the different clones that were isolated as probes on PFGE separated chromosomes of *T. b. brucei.* (TREU 927/4), only chromosome XI was hybridizing [Brossard, diploma work, 1998]. Thus, if the H1 gene organization really reveals to be constituted of 2 separated H1 clusters, both of them must be situated on the same chromosome. Using the Genome Project data as source of information, it also appears that genes coding for the other histones are located on different chromosomes like it had been described for *T. cruzi* [Åslund, 94]. This represents quite a unique type of histone gene organization among eucaryotes.

TRANSCRIPTION AND MRNA PROCESSING

In trypanosomes, the majority of genes appear to be clustered in polycistronic clusters, therefore, any control at the level of transcription initiation or elongation would imply that genes whose expression must be controlled in terms of developmental stage and/or cell cycle should be grouped in the same unit. This is partially verified for the expression units of the two major surface antigens, VSG and procyclin from which promoters have been described. However, even in this case, additional events of regulation have been reported that are implicated in stage specificity of expression [Berberof, 95]. Actually, the particular organization of the genome of trypanosomes also displays several additional levels of regulation mechanisms influencing mRNA maturation or stability. In the case of VSGs and procyclin, the 3' untranslated region (3'-UTR) was shown to contain elements that have different effects depending on the life stage of the parasite. In bloodstream forms, the 3'-UTR of VSG confers an increased mRNA stability while the same region showed to reduce the mRNA maturation efficiency in the procyclic form. An inverse pattern was obtained by analyzing the effect of the procyclins 3'-UTR [Berberof, 95]. Furthermore, it was also described that genes belonging to the same polycistronic transcription unit can exhibit drastically different mature mRNA levels at a given developmental stage and that they can be differentially stage-regulated [Revelard, 93; Graham, 95].
All this did not make it easier to identify the H1 variant that was shown to be over-expressed in bloodstream forms, since all ORFs of the different variants are separated by very similar IRs. In addition, the regulating elements described for other systems do not appear to follow consensus sequences [Graham, 95].

In higher eucaryotes, except some basal histones that are transcribed throughout the cell cycle [Wells, 85], histone mRNAs are not polyadenylated and present, instead, a stem-loop structure at their 3' ends.
In *T. brucei*, like it has been postulated as being a general feature in trypanosomatids, histone H1 as well as the core histones that were analyzed in details are polyadenylated. In fact, the sequences separating the different ORFs (IR) that were obtained do correlate very well with the typical polycistronic way of transcription observed in trypanosomes [for review, see Vanhamme, 95] and the presence of the typical stretch of pyrimidine which is necessary for polyadenylation at a site located upstream this region was always present. The used site is normally an adenosine and this was always verified on the obtained

cDNAs. Like it was observed for other genes in trypanosomes [Matthews, 94], multiple closely spaced 3'-ends are present on the different polyadenylated cDNA clones that were isolated. However, one clone was also showing a site that was much more far away from the polypyrimidine tract, thus leaving the processed mRNA with a truncated 3' sequence if compared to the other. By observing the other clones around this location, nothing could explain an eventual false annealing of the oligo-dT that was used during cDNA synthesis and thus, the used polyadenylation site must represent a real alternative. In line with this observation, the 3'-UTR was shown to contain positive as well as negative regulatory elements that can have potentially great effect on steady state mRNA levels and/or on translation [Schürch, 97]. Furthermore, alternative polyadenylation was shown to be able to differentially influence expression of a single gene between the two stages of the parasitic life cycle [Erondu, 92]. Unfortunately, the alternatively polyadenylated clone was a partial cDNA missing the 5'-end and thus, it cannot be said with certainty to which size class, nor to which group among the three different 5'-ends it belongs. Its 3'-UTR resembles more the kind of sequence following the variant bearing the different 5'-end but alignment of the partial deduced amino acid sequence to the other variants shows 3 amino acids that would assign it to the other variants lacking the special N-terminus. Even if it is a cDNA clone of *T. b. gambiense* from which a less exact representation of genes and gene organization was gained, it would have been interesting to be able to say if it was an MNNTT gene or another type of genes.

The polypyrimidine tract is also necessary for trans-splicing which occurs downstream on an AG dinucleotide [Huang, 91]. This mechanism appears to be closely linked to polyadenylation and is thought to occur immediately before poly-A tail addition or concomitantly. The spliced leader (or mini-exon) is added to the 5'-end of the following ORF and this way leaves the mRNA in a mature form ready to be translated [Huang, 91; Matthews, 94]. Alternative splicing influencing regulation has also been described in trypanosomes [Revelard, 93], however, all the spliced H1 sequences that were obtained by selective RT-PCR exhibited the same site giving good evidence that this regulation mechanism does not play a role for histone H1 even if only about ten such clones were sequenced.

Global H1 transcription analysis between the two life-stages revealed no significant differences.

Attempts were made to analyze the transcription rate of the MNNTT-80 variant separately. Indeed, before the isolation of MAKTT-variants, this variant was the only one which displayed clear coding sequence differences which could have let suppose additional functionality. The IR following this gene is also presenting the most substitutions if compared to the IR separating the other ORFs and it was also presenting a theoretical stem-loop structure absent on the 3'-UTR of the other variants, letting suppose that MNNTT-variants presented some potential differential transcription/translation abilities. However, the testing of four probes corresponding to the four size classes (complete ORFs) showed no specificity. The very conserved sequences of the different ORFs would rather have let to predict cross hybridization of all variants, so the special situation with the two longest probes hybridizing to each other corresponding ORFs and the two shorter ones doing as well is then probably due to the insertions/deletions that are located inside the different coding sequences. Indeed, proper alignment of the four variants needs the use of gap insertions. The different probes are then probably hindered to match properly the other variants because of the shift introduced by the insertions or deletions. From this, the idea of testing the transcription of the MNNTT-80 variant appeared impossible by the use of this approach since the probe for this variant was not specific. Indirect analysis was also not possible since the probe for MAKAS-76-variants matches the ORF of MNNTT-80-variants and the two shorter probes miss to hybridize to the MAKAS-76 variant. Though, no probe and no combination of probes would allow to differentiate MNNTT-80 variants transcription from the other.

By the use of an oligonucleotide probe matching the specific 5'-end of MNNTT-variants, it appeared that this kind of transcripts were present in bloodstream forms either in equal or in lesser amounts if compared to procyclic forms. Since the only difference of expression that is known is related to an over-expression of one of the four variants in bloodstream forms, it seems that there is no significant transcription difference in this variant.

Qualitative RT-PCR experiments showed that transcripts containing the MAKTT ORF are present in both stages, thus indicating that if the MAKTT variant is involved in differential expression, regulation should take place at higher levels of mRNA processing. Remembering the alternatively polyadenylated cDNA clone from *T. b. gambiense* (see above) which was followed by a 3'-UTR similar to the one following the differently N-terminated variant, but whose partial deduced amino acid sequence was rather corresponding to another classic variant, the parallel with the MAKTT gene presenting the

same features became intriguing. Thus, it could be that the already mentioned effect of alternative polyadenylation mechanism on expression modulation is also implicated here. However, since the RT-PCR approach was based on a H1-specific and conserved primer located a few tenths of bases beyond the stop codon, the eventual different length of the MAKTT transcript cannot be seen. Additional efforts should be made in order to obtain full length transcripts of all variants, especially in bloodstream forms. It should also be noted here that the existence of different transcript sizes were not seen on northern blots of any stage.

The role of a stem-loop structure in the 3'-UTR of procyclin mRNAs was analysed by mutations and/or deletions [Hehl, 94]. The resultant expression of the different constructions clearly showed to be affected even by minor changes. To this date nothing is known about the potential regulatory effect of the theoretical stem-loop structure found on the IR downstream MNNTT-80 and MAKTT-76 variants. However, it is unlikely that this structure plays a regulatory role in H1 expression since only one size class is present more abundantly in bloodstream forms.

EXPRESSION

Among the four *T. b. brucei* H1 variants, the one which is over-expressed in bloodstream forms [Burri, 94] could not be identified according to previous studies in which they were separated by HPLC and characterized by amino-acid composition [Burri, 93]. Even if gross repartition of the different amino-acids involved are well verified, these former results do no reflect the real composition of the gene products. The same kind of divergence was seen in *T. cruzi* [Toro, 93; Åslund, 94] and is probably due to the heterogeneity of this gene family and perhaps to baseline contaminants. However, the peptide fragment H1.3 sequenced by Burri [Burri, 95] which is correlated to the over-expressed variant best aligns to the N-terminal region of MAKAS-76, -71 and -61 variants. If considering the separation according to the molecular weight in 2D gels [Burri, 94], the over-expressed variant appears to be the second in size, that means a 76 amino acid variant and in all the gene/transcript isolation approaches, the 76 amino acid variants were always the best represented.

In addition, the recent isolation of H1 genes that appear to be otherwise clustered goes further in the direction that it is not the MNNTT-80 variant carrying the extra N-terminal region that is over-expressed. According to this, it must be a common H1, displaying no special coding sequence features, that is observed more abundantly relatively to the other variants in bloodstream forms.

The fact that there are probably several clusters that are perhaps differently regulated over the cell-cycle or parasitic life-cycle could bring to a better explanation of the observed differential expression between the two life-stages. Actually, the fact that the MAKTT coding sequence was never isolated in any approach that was made to obtain H1 genes from procyclic material (cDNA screening / RT-PCR) makes it a good candidate to explain the observed over-expression seen in bloodstream forms. It should be noted here that when Burri sequenced the peptide fragments of the four variants [Burri, 95], he used FPLC separated proteins extracted from procyclic forms. If bloodstream forms had been used instead of procyclic forms, the H1.3 fragment may have been different, perhaps corresponding to a MAKTT motif.

However, if the over-expression is due to the MAKTT cluster, it would need some detailed analysis. Actually, this cluster is formed of two ORFs, one which would correspond to the over-expressed variant of 76 amino acids and a second one coding for a 71 amino acid variant. This latter does not appear to modify notably the strength of the band corresponding to this size class in bloodstream forms since Burri did not observe any other intensity changes than the H1.3 spot [Burri, 94].

If considering the tandems of the first cluster containing MNNTT/MAKAS variants, there is a very peculiar variant of as few as 36 amino acids that was never seen otherwise, neither during cDNA screening / RT-PCR isolation, nor on SDS-PAGE gel / Immunoblots. It could be possible that the very small size of this variant of theoretical 3550 kDa makes it migrate out of the gel, but considering the difference between theoretical molecular mass and real migration of the other variants, one could estimate this variant to migrate along the 6 kDa marker band on 15% SDS gel. However, never the faintest band or shadow was observed at this level. It could be that this variant constitutes the end of the polycistronic cluster and that it lacks some features that would allow it to be properly processed and/or translated. Since the IR preceding this ORF does not denote from the other IR that were analyzed, this hypothesis would need knowledge of its 3’ environment. According to this hypothesis, if admitting that the cluster closing ORF cannot be transcribed, processed or

translated correctly, the first ORF of the MAKTT cluster which appears to be constituted only by two ORFs would be the only gene that is expressed and thus bring to the observed enhancement of expression of this size class. Although, the existence of further genes and/or clusters cannot be excluded yet.

In order to determine if the new H1 cluster is responsible for the differential expression over the parasitic life-cycle, the need of a quantitative analysis would be necessary since MAKTT transcripts could be shown to be present in both procyclic and bloodstream forms (see outlook).

The question if MAKTT genes are otherwise regulated during the cell cycle like it has been shown for a fraction of histone H1 in *T. cruzi* [Sabaj, 1997] appears difficult to answer. These authors measured the incorporation of [^{3}H]-leucine, -lysine or –arginine into histones upon Hydroxyurea-induced synchronization of DNA replication [Galanti, 94]. While all core histone and a fraction of H1 were synthesized concomitantly to DNA replication, another fraction of H1 was shown to be constitutively synthesized. These authors also showed that the H1 pattern was not identical before and after DNA replication initiation, an additional band appearing upon replication. However, they could not say if this was due to posttranslational modification or to allelic variants.

In the case of *T. b. brucei*, the MAKTT cluster contains variants that are equally sized to the other variants and thus it would be impossible to discern among products of both clusters by migration on gel and subsequent fluorogram like it was made by Sabaj (see above). However, it would be interesting to know if H1, or a fraction of it, is also synthesized constitutively in *T. brucei*.

Finally, the possible differential transcription analysis of the different H1 variants during the cell cycle, for example by FISH, like it was made for global transcription of core histones in *T. b. brucei* [Ersfeld, 1996], appears difficult because of their strong sequence similarity that do not allow the design of long and specific probes for each variants.

In higher eucaryotes, histone H2A, H2B, H3, H4 and H1 genes are grouped into complete sets which can be repeated many times in the genome depending on organism [van Holde, 89]. The dislocation of such complete sets of histone genes as it is found in trypanosomes could also suggest that stoichiometry of histones may be very different in these organisms. Location of histone genes over several chromosomes and even disruption into more than one cluster for the same histone could allow differential regulation of certain variants as

well as modulation of relative amounts of each histone. Since the over-expressed H1 variant appears to be a common H1 presenting no significant differences at the amino acid level, differences in compaction observed between the two stages of *T. brucei* could then simply rely on production of higher amount of H1 rather than on expression of a variant with different properties. Variation of histone H1 levels was already described in *Leishmania major* in which only two very similar functional H1 genes are found [Noll, 97]. Also in line with this hypothesis, homologous reconstitution experiments in which increasing ratios of H1 to core histones were applied clearly showed an increase in chromatin compaction levels [Schlimme, 95]. Such a mechanism may be related to the proposed existence of two H1 clusters in which the MAKTT variant could be brought to efficient translation in bloodstream forms. However, lots of additional data and experimentation is needed to gain insight into this question.

PRIMARY STRUCTURE AND AMINO ACID MOTIFS

If compared to H1s of *T. cruzi*, the relative simple amino acid composition and primary structure of African trypanosome H1 histones already give some interesting indications upon their ability to undergo posttranslational modification. Threonine and serine (only one in *T. b. brucei*, three to four in *T. cruzi*) exhibit a hydroxyl moiety which allows phosphorylation, a common posttranslational modification well described in histone H1 [van Holde, 89]. This kind of phosphate linkage is very stable to even strongly acidic solutions and thus resist to acid extraction and purification of histones. Lysine, histidine and probably arginine, can also be subjected to phosphorylation, but in a more unstable and very acid-labile way.

In *T. b. brucei*, H1 sequences do not exhibit the well defined "(S/T)-P-X-(K/R)" phosphorylation motif [Hill, 90; 91] present one to three times in *T. cruzi*. In addition, SPKK motifs have also been shown to be nucleic acid-binding units of proteins [Churchill, 89]. In *T. b. brucei*, only the MNNTT-80 variants do exhibit such a kind of motif, at the end of its specific N-terminus, with a threonine instead of the serine [Suzuki, 89]. The other threonines in these variants do not resemble typical phosphorylation motifs. The unique serine situated at position 5 in MAKAS-76, -71 and –61 variants corresponds to other motifs proposed to be phosphorylated, the serine being flanked by two hydrophobic residues and immediately preceded by a basic residue (lysine or arginine) [Hohmann, 83].

The same kind of analysis can be made when considering the 7 H1 variants of *T. b. gambiense* since no additional amino acids susceptible to be phosphorylated are introduced. The two threonines present in the recently isolated MAKTT-76 variants do not resemble classic phosphorylation sites.

The dephosphorylation and subsequent HPLC-elution made by Burri also showed that only three of the four H1 variants were phosphorylated since the peak corresponding to the smallest variant H1.4 was absolutely not modified even after a prolonged incubation with alkaline phosphatase [Burri, 95]. This observation remains difficult to explain in regard to the deduced peptide sequences. All variants do present the same general distribution of plausible phosphorylation sites and the deletions being responsible for the different size classes do not contain any serines or threonines.
However, the absence of multiple phosphorylation sites and of nucleic acid-binding SPKK motifs in *T. brucei* H1s could explain the differences seen in chromatin compaction if compared to *T. cruzi* [Hecker, 94].

The role of the specific N-terminal end of the MNNTT variants is unknown. However, one could speculate that it gives some other properties to these variants in terms of interaction with chromatin and/or with the core histones. Another hypothesis could be an implication as a signaling motif in transcription, in maturation of mRNA precursors or translation.

PHYLOGENY AND EVOLUTION

Comparison between histone H1 of different lower eucaryotes appears to be very difficult. Blast searches against *T. b. brucei* H1 genes indeed give mainly histone H1 to be the best related sequences, but there are also astonishing gaps in the obtained listing between organisms that are considered to be closely related. On the other hand, higher eucaryote H1s are not only very divergent from their lower eucaryote counterparts, but they also do present domains that are absent, especially in trypanosomes. It has already been observed that trypanosomatids H1 do only correspond to the C-terminal domain of classic H1 but even when considering only this portion of protein, conservation of primary structure is not evident. Between H1 of *T. cruzi* and Human, a homology of only 43% has been reported and closer analysis shows no longer stretches or amino acid motifs to be conserved

[Galanti, 98]. A general fact that appears evident when trying to align trypanosomatid core histones with higher eucaryotes histones is that even if their central globular domains are relatively conserved, their tails are greatly divergent. For all these reasons and considering that *T. b. brucei* H1s correspond only to the C-tail, the observed lack of similarity becomes obvious. In the evolutionary point of view, by analyzing the N-terminal domain of *T. cruzi* H4, it was proposed that there were two evolutionary steps with a high rate of amino acid substitutions in this domain up to ciliates and then a slow-down in the metazoa branch. This slow-down being perhaps related to the apparition of novel functions for this domain [Toro, 92]. This theory could be applicable as a general view of histone evolution since lots of parallels can be drawn from data concerning all core histones and also H1 when they are compared with their higher eucaryote counterparts [Galanti, 98]. The heterogeneity of H1 genes among different African trypanosome subspecies and strains is in line with this. It seems reasonable to admit that the needs of histones like they appeared later with gain of novel domains and functions were not a necessity for organisms which stayed at the unicellular level and which genome is clearly smaller.

Resulting from this observation of poorer histone sequence conservation in protozoa, these proteins were proposed to give potentially a new tool for phylogenetic investigations among lower eucaryotes. Indeed, histone H4 shows more divergence between the American and African trypanosomes than between Man and Sea urchin [Hecker, 93]. However, this divergence is so pronounced when looking at trypanosomatids H1 that they clearly would be an unadapted tool for such investigations. The sequences are to divergent and probably reflect a complex moment in evolution with lots of branching events. The unexpected divergence of 7,7% observed between H4 from *Leishmania infantum* and *Leishmania tarentolae* which is considered to be one of the most conserved gene at all provides a good illustration of this [Lukes, 00].

Therefore, no tree constructions that would have compared too different sequences and/or domains that are not really homologue were made. The establishing of similarity rates also appears inappropriate, mainly because of the same reasons. Thus, only the objective approach of direct alignments that clearly shows what is comparable was considered.

In *T. cruzi*, the global amino acid composition, the length and the basic properties from one end to the other of the protein are maintained if compared to *T. brucei*. However, the alignment does also show that these two species are much more far away from each other than would have been expected. This clearly indicates that the African and American trypanosomes must have separated very early during evolution not only from the other

eucaryotes [Woese, 90], but also from each other. This is in line with phylogenetic studies based on rRNA genes showing that *T. brucei* constitutes the earliest divergent branch in trypanosomatids evolution [Maslov, 96] and an early branching of Salivaria in general [Haag, 98].

Unlike the H1 gene discovered in *L. major*, the *L. brasiliensis* H1 also shows great amino acid motifs similarity with *T. brucei* H1 and the observed array of 2 basic amino acids separated by 3 non-polar residues is perfectly identical. It appears rather astonishing that this H1 of an American leishmania matches so well African trypanosome H1 while the comparison with Old World leishmania is less obvious. This is especially true when considering that leishmania are always located far away from *T. brucei* on evolutionary trees based on rRNA subunits nucleotide sequences [Maslov, 96; Haag, 98].

The second non-trypanosome H1 that showed highest similarity with *T. brucei* is an H1-homologue (BpH1) found in the procaryote *Bordetella pertussis* [Scarlato, 95]. This protein is about 2 fold the size of *T. brucei* H1 and the match is located along its middle-end region. Here, the similarity goes still further; the lysine doublets being strictly conserved while some arginines were substituted in *L. brasiliensis*.
The amino acids used between the lysine pairs are the same than in the two other H1 and if the global composition of the aligned regions of these three H1 was compared, it would be difficult to see any difference. However, while the *L. brasiliensis* H1 is quite of the same size than the *T. brucei* H1 with an addition of a short 25 residue long N-terminal region that could be analogue to the N-terminus of *T. b. brucei* MNNTT-80 variants, in *B. pertussis*, the protein presents two extremities of different biochemical properties and much more complex amino acid sequences. Such H1-homologues were shown to bind DNA in a non-specific fashion, to be able to compact DNA and perhaps also to modulate expression [Scarlato, 95]. Dnase I protection assays also showed that BpH1 may cover DNA fragments from end to end and to have important effects on DNA flexibility [Zu, 96]. Even if some organisms closely related to Bordetella display an endosymbiontic life style in several trypanosomatids (genus Crithidia and Blastocrithidia) [Du, 94], it appears difficult to imagine an evolutionary relation or some kind of transfer between *B. pertussis* and *T. b. brucei*, since African trypanosomes seem to have branched very early in evolution [Maslov, 96; Haag, 98] and also because the hypothetical entry for symbiosis into kinetoplastids is thought to have occurred far after this branching event [Du, 94]. More

data upon phylogeny among and/or between these organisms would be necessary and in particular some hints concerning histone H1 (homologues) should be investigated.

BpH1 as well as another H1 homologue (Hc1) from *Chlamydia trachomatis* [Hackstadt, 91] showing less homology to *T. b. brucei* differ from the known procaryotic H1-like proteins because their amino acid composition and primary structure show high degree of homology with histone H1 proteins. Hc1 was also shown to be involved in condensation of the chlamydial nucleoid [Barry III, 92]. Interestingly, *C. trachomatis* is also an obligate intracellular parasite, but this could be due to the fact that pathogens are, indeed, much more studied than free living and harmless organisms. In this organism, Hc1 seems to be necessary to condense DNA upon intracellular differentiation into basal bodies, the dispersion form of the bacteria.

Bacterial endosymbionts have also gained attention because of possible relevance in regard to origins of mitochondria and chloroplasts [Gray, 88]. However, because of the large clefts appearing in the present understanding of phylogeny, the presence in procaryotes of histone homologues resembling mainly the C-terminal DNA-binding domain of “classic” H1 from which the one found in Bordetella is also well aligned to the H1 of *T. brucei* may also represent a convergent evolution of efficient DNA-binding structures.

FUNCTIONAL ASPECTS

Beside the histone H1 sequences discussed above, blast searches also revealed an entry that presented quite an absolute match to *T. brucei* H1. This entry is referred to as a *T. b. brucei* microtubule associated protein. The surprising assignation of this perfectly aligned sequence found by blast search could be due to the used microtubule purification procedure which perhaps allowed co-extraction of H1. Another hypothesis could be that histone H1 displays additional functions in trypanosomes, since chromatin condensation does not occur in a significant way. In agreement with this hypothesis is the finding that H1 of sea urchin sperm cells was shown to stabilize microtubules and antibodies raised against this protein also recognized organelles like the flagellum or cilia in different lower eucaryotes [Multigner, 92].

In addition, during the past decade, lots of studies were showing that histones have significant effects on gene regulation and this research direction will still continue to prove that histones are not only structural proteins only linked to the first levels of chromatin

condensation. However, the difficulty to make an interpretation of these results that come from different *in-vitro* and *in-vivo* models as well as from different organisms will probably generate lots of additional experimentation and discussion [for review, see Grunstein, 92; Wolffe, 92; 97; Widom, 98].

In *T. b. brucei*, it appears difficult to make an approach of these questions, mainly because of the complexity of the gene family in question and also because of the numerous copies that are present and that are probably dispersed into at least two transcriptional units. For example, knockout experiments like they were made in *Tetrahymena thermophila* [Shen, 95] and which showed that H1 was not essential for cell survival would be very difficult to perform in *T. brucei*. A prerequisite for this kind of approaches would be an exact knowledge of the distribution and the surrounding sequences of *T. brucei* H1 genes. However, the rapid advances made by the Trypanosome Genome Project should provide with this in a close future. To date, the only indirect information that could be interpreted in the way of an implication of *T. brucei* H1 in gene regulation comes from chromatin reconstitution experiments. H1-depleted chromatin supplemented with the H1 containing fraction obtained by sucrose gradients which also contains additional nuclear proteins do not allow to reach the same chromatin condensation degree than with purified H1. This can be explained by the fact that the used fraction also contains transcription factors which may act antagonistically to H1 by mediating a decondensation of chromatin for access to DNA information [Burri, 95].

The idea that H1 could also act as a structural protein by stabilizing microtubules or could even be located elsewhere, for example in the kinetoplast for stabilizing and processing kDNA was not, in our knowledge, investigated yet in trypanosomes. In *T. cruzi*, no report of extranuclear H1 protein could be found and all the work on *T. b. brucei* was made on H1 extracted from isolated nuclei. No special effort was made to isolate the other cellular compartments in order to highlight eventual other histone H1 locations. However, the use of anti-H1 antibodies revealed no cytoplasmic signal and no other cell structure than the nucleus could be highlighted by immunofluorescence. In *Leishmania major*, the H1 gene products also showed to be exclusively nuclear [Noll, 97].

Within the nucleus, the used fixation procedure allowed to gain an astonishing specific signal and access to H1 appears to be fulfilled throughout the whole cell-cycle without great differences between procyclic culture forms and bloodstream forms. Previous studies, although, suggested an internal localization of H1 upon chromatin compaction in *T. brucei*. Proteolysis with immobilized trypsin showed that at low ionic strength which means in an

open nucleosome filament, H1 was accessible to the enzyme, whereas at higher ionic strength, no degradation was seen [Burri, 95]. This indicated that in a condensed nucleosome filament, H1 is not exposed at the surface. However, it may be that when taking a whole nucleus, chromatin presenting different levels of compaction is present in a random distribution, giving rise to the observed patterns with anti-H1 antibodies. Indeed, early electron microscope studies already showed the loose chromatin without any compaction and its distribution over the whole nucleus during mitosis [Vickermann, 70]. It may also be that the used cell preparation technique is denaturing the chromatin structure, giving access to H1.

The negative results obtained by the use of the PEPserum which should recognize the special N-terminal region of the MNNTT-80 variant could suggest that this little portion of H1 is inaccessible, being hidden or enclosed by other structures. Since the function of this H1 portion is unknown, it is difficult to make an interpretation of this. It may also be that this N-terminus becomes modified even by phosphorylation or by any other posttranslational modification, preventing access to the antibody. In line with this is that, on western blots, a response was only obtained with proteins induced in bacteria but the serum always failed to recognize proteins that were obtained from trypanosome material. However, attempts made on dephosphorylated perchloric acid extracted H1 also revealed to stay negative.

Ultrastructural observation in which dividing nuclei thin sections of *T. cruzi* were 3D-reconstituted showed that trypanosomes undergo a series of nuclear processes for division that greatly differ from those observed in higher eucaryotes [Solari, 80]. Again, the same general features already described by Vickermann [Vickermann, 70] were observed but in more details. While chromatin is packed in lumps lying on the inner side of the nuclear membrane at interphase, it becomes dispersed throughout nuclear division. This author also described a complex microtubular structure which, in cooperation with some other dense aggregates called *dense plaques*, was implicated in chromosome separation.

Therefore, the fine nuclear localization of these short H1s and their interactions with other structures should now be investigated in more details. It remains unclear how these simple proteins interact with DNA, the nucleosome or other structural and/or regulatory proteins. Is the absence of chromatin compaction observed in these parasites during mitosis related to the short and simple structure of histone H1 coupled to some other divergent characteristics of trypanosomatids core histones? Or, is the nuclear organization and DNA

processing so divergent in these organisms that the parallel with higher eucaryotes histone functionality is impossible? Or, finally, is it something in between, trypanosomatids displaying some kind of primitive DNA organization system without the ability or need to compact tightly their DNA upon division, but using partial or different histone potentials for other purposes?

Actually, increasing evidence comes up indicating that the three-dimensional organization in the nucleus and the proximity of genes to protein structures like the nuclear envelope or centromeric complex is involved in gene regulation [Ersfeld, 99]. According to authors whose results suggest some kind of interaction of H1 with structural components like microtubules [Multigner, 92] or specific DNA sequences implicated in scaffold anchoring [Ivanchenko, 1992] and the fact that some H1 transcripts in trypanosomatids are not always coordinated with DNA replication [Sabaj, 97], the question arises whether these small and apparently primitive histone H1 do only display a primordial structural function in chromatin condensation or have an absolutely different primary role.

The special features of nuclear division in trypanosomes, with the loose DNA and the persistence of the nuclear membrane throughout division could represent a primitive model for mechanisms involving histones that are still unclear in higher eucaryotes.

However, despite the lack of pronounced chromosome condensation, trypanosomes do express histone H1 and regulate it throughout their life-cycles and cell-cycles.

V. OUTLOOK

The general organization of histone H1 genes in African trypanosome has been examined and lots of parallels but also differences were highlighted when comparing these results with the situation encountered in other kinetoplastids. Some additional investigations should now be made in order to gain complete comprehension of H1 genes before going further into the analysis of histone functions in these organisms. First, it would be advisable to clarify the astonishing flexibility and divergence of H1 genes seen in different species, subspecies and strains since evidence accumulated in this direction. Investigations were made on different parasitic sources, including cDNA libraries and genomic DNA of several origins. In order to be able to benefit from the advances made in the Trypanosome Genome Project, this work soon focused more precisely on *Trypanosoma brucei brucei* (Strain 927/4) which was selected for genome sequencing. However, the additional intermediate H1 variant that was isolated from the cDNA library of *T. b. brucei* (Strain 427) strongly suggests that the situation could be very complex when trying to compare data from different origins. The numerous studies made on African trypanosomes H1 on the protein level by biochemical characterization and in vitro chromatin reconstitution was made on strain STIB 345. Even if the results presented here are in good correlation with these information, the proof that they can be compared should be gained at the molecular level. Genomic isolation of H1 genes from *T. b. brucei* STIB 345 appears to be the most efficient approach and would not imply important investment of time or of material. This is particularly important for the elucidation of the problem of what variant is over-expressed in bloodstream forms and seems to be a prerequisite for further approaches of the function and implication of this differential expression between the two life stages.

A further question which should be answered is the sequential organization of the clustered genes in other subspecies and species. In *Trypanosoma cruzi*, one variant displaying a 20 amino acid long N-terminal region which is probably homologue to the one found on *T. b. brucei* MNNTT-variants was described. If this variant does also constitute the head of the gene cluster, the hypothesis that this element is involved in a particular mechanism would appear strengthened. The discovery of two such genes of different length in *Trypanosoma brucei gambiense*, however, tends to increase the complexity of the problem.

The isolation of the MAKTT gene which is not physically closely linked to the other genes should also be investigated in more details. RFLP experiments on genomic DNA and subsequent hybridization against oligonucleotide probes specific to each group of genes could be made in order to highlight the different clustering of these genes.

The decision to focus on the *T. b. brucei* strain which is used as model in the Genome Project appeared not to be the best solution since to this date, sequencing of chromosome XI is not started. However, it is possible to obtain genomic clones after the screening of high density filter carrying several thousands of available genomic clones [Melville; 98a, b]. This approach could be very useful in order to map the H1 clusters and gain a better view of their surrounding sequences and the relative positioning of MAKTT genes to the other H1 genes.

If MAKTT genes really constitute a different cluster, it would also be interesting to know if other genes are associated in the transcription unit and to see in what kind of mechanisms they are implicated.

MAKTT variants could be responsible for the different expression levels seen between procyclic and bloodstream forms but the fact that its coding gene is followed by a 71 variant ORF makes it difficult to explain. The hypothesis that the last ORF of an H1 cluster is unable to be transcribed and/or processed properly is difficult to verify since the second ORF of the MAKTT cluster resembles in all points the other genes of the same size. It was also shown that transcripts of MAKTT genes are present in both stages. However, all this and the additional fact that no significant difference in global H1 transcription between the two stages can be highlighted is not really contradictory to the implication of MAKTT in differential expression. Actually, most regulation events must take place posttranscriptionnally. This latter fact makes it questionable if it would be useful to investigate the relative transcription rates of the different H1 genes even by the use of competition RT-PCR or real time PCR experiments. However, these approaches could be very powerful and some real time PCR designs can allow to differentiate between products differing by only one nucleotide [Marras, 99].

Another aspect which should be controlled is the comparative global expression level between procyclic and bloodstream forms. All previous studies compared the relative abundance of the four variants but the two stages were not compared to each other, even if some differences were noted on Triton acid-urea gels [Schlimme, 93]. Knowledge of the expression levels according to life-stage could be useful for further comprehension of what

is different between them. Phosphorylation analysis between the two stages could also be informative since only procyclic forms were analyzed to date [Burri, 95].

For all these additional investigations, it would also be advisable to care about proliferation state and if possible also about the cell-cycle. As it was shown in *T. cruzi*, a fraction of H1 is constitutively expressed and is not linked to DNA replication [Sabaj, 97]. Furthermore, differentiated trypomastigote forms resting in G0 phase do not express any subfraction of H1 and display an increased stability and compaction level of chromatin.

In *T. b. brucei*, it was observed that the differentiation of replicating slender bloodstream forms to non-dividing stumpy forms limits the parasite population and allows survival of the mammalian host [Vassella, 97]. This differentiation is triggered by the cell density and is temporally correlated to cell-cycle arrest and competence of a bloodstream form population to initiate transformation to procyclic forms [Matthews, 94; Reuner, 97]. It may be that the over-expression of H1 is also correlated to this moment in the trypanosome population development, participating in stabilization of the chromatin upon cell cycle arrest.

In order to verify if the over-expression of one variant in *T. b. brucei* is linked to the higher temperature of the vertebrate host environment [Schlimme, 93] or if other factors like basal expression, cell density and/or cell-cycle [Sabaj, 97] are involved, it is necessary to characterize more precisely the cells that are used. Indeed, when bloodstream forms were propagated in mice, there was no real control of the obtained cell population and the collection moment was only a matter of number of cells. In addition, different strains appear to behave differently in the same growth conditions.

Synchronization procedures in trypanosomes are not really available, however, it is possible to synchronize pleomorphic trypanosomes as stumpy forms so that they are also synchronized in G1/G0 [Vassella, 97]. It would be very interesting to see if there are also differences between these different cell populations by performing two dimensional H1 separation as it was made by Burri [Burri, 93, 94].

The most important problem when trying to analyze which variant is over-expressed is the very strong sequence and size similarity between the different variants. If it reveals impossible to point out an eventual differential transcription rate by already mentioned techniques, it would be very difficult to show, for example, that MAKTT variants are involved in this phenomenon. Indeed, it is very unlikely to rise antibodies specific for these variants and the only approach which could be envisaged would then be to sequence the N-

terminal region of the over-expressed protein taking care to perform histone H1 extraction on bloodstream forms and not on procyclic forms like it was made by Burri [Burri, 95].

Once these elements are investigated and a better general view of histone H1 in African trypanosomes is gained, a wide range of functional studies could be envisaged.

First of all, it would be possible to reach a better understanding of H1-chromatin interaction by performing reconstitution experiments on purified trypanosome chromatin with all variants produced separately in an expression system. In previous studies, these experiments were made with purified H1 obtained from cultured cells. Even if HPLC separation of all variants was finely tuned [Burri, 95], no attempt was made to reconstitute nucleosome filaments with the different fractions and it would be interesting to see if there is any difference in compaction ability, principally between MNNTT variants and the other variants. However, one should also keep in mind that H1 variants are known to be modified at least by phosphorylation which are not accomplished in procaryotic expression systems.

The use of the specific antibody should also allow, in complement to electron microscopy, to study the positioning of these small H1 over the nucleosomic filament. It is known that trypanosomatid H1 do only correspond to the C-terminal tail of higher eucaryote H1s which is thought to bind with linker DNA. However, most described H1 in lower eucaryotes display a short N-terminal sequence diverging greatly from the rest of the protein. These sequences which are present only on one variant in trypanosomes could display additional function and represent an evolutionary step in the direction of developing binding structures for other chromatin components. If this was true, trypanosomes would be good models since they possess both types of H1 proteins, bearing or not an additional N-terminal sequence.

The affinity of *T. brucei* H1 for DNA or some preferences for particular DNA sequences like the AT-rich motifs which are known to be scaffold binding in higher eucaryotes [Izaurralde, 89; Ivanchenko, 92] would also be of interest. This can be performed in vitro even by simple incubation of the different variants with southern blots of several DNA fragments of different nature or conformational states [Scarlato, 95] or by the use of more complex assays like One-Hybrid systems that are commercially available. As a correlative to this, the study of competition with other proteins for their chromatin access could also give precious indications concerning their implication in regulatory mechanisms of gene transcription. In the same line of preoccupation, the affinity of H1 for structural

components like scaffold proteins or other structures involved in nuclear division should also be examined. The Two-Hybrid system could be a way to elucidate some of these questions but it should be made with special care, since the extreme biochemical properties of H1 and especially their high positive charge could give results that are not corresponding to real *in vivo* mechanisms.

H1 gene knock-out experiments in other protozoa have been shown to have little effect on cell survival and growth [Shen, 95]. This kind of approach appears rather difficult in trypanosome because of the multiplicity of the H1 genes and the probable distribution in at least two clusters on one chromosome. However, trypanosomes reveal to be good subjects for targeted homologous recombination and this technique could be used to study the importance of H1 in the biology of trypanosomes [Asbroek, 90; Lee, 90; Eid, 91]. If MAKTT genes reveal to constitute a separate cluster, homologous recombination experiments in which the MAKTT ORF is replaced by a reporter gene could allow to monitor the transcription/expression of this variant and its implication in the H1 over-expression and/or the cell-cycle/life-cycle of the parasite.

Over-expression on the other hand would be feasible by the use of stable and inducible vectors that are now available for African trypanosomes [Biebinger, 97]. One could imagine to perform such kind of transformation with different variants and subsequent observation of cell growth in culture or fine analysis of chromatin condensation which would then give relevant indication on H1 function. These kind of vectors are also suitable for many other approaches like antisense RNA for down regulating the expression level or for the production of stage-specific gene products at atypical stages.

Finally, it could also be envisaged to analyze the importance of the 3'-UTR separating H1 genes. The use of reporter genes fused to these sequences or parts of them and ligated into an inducible vector like it was made for other trypanosome gene systems [Huang, 91; Matthews, 94; Vassella, 94; Berberof, 95; Schürch, 97; Hehl, 94] would give further information on nucleotide motifs involved in regulation. Indeed, gene transcription and translation regulation is poorly understood in trypanosomes and appears not to depend on consensus sequences. Similarly, once the H1 gene cluster(s) are mapped, the 5' sequence of MNNTT and MAKTT genes could also be experimented.

The potential justification of such kind of research which proposes that the different genome organization and processing as well as the disparities found between histones of higher eucaryotes and trypanosomes could open up new targets for parasite control remains

valuable [Hecker, 94; Galanti, 98]. However, it appears difficult to be able to gain access to such intimate constituents inside the nucleus and H1 has not yet been shown to be a central element for survival. On the other hand, the advances made in stable antisense molecules production at the level of therapeutic integration for potential treatment of several metabolic diseases (psoriasis, Alzheimer…), viral problems (AIDS, hepatitis…) or cancer [Forster, 97] could represent such kind of possibilities. It may be that a reduced H1 expression is not critical for cell survival but rather on population kinetics and differentiation into stumpy forms. The question if H1 is implicated in the arrest of stumpy forms in G0/G1 could be very important in a parasitological view. It was shown that only stumpy forms which are non-proliferating and cell-cycle arrested cells can differentiate promptly into procyclic forms [Matthews, 94]. The function of H1 in the interplay between the cell-cycle and the parasitic life-cycle should be investigated in details.

As a general conclusion, it could be said that this work opens investigation ways which were not possible without the knowledge of H1 gene structure and organization. A better understanding of histones functionality in trypanosomes remains of great interest principally because they constitute one of the most central element in living cells. In trypanosomes, histone H1 expression has been shown to be regulated at different levels, including cell-cycle and life-stages. Therefore, knowledge of their genes, their organization and surrounding sequences give, beside other gene systems that were described, an additional model for the understanding of mechanisms which lead to expression at precise moments. Furthermore, in the particular case of histones, the gene products are going to wrap the genetic information and thereby potentially influence all the other cellular mechanisms at their origin.

VI. REFERENCES

Adachi Y, Käs e, Laemmli U K. Preferential cooperative binding of DNA topoisomerase II to scaffold associated regions. EMBO J 1989; 8 (13): 3997-4006.

Allsopp R, Hall D and Jones T. Fatal attraction for the tsetse fly. New Sci. 1985; 108: 40-43.

Asbroek A, Ouellette M, Borst P. Targeted insertion of the neomycin phosphotransferase gene into the tubulin gene cluster of *Trypanosoma brucei*. Nature 1990; 348: 174-5.

Åslund L, Carlsson L, Henriksson J et al. A gene family encoding heterogeneous histone H1 proteins in *Trypanosoma cruzi*. Mol Biochem Parasitol 1994 ; 65 : 317-30

Ausubel F M, Brent R, Kingston R E et al. Current protocols in molecular biology. John Wiley & sons Inc. 1994.

Barnes D A, Mottram J, Selkirk M, Agabian N. Two variant surface glycoprotein genes distinguish between different substrains of *Trypanosoma brucei gambiense*. Mol Biochem Parasitol 1989; 34: 135-46.

Barry III C E, Hayes S F, hackstadt T. Nucleotid condensation in *Escherichia coli* that express a chlamydial histone homolog. Science 1992; 256: 377-79.

Barry J D, Crowe J S, Vickerman K. Instability of the *Trypanosoma brucei rhodesiense* metacyclic variable antigen repertoire. Nature 1983; 306 (5944): 699-701.

Barry J D. The relative significance of mechanisms of antigenic variation in African trypanosomes. Parasitol Today 1997; 13 (6): 212-18.

Belli S, Formenton A, Noll T et al. *Leishmania major*: Histone H1 gene expression from the *sw3* locus. Exp Parasitol 1999; 91: 151-160.

Bender K, Betschart B, Hecker H. Histone-DNA interactions in the chromatin of procyclic *Trypanosoma brucei brucei.* Parasitol Res 1992(b); 78: 495-500.

Bender K, Betschart B, Marion C, Michalon P, Hecker H. Structural differencies between the chromatin of procyclic *Trypanosoma brucei brucei* and of higher eucaryotes as probed by immobilized trypsin. Acta Tropica 1992(c); 52; 69-78.

Bender K, Betschart B, Schaller J, Kämpfer U, Hecker H. Biochemical properties of histone-like proteins of procyclic *Trypanosoma brucei brucei.* Acta Tropica 1991; 50: 169-84.

Bender K, Betschart B, Schaller J, Kämpfer U, Hecker H. Sequence differencies between histones of procyclic *Trypanosoma brucei brucei* and higher eucaryotes. Parasitology 1992(a); 105; 97-104.

Berberof M, VanHamme L, Tebabi P et al. The 3'-terminal region of the mRNAs for VSG and procyclin can confer stage specificity to gene expression in *Trypanosoma brucei*. EMBO J 1995; 14: 2925-34.

Biebinger S, Rettenmaier S, Flaspohler J et al. The PARP promoter of *Trypanosoma brucei* is developmentally regulated in a chromosomal context. Nucl Acids Res 1996; 24: 1202-11.

Biebinger S, Wirtz L E, Lorenz P, Clayton C. Vectors for inducible expression of toxic gene products in bloodstream and procyclic *Trypanosoma brucei*. Mol Biochem Parasitol 1997; 85: 99-112.

Bontempi E J, Porcel B M, Henriksson J et al. Genes for histone H3 in *Trypanosoma cruzi*. Mol Biochem Parasitol 1994; 66: 147-51.

Bradbury E M, Inglis R J, Matthews H R. Control of cell division by very lysine rich histone (F1) phosphorylation. Nature 1974; 247: 257-61.

Bringaud F and Baltz T. Differential regulation of two distinct families of glucose transporter genes in *Trypanosoma brucei*. Mol Cel Biol 1993; 1146-54.

Brossard P. Localisation de l'histone H1 chez *Trypanosoma brucei*. Diplôme de biologie, Institut de zoologie, Université de Neuchâtel; 1998.

Burri M, Schlimme W, Betschart B, Hecker H. Characterization of the histones of *Trypanosoma brucei brucei* bloodstream forms. Acta Tropica 1994;.58: 291-305.

Burri M, Schlimme W, Betschart B, Kämpfer U, Schaller J, Hecker H. Biochemical and functional characterization of histone H1-like proteins in procyclic *Trypanosoma brucei brucei*. Parasitol Res 1993; 79: 649-59.

Burri M, Schlimme W, Betschart B, Lindner H, Kämpfer U, Schaller J, Hecker H. Partial amino acid sequence and functional aspects of histone H1 proteins in *Trypanosoma brucei brucei*. Biol Cell 1995; 83: 23-31.

Churchill M E A, Suzuki M. "SPKK" motifs prefer to bind to DNA at A/T-rich sites. EMBO J 1989; 8: 4189-95.

Cunningham I. New culture medium for maintenance of tsetse tissues and growth of trypomastigotes. J Protozool 1977; 24 (2): 325-9.

Dasso M, Dimitrov S, Wolffe A P. Nuclear assembly is independent of linker histone. Proc Natl Acad Sci USA 1994; 91: 12477-81.

De Souza W and Meyer H. On the fine structure of the nucleus in *Trypanosoma cruzi* in tissue culture forms. Spindle fibers in the dividing nucleus. J Protozool 1974; 21 (1): 48-52.

Denise K, Matthews K, Lindergard G, Croft S, Barrett M P. Trypanosomiasis and leishmaniasis: between the idea and the reality of control. Parasitol Today 1999; 15 (2): 43-45.

Devereux J, Haeberli P, Smithies O. A comprehensive set of sequence analysis programs for the VAX. Nucl Acids Res 1984; 12(1): 387-95.

Du Y, Maslov D A, Chang K-P. Monophyletic origin of β-division proteobacterial endosymbionts and their coevolution with insect trypanosomatid protozoa *Blastocrithidia culicis* and *Crithidia* spp. Proc Natl Acad Sci USA 1994; 91: 8437-41.

Duschak V G, Cazzulo J J. The histones of the insect trypanosomatid, Crithidia fasciculata. Biochim Biophys Acta 1990; 1040(2): 159-66.

Eid J, Sollner-Webb B. Stable integrative transformation of *Trypanosoma brucei* that occurs exclusively by homologous recombination. Proc Natl Acad SCI USA 1991; 88: 2118-21.

El-Sayed, N M A, Alarcon C M, Beck J C, Sheffield V C, Donelson J E. cDNA expressed sequence tags of *Trypanosoma brucei rhodesiense* provide new insight into the biology of the parasite. Mol Biochem Parasitol 1995; 73: 75-90.

Erondu N E, Donelson J E. Differential expression of two mRNAs from a single gene encoding an HMG1-like DNA binding protein of African trypanosomes. Mol Biochem Parasitol 1992 ; 51 : 111-18.

Ersfeld K, Docherty R, Alsford S and Gull K. A fluorescence in situ hybridisation study of the regulation of histone mRNA levels during the cell cycle of *Trypanosoma brucei*. Mol Biochem Parasitol 1996; 81: 201-9.

Ersfeld K, Melville S E, Gull K. Nuclear and genome organization of *Trypanosoma brucei*. Parasitol Today 1999; 15: 58-63.

Fasel N J, Robyr D C, Mauel J, Glaser T A. Identification of a histone H1-like gene in *Leishmania major*. Mol Biochem Parasitol. 1994; 62: 321-4.

Felsenfeld G. Chromatin as an essential part of the transcriptional mechanism. Nature 1992; 355: 219-24.

Ferguson M, Torri A F, Ward D C, Englund P T. In situ hybridization to the *Crithidia fasciculata* kinetoplast reveals two antipodal sites involved in kinetoplast DNA replication. Cell 1992; 70: 621-29.

Forster H and Gwinner E. Die modulation der Genexpression durch Antisense. Bioworld 1997; 3:3-6.

Galanti N, Dvorak J A, Grenet J, McDaniel J P. Hydroxyurea-induced synchrony of DNA replication in the kinetoplastida. Exp Cell Res 1994; 214 (1); 225-30.

Galanti N, Galindo M, Sabaj V, Espinoza I, Toro G C. Histone genes in trypanosomatids. Parasitol Today 1998; 14(2): 64-70.

Gibson W C, Swinkels B W, Borst P. Post-transcriptional control of the differential expression of phosphoglycerate kinase genes in *Trypanosoma brucei*. J Mol Biol 1988; 201: 315-25.

Golvan Y-J. Elements de parasitologie médicale (4ème édition). Flammarion Médecine-Sciences 1983.

Graham S V, Barry J D. Polysomal, procyclin mRNAs accumulate in bloodstream forms of monomorphic and pleomorphic trypanosomes treated with protein synthesis inhibitors. Mol Biochem Parasitol 1996; 80: 179-91.

Graham S V. Mechanisms of stage-regulated gene expression in Kinetoplastida. Parasitology Today 1995; 11: 217-23.

Gray M W. Organelle origins and ribosomal RNA. Biochem Cell Biol 1988; 66: 325-48.

Grunstein M. Histones as regulators of genes. Scientific American 1992; october: 40-47.

Haag J, O'h Uigin C, Overath P. The molecular phylogeny of trypanosomes: evidence for an early divergence of the salivaria. Mol Biochem Parasitol. 1998; 91: 37-49.

Hackstadt T, Baehr W, Ying Y. *Chlamydia trachomatis* developmentally regulated protein is homologous to eucaryotic histone H1. Proc. Natl. Acad. Sci. USA 1991; 88: 3937-3941.

Hecker H and Gander E S. The compaction pattern of the chromatin of Trypanosomes. Biol Cell 1985; 53: 199-208.

Hecker H, Bender K, Betschart B, Modespacher U P. Instability of the nuclear chromatin of procyclic *Trypanosoma brucei brucei*. Mol Biochem Parasitol 1989; 37: 225-34.

Hecker H, Betschart B, Bender K, Burri M, Schlimme W. The chromatin of Trypanosomes. Int J Parasitol 1994; 24 (6): 809-19.

Hecker H, Betschart B, Burri M, Schlimme W. Functional morphology of Trypanosome chromatin. Parasitology Today 1995; 11 (2): 79-83.

Hecker H. Man and Sea Urchin - More closely related than African and American Trypanosomes? Parasitology Today 1993; 9 (2), 57.

Hehl A, Vassella E, Braun R, Roditi I. A conserved stem-loop structure in the 3' untranslated region of procyclin mRNAs regulates expression in *Trypanosoma brucei*. Proc Natl Acad Sci. 1994 ; 91 : 370-4.

Henriksson J, Aslund L, Petersson U. Karyotype variability in *Trypanosoma cruzi*. Parasitology Today 1996; 12: 108-14.

Hide G, Mottram J C, Coombs G H, Holmes P H. Trypanosomiasis and leishmaniasis: biology and control. CAB International; University Press Cambridge 1997.

Hill C S, Packman L C, Thomas J O. Phosphorylation at clustered -Ser-Pro X-Lys/Arg- motifs in sperm-specific histones H1 and H2B. EMBO J 1990; 9: 805-13.

Hill C S, Rimmer J M, Green B N, Finch J T, Thomas J O. Histone-DNA interactions and their modulation by phosphorylation of Ser-Pro-X-Lys/Arg- motifs. EMBO J 1991; 10: 1939-48.

Hirschy A. Reconnaissance de l'histone H1 de *Trypanosoma brucei brucei* par la production d'anticorps polyclonaux: quelques essais. Diplôme de biologie, Institut de zoologie, Université de Neuchâtel; 1998.

Hirst S I and Stapley L A. Parasitology: the dawn of a new Millenium. Parasitol Today 2000; 16 (1): 1-3.

Hohmann P. Phosphorylation of H1 histones. Mol Cell Biochem 1983; 57: 81-92.

Horn D, Cross G A M. Position-dependent and promoter-specific regulation of gene expression in *Trypanosoma brucei*. EMBO J. 1997; 16: 7422-31.

Huang J, Van der Ploeg L H T. Requirement of a polypyrimidine tract for trans-splicing in trypanosomes : discriminating the PARP promoter from the immediately adjacent 3' splice acceptor site. EMBO J 1991 ; 10 : 3877-85.

Hynek M. Isolation et séquençage de clones d'ADNc codants pour l'histone H1 chez *Trypanosoma brucei brucei* GARP 16. Travail de Certificat en Parasitologie, Institut de zoologie, Université de Neuchâtel; 1999.

Ivanchenko M, Avramova Z. Interaction of MAR-sequences with nuclear matrix proteins. J Cell Biochem 1992; 50 (2): 190-200.

Izaurralde E, Kas E, Laemmli U K. Highly preferential nucleation of histone H1 assembly on scaffold associated regions. J Mol Biol 1989; 210 (3): 573-85.

Jenni L, Marti S, Schweizer J et al. Hybrid formation between African trypanosomes during cyclical transmission. Nature 1986; 322 (6075): 173-175.

Johmann C A, Gorovsky M A. Purification and characterization of the histones associated with the macronucleus of Tetrahymena. Biochemystry 1976; 15(6): 1249-56.

Katz M, Despommier D D, Gwadz R W. Parasitic diseases (second edition). Springer Verlag 1989.

Kohl L, Gull K. Molecular architecture of the trypanosome cytoskeleton. Mol Biochem Parasitol 1998; 93: 1-9.

Kuzoe F A S. Current situation of African trypanosomiasis. Acta Tropica 1993; 54: 153-62.

Lanham S M. Isolation of salivarian trypanosomes from Man and other mammals using DEAE-cellulose. Exp. Parasitol 1970; 28: 521-534.

Lee M G-S and Van der Ploeg L H T. Homologous recombination and stable transfection in the parasitic protozoan *Trypanosoma brucei*. Science 1990; 250: 1583-87.

Lukes J and Maslov D A. Unexpectedly high variability of the histone H4 gene in leishmania. Parasitol Res 2000; 86 (3): 259-61.

Marchuk D, Drumm M, Saulino A, Collins F S. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. Nucl Acids Res 1991; 19: 1154.

Marras S A, Kramer F R, Tyagi S. Multiplex detection of single nucleotide variations using molecular beacons. Genet Anal 1999; 14 (5-6): 151-56.

Maslov D A, Lukeš J, Jirku M, Simpson L. Phylogeny of trypanosomes as inferred from the small and large subunit rRNAs: implications for the evolution of parasitism in the trypanosomatid protozoa. Mol Biochem Parasitol 1996; 75: 197-205.

Matthews K R, Gull K. Evidence for an interplay between cell cycle progression and the initiation of differentiation between life cycle forms of African trypanosomes. J Cell Biol 1994; 125:1147-56.

Matthews K R, Tschudi C, Ullu E. A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. Genes Dev 1994; 8: 491-501.

Melville S E, Majiva P, Donelson J. Resources available from the African Genome Project. Parasitology Today 1998(a); 14: 3-4.

Melville S E. The African trypanosome Genome Project: focus on the future. Parasitology Today 1998(b); 14: 129-31.

Multigner L, gagnon J, Van Dorsselaer A, Job D. Stabilization of sea urchin flagellar microtubules by histone H1. Nature 1992; 360: 33-9.

Noll T M, Desponds C, Belli S I, Glaser T A, Fasel N J. Histone H1 expression varies during the *Leishmania major* life cycle. Mol Biochem Parasitol 1997; 84: 215-27.

Ohsumi K, Katagiri C, Kishimoto T. Chromosome condensation in Xenopus mitotic extracts without histone H1. Science 1993; 262: 2033-5.

Pays E, Coquelet H, Tebabi P et al. *Trypanosoma brucei*: constitutive activity of the VSG and procyclin gene promoters. EMBO J. 1990; 9: 3145-51.

Pays E, Nolan D P. Expression and function of surface proteins in *Trypanosoma brucei*. Mol Biochem Parasitol 1998; 91: 3-36.

Pruss D, Hayes J J, Wolffe A P. Nucleosomal assembly – where are the histones? Bioessays 1995; 17(2): 161-70.

Puerta C, Martin J, Alonso C, López M C. Isolation and characterization of the gene encoding histone H2A from *Trypanosoma cruzi*. Mol Biochem Parasitol 1994; 64: 1-10.

Reuner B, Vassella E, Yutzy B, Boshart M. Cell density triggers slender to stumpy differentiation of Trypanosoma brucei bloodstream forms in culture. Mol Biochem Parasitol 1997; 90 (1): 269-80.

Revelard P, Lips S, Pays E. Alternative splicing within and between alleles of the ATPase gene 1 locus of *Trypanosoma brucei*. Mol Biochem Parasitol 1993 ; 62 : 93-102.

Roditi I, Pearson T W. The procyclin coat of African trypanosomes. Parasitol Today 1990; 6: 79-82.

Roditi I, Schwarz H, Pearson T W et al. Procyclin gene expression and loss of the variant surface glycoprotein during differentiation of *Trypanosoma brucei*. J Cell Biol 1989; 108: 737-46.

Roditi I. The VSG-Procyclin switch. Parasitology Today 1996; 12: 47-9.

Roth S Y, Allis C D. Chromatin condensation: does histone H1 dephosphorylation play a role? Trends Biochem Sci 1992; 17(3): 93-8.

Sabaj V, Díaz J, Toro G C, Galanti N. Histone synthesis in *Trypanosoma cruzi*. Experimental Cell Res. 1997; 236: 446-52.

Sambrook J, Fritsch E F, Maniatis T. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory Press, Second Edition 1989.

Sanders C. A method for the fractionation of the high-mobility group non-histone proteins. Biochem Biophys Res Comm 1977; 78: 1034-42.

Scarlato V, Aricò B, Goyard S et al. A novel chromatin-forming histone H1 homologue is encoded by a dispensable and growth-regulated gene in *Bordetella pertussis*. Mol Microbiol 1995; 15: 871-81.

Schlimme W, Burri M, Bender K, Betschart B, Hecker H. *Trypanosoma brucei brucei*: differences in the nuclear chromatin of bloodstream forms and procyclic culture forms. Parasitology 1993; 107: 237-47.

Schlimme W, Burri M, Betschart B, Hecker H. Properties of the histones and functional aspects of the soluble chromatin of epimastigote *Trypanosoma cruzi*. Acta Tropica 1995; 60: 141-54.

Schmidt G D and Roberts L S. Foundations of parasitology (3 Edition). Times Mirror/Mosby College Publishing. 1985.

Schulze E, Trieschmann L, Schulze B, et al. Structural and functional differences between histone H1 sequence variants with differential intranuclear distribution. Proc Natl Acad Sci USA 1993; 90(6): 2481-5.

Schürch N, Furger A, Kurath U, Roditi I. Contributions of the procyclin 3' untranslated region and coding region to the regulation of expression in bloodstream forms of *Trypanosoma brucei*. Mol Biochem Parasitol 1997 ; 89 : 109-21.

Shapiro S Z and Doxsey S J. Purification of nuclei from a flagellate protozoan, *Trypanosoma brucei*. Anal Biochem 1982; 127: 112-15.

Shen X, Gorovsky M A. Linker histone H1 regulates specific gene expression but not global transcription in vivo. Cell 1996; 86(83): 475-83.

Shen X, Yu L, Weir J W, Gorovsky M A. Linker histones are not essential and affect chromatin condensation in vivo. Cell 1995; 82(1): 47-56.

Solari A J. The 3-dimensional fine structure of the mitotic spindle in *Trypanosoma cruzi*. Chromosoma 1980; 78: 239-55.

Soto M, Quijada L, Alonso C, Requena J M. Molecular cloning and analysis of expression of the *Leishmania infantum* histone H4 genes. Mol Biochem Parasitol 1997; 90 (2): 439-47.

Spinoza I, Toro G C, Hellman U, Galanti N. Histone H1 and core histones in Leishmania and Crithidia: comparison with Trypanosoma. Exp Cell Res 1996; 224: 1-7.

Suzuki M. SPKK, a new nucleic acid-binding unit of protein found in histone. EMBO J 1989; 8: 797-804.

Tait A, Buchanan N, Hide G, Michael C, Turner R. Self-fertilization in *Trypanosoma brucei*. Mol Biochem Parasitol 1996; 76: 31-42.

Thatcher T H, Gorovsky M A. Phylogenetic analysis of the core histones H2A, H2B, H3 and H4. Nucl Acids Res 1994; 22: 174-9.

Thoma F, Koller T, Klug A. Involvement of histone H1 in the organization of the nucleosome and the salt-dependent superstructures of chromatin. J Cell Biol 1979; 83: 403-27.

Thomas J O. The higher order structure of chromatin and histone H1. J Cell Sci Suppl 1984; 1: 1-20.

Thomas M C, Olivares M, Escalante M et al. Plasticity of the histone H2A genes in a Brazilian and six Colombian strains of *Trypanosoma cruzi*. Acta Tropica 2000; 75: 203-10.

Thompson J D, Higgins D G, Gibson T J. Improving the sensitivity of progressive multiple sequence alignment through sequence weighing, positions-specific gap penalties and weight matrix choice. Nucl Acid Res 1994; 22: 4673-4680.

Toro G C and Galanti N. Further characterization of *Trypanosoma cruzi* histones. Biochem Int 1990; 21: 481-490.

Toro G C and Galanti N. Histone H1 and histone variants in *Trypanosoma cruzi*. Exp Cell Res 1988; 174: 16-24.

Toro G C, Galanti N, Hellman U, Wernstedt C. Unambiguous identification of histone H1 in *Trypanosoma cruzi*. J Cell Biochem 1993; 52: 431-9.

Toro G C, Wernstedt C, Medina C, Jaramillo N, Hellman U, Galanti N. Extremely divergent histone H4 sequence from *Trypanosoma cruzi*: evolutionary implications. J Cell Biochem 1992; 49: 266-71.

Trieshmann L, Schulze E, Schulze B, Grossbach U. The histone H1 genes of the dipteran insect, *Chironomus thummi*, fall under two divergent classes and encode proteins with distinct intranuclear distribution and potentially different functions. Eur J Biochem 1997; 250(1): 184-96.

Van der Ploeg L H T, Valerio D, De Lange T, Bernards A, Borst P, Grosveld F G. An analysis of cosmid clones of nuclear DNA from *Trypanosoma brucei* shows that the genes for variant surface glycoproteins are clustered in the genome. Nucl Acids Res 1982; 10: 5905-23.

Van Holde K E. Chromatin. Springer-Verlag New York Inc. 1989.

Van Holde K, Zlatanova J. Chromatin architectural proteins and transcription factors. A structural connection. Bioessays 1996(b); 18(9): 697-700.

Van Holde K, Zlatanova J. The nucleosome core particle: does it have structural and physiologic relevance? Bioessays 1999; 21(9): 776-80.

Van Holde K, Zlatanova J. What determines the folding of the chromatin fiber? Proc Natl Acad Sci USA 1996(a); 93(20): 10548-55.

Vanhamme L, Pays E. Control of gene expression in Trypanosomes. Microbiol Rev 1995 ; 59 : 223-40.

Vassella E, Braun R, Roditi I. Control of polyadenylation and alternative splicing of transcripts from adjacent genes in a procyclin expression site : a dual role for polypyrimidine tracts in trypanosomes? Nucleic Acids Res 1994; 22: 1359-64.

Vassella E, Reuner B, Yutzy B, Boshart M. Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. J Cell Sci 1997; 110 (Pt21): 2661-71.

Vickermann K, Preston T M. Spindle microtubules in the dividing nuclei of trypanosomes. J Cell Sci 1970; 6: 365-83.

Wells D and Kedes L. Structure of a human histone cDNA: evidence that basally expressed histone genes have intervening sequences and encode polyadenylylated mRNAs. Proc Natl Acad Sci USA 1985; 82: 2834-38.

Widom J. Chromatin structure: Linking structure to function with histone H1. Current Biology 1998; 8: R788-R791.

Woese C R, Kandler O, Wheelis M L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eucarya. Proc Natl Acad Sci USA 1990; 87: 4576-9.

Wolffe A P, Khochbin S, Dimitrov S. What do linker histones do in chromatin? Bioessays 1997; 19(3): 249-55.

Wolffe A P. Transcription: in tune with the histones. Cell 1994; 77: 13-16.

Wolffe A. Chromatin structure and function. Academic Press, Harcourt Brace Jovanovich Publishers 1992.

Zlatanova J, van Holde K. The linker histones and chromatin structure: new twists. Prog Nucleic Acid Res Mol Biol 1996; 52: 217-59.

Zu T, Goyard S, Rappuoli R, Scarlato V. DNA binding of the *Bordetella pertussis* H1 homolog alters in vitro DNA flexibility. J Bacteriol 1996; 178: 2982-5.

VII. ANNEX

ANNEX A

***Trypanosoma brucei gambiense* (strain 1257) cDNA clones:**

AJ272459	TbgH1A85	(cDNA clone c1.1)
AJ272460	TbgH1B81	(cDNA clone c2.4)
AJ272461	TbgH1A81	(cDNA clone c8.1)
AJ272462	TbgH1A76	(cDNA clone c7.9)

***Trypanosoma brucei gambiense* (strain STIB755) genomic clones:**

AJ287591	TbgH1B85
AJ287592	TbgH1A75
AJ287593	TbgH1A91
AJ287594	TbgH1A71
AJ287595	TbgH1B76
AJ287596	TbgH1C81
AJ287597	TbgH1A61

***Trypanosoma brucei brucei* (strain TREU927/4) genomic clones:**

AJ287598	TbbH1A80 and TbbH1A76
AJ287599	TbbH1B76 and TbbH1C76
AJ287600	TbbH1D76 and TbbH1A71
AJ287601	TbbH1F76
AJ287602	TbbH1B71
AJ287603	TbbH1A61
AJ400880	TbbH1H76 and TbbH1D71
AJ400881	TbbH1I76 and TbbH1E71

***Trypanosoma brucei brucei* (strain TREU927/4) cDNA clones:**

AJ287604	TbbH1B80
AJ287605	TbbH1E76
AJ287606	TbbH1C71

***Trypanosoma brucei brucei* (strain 427) cDNA clones:**

AJ287607	TbbH1G76
AJ287608	TbbH1A66

Remark: All sequences that are discussed in the present work were submitted to the EMBL Nucleotide Sequence Database. Some sequences may have identical amino acid sequences, differences being on the level of nucleotide sequences, neighboring ORFs or origin (subspecies or strains).

ANNEX B

A

```
TbgH1B85        CTTTATCGACTCCCCACAAGAATGAACAACACAACCGATACTGTGAAGGCCACCCCAAAG
TbgH1A75        CTTTATCGACTCCCCACAAGAATGAACAACACAACCGATACTGTGAAGGCCACCCCAAAG
                ************************************************************

TbgH1B85        AAGGTTGCAGCCAAAAAGGCTGTCGCTAAGAAGGCTGCTCCAAAGAAGGCTGTCGCCAAG
TbgH1A75        AAGGTTGCAGCCAAAAAGGCTGTCGCTAAGAAGGCTGCTCCAAAGAAGGCTGTCGCTAAG
                ************************************************************ ***

TbgH1B85        AAGGCTGCTCCCAAAAAGGCTGTCGCTAAGAAGGCTGCTCCCAAAAAGGCTGTCGCCAAG
TbgH1A75        AAGGCTGCTCCCAAAAAGGCTGTCGCTAAGAAGGCTGCTCCCAAAAAGGCTGTCGCCAAG
                ************************************************************

TbgH1B85        AAGGTTGTCGCCAAAAAGGTTGCCCCCAAGAAGGTTGTCGCCAAAAAGGTTGCCCCCAAG
TbgH1A75        AAGGTTGTCGCCAAAAAGGTTGCCCCCAAGAAG--------------------------
                ********************************

TbgH1B85        AAGGTTGCCGGCAAGAAGGCCGCCGCTAAGAAGGCGTGAGCGCATCCGCTGCTGCCCGCT
TbgH1A75        ---GTTGCCGGCAAGAAGGCCGCCGCTAAGAAGGCGTGAGCGCATCCGCTGCTGCCCGCT
                   ************************************************************

TbgH1B85        ATTAGACAAGCTATGAGGTTTACC------
TbgH1A75        ATTAGACAAGCTATGAGGTTTACCTGAGTG
                ************************
```

Annex B: Complete nucleotide sequence of all variants of *T. b. gambiense* (strain STIB 755) submitted to GenBank. **A:** Sequences obtained with primers 5'-MNNTT / H1orf3'. **B:** Sequences obtained with primers 5'-MAKAS / H1orf3' (next page).

B

```
TbgH1A71          CAATCTTATCAACACTCGGAAGTATGGCGAAGGCATCTGCTGCTCCCAAGAAAGCTGTG
TbgH1B76          CAATCTTATCAACACTCGGAAGTATGGCGAAGGCATCTGCTGCTCCCAAAAAGGCTGTC
TbgH1C81          CAATCTTATCAACACTCGGAAGTATGGCGAAGGCATCTGCTGCTCCCAAGAAAGCTGTG
TbgH1A91          CAATCTTATCAACACTCGGAAGTATGGCGAAGGCATCTGCTGCTCCCAAAAAGGCTGTC
TbgH1A61          CAATCTTATCAACACTCGGAAGTATGGCGAAGGCATCTGCTGCTCCCAAAAAGGCTGTC
                  ********************************************** ** *****

TbgH1A71        GCTAAGAAGGCAGCCCCCAAAAAGGCTGTCGCTAAGAAGGCTGCTCCAAAGAAGGCTGTC
TbgH1B76        GCTAAGAAGGCAGCTCCCAAGAAAGCTGTGGCTAAGAAGGCTGCTCCAAAGAAGGCTGTC
TbgH1C81        GCTAAGAAGGCAGCCCCCAAAAAGGCTGTCGCTAAGAAGGCTGCTCCAAAGAAGGCTGTC
TbgH1A91        GCTAAGAAGGCAGCCCCCAAAAAGGCTGTCGCTAAGAAGGCTGCTCCAAAGAAGGCTGTC
TbgH1A61        GCTAAGAAGGCTGCTCCCAAAAAGGCTGTTGCCAAGAAG--------------------
                ********** ** ***** ** ***** ** ******

TbgH1A71        GCTAAGAAGGCAGCCCCCAAAAAGGCTGTCGCTAAGAAGGCTGCTCCC------------
TbgH1B76        GCTAAGAAGGCTGCTCCAAAGAAGGCTGTCGCTAAGAAGGCTGCTCCCAAAAAGGCTGTC
TbgH1C81        GCTAAGAAGGCAGCCCCCAAAAAGGCTGTCGCTAAGAAGGCTGCTCCCAAAAAGGCTGTC
TbgH1A91        GCTAAGAAGGCAGCCCCCAAAAAGGCTGTCGCTAAGAAGGCTGCTCCCAAAAAGGCTGTT
TbgH1A61        ------------------------------------------------------------

TbgH1A71        ------------------AAAAAGGCTGCCCCC---------------------------
TbgH1B76        GCT---------------AAGAAGGCTGCTCCC---------------------------
TbgH1C81        GCTAAGAAGGCTGCTCCCAAAAAGGCTGCCCCC---------------------------
TbgH1A91        GCCAAGAAG-----------------------GTTGCCGCCACGAAACCTTTAGCCAAG
TbgH1A61        ----------------------------------GTTGCCGCCACGAAACCTTTAGCCAAG

TbgH1A71        --------------------------AAGAAGGTTGCCCCCAAGAAGGTTGCCGGCAAG
TbgH1B76        --------------------------AAAAAGGTTGCCCCCAAGAAGGTTGCCGGCAAG
TbgH1C81        --------------------------AAGAAGGTTGCCCCCAAGAAGGTTGCCGGCAAG
TbgH1A91        AAGGTTGCCGGCAAGAAGGTTGTCGCCAAAAAGGTTGCCCCCAAGAAGGTTGCCGGCAAG
TbgH1A61        AAGGTTGCCGGCAAGAAGGTTGTCGCCAAAAAGGTTGCCCCCAAGAAGGTTGCCGGCAAG
                                              ***************************

TbgH1A71        AAGGCCGCCGCTAAGAAGGCGTGAGCGCATCCGCTGCTGCCCGCTATT---AGACACGCT
TbgH1B76        AAGGCCGCCGCTAAGAAGGCGTGAGCGCATCCGCTGCTGCCCGCTATT---AGACAAGCT
TbgH1C81        AAGGCCGCCGCTAAGAAGGCGTGAGCGCATCCGCTGCTGCCCGCTATT---AGACACGCT
TbgH1A91        AAGGCCGCCGCTAAGAAGGCGTGAGCGCATCCGCTGCTGCCCGCTATTATTAGACACGCT
TbgH1A61        AAGGCCGCCGCTAAGAAGGCGTGAGCGCATCCGCTGCTGCCCGCTATTATTAGACACGCT
                ********************************************    **** ***

TbgH1A71        ATGAGGTTTACCTGAGTG
TbgH1B76        ATGAGGTTTACCTGAGTG
TbgH1C81        ATGAGGTTTACCTGAGTG
TbgH1A91        ATGAGGTTTACCTGAGTG
TbgH1A61        ATGAGGTTTACCTGAGTG
                ******************
```

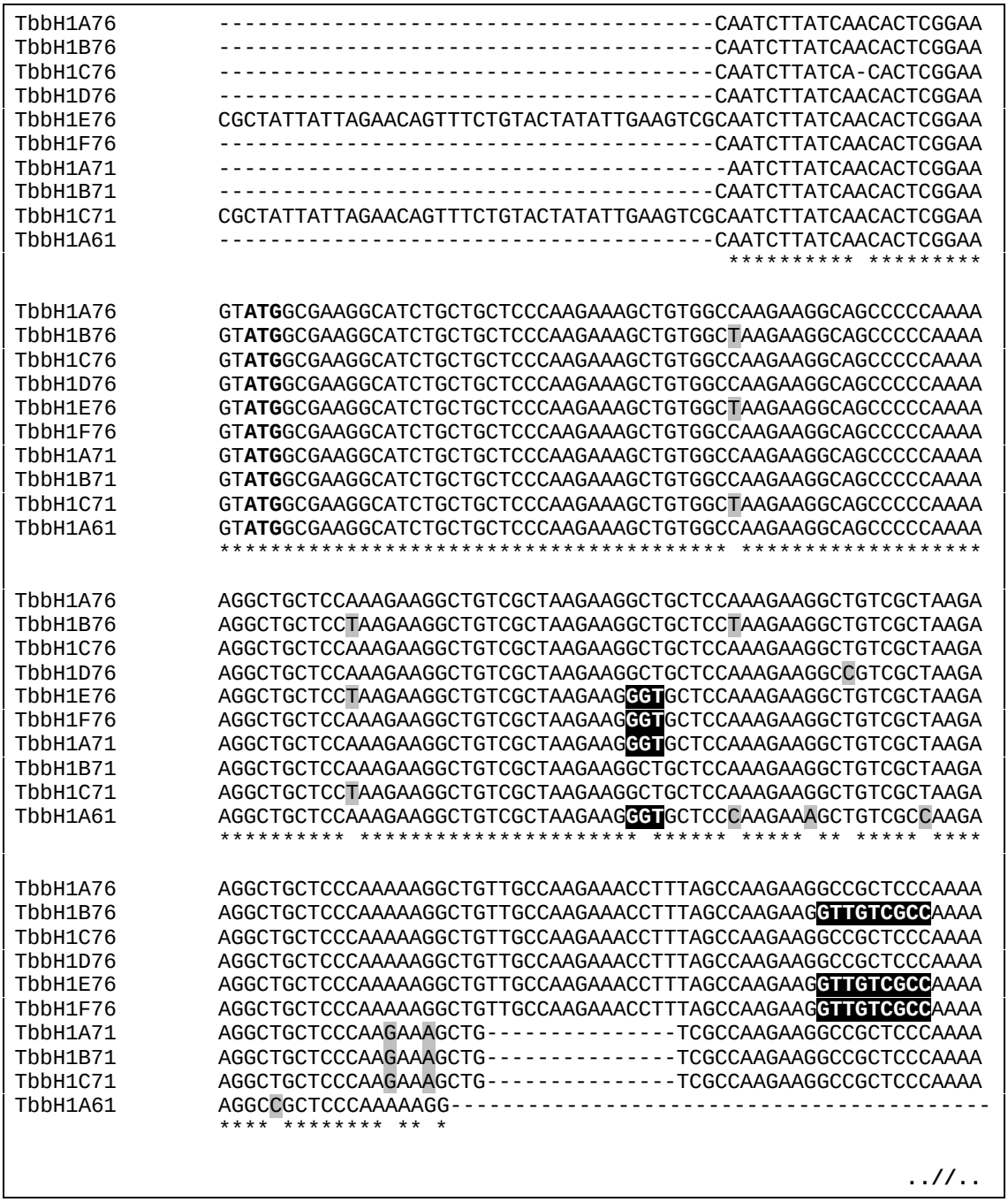
ANNEX C

A

```
TbbH1A80        ------------------------------------CTTTATCGACTCCCCACAAGA
TbbH1B80        CGCTATTATTAGAACAGTTTCTGTACTATATTGAAAGCTCTTTATCGACTCCCCACAAGA
                                                    ********************

TbbH1A80        ATGAACAACACAACCGCTACTGTGAAGGCCACCCCAAAGAAGGTTGCAGCCAAAAAGGCC
TbbH1B80        ATGAACAACACAACCGCTACTGTGAAGGCCACCCCAAAGAAGGTTGCAGCCAAAAAGGCC
                ************************************************************

TbbH1A80        GCTCCAAAGAAAACTGTGGCCAAAAAGGCTGCTCCAAAGAAGGCTGTTGCCAAGAAACCT
TbbH1B80        GCTCCAAAGAAAACTGTGGCCAAAAAGGCTGCTCCAAAGAAGGCTGTTGCCAAGAAACCT
                ************************************************************

TbbH1A80        TTAGCTAAGAAGGTTGTCGCTAAGAAGGCTGCTCCCAAGAAAGCTGTCGCCAAGAAGGCC
TbbH1B80        TTAGCTAAGAAGGTTGTCGCTAAGAAGGCTGCTCCCAAGAAAGCTGTCGCCAAGAAGGCC
                ************************************************************

TbbH1A80        GCTCCCAAAAAGGTTGCCCCGAAGAAGGTTGCCGGCAAGAAGGCCGCCGCTAAGAAGGCG
TbbH1B80        GCTCCCAAAAAAGTTGCCCCGAAGAAGGTTGCCGGCAAGAAGGCCGCCGCTAAGAAGGCG
                *********** ************************************************

TbbH1A80        TGAGCGCATCCGCTGCTGCCCGCTATTAGACACGCTATGAGGTTTACCTGAGTG
TbbH1B80        TGAGCGCATCCGCTGCTGCCCGCTATTAGACACGCTATGAGGTTTACCTGAGTG
                ******************************************************
```

Annex C: Complete nucleotide sequence of all variants of *T. b. brucei* (strain TREU 927/4) submitted to GenBank. **A:** Sequences obtained with primers 5'-MNNTT / H1orf3'. **B:** Sequences obtained with primers 5'-MAKAS / H1orf3' (next two page).

B


TbbH1A76 ----------CAATCTTATCAACACTCGGAA
TbbH1B76 ----------CAATCTTATCAACACTCGGAA
TbbH1C76 ----------CAATCTTATCA-CACTCGGAA
TbbH1D76 ----------CAATCTTATCAACACTCGGAA
TbbH1E76 CGCTATTATTAGAACAGTTTCTGTACTATATTGAAGTCGCAATCTTATCAACACTCGGAA
TbbH1F76 ----------CAATCTTATCAACACTCGGAA
TbbH1A71 ----------AATCTTATCAACACTCGGAA
TbbH1B71 ----------CAATCTTATCAACACTCGGAA
TbbH1C71 CGCTATTATTAGAACAGTTTCTGTACTATATTGAAGTCGCAATCTTATCAACACTCGGAA
TbbH1A61 ----------CAATCTTATCAACACTCGGAA
TbbH1A76 GTATGGCGAAGGCATCTGCTGCTCCCAAGAAAGCTGTGGCCAAGAAGGCAGCCCCCAAAA
TbbH1B76 GTATGGCGAAGGCATCTGCTGCTCCCAAGAAAGCTGTGGCTAAGAAGGCAGCCCCCAAAA
TbbH1C76 GTATGGCGAAGGCATCTGCTGCTCCCAAGAAAGCTGTGGCCAAGAAGGCAGCCCCCAAAA
TbbH1D76 GTATGGCGAAGGCATCTGCTGCTCCCAAGAAAGCTGTGGCCAAGAAGGCAGCCCCCAAAA
TbbH1E76 GTATGGCGAAGGCATCTGCTGCTCCCAAGAAAGCTGTGGCTAAGAAGGCAGCCCCCAAAA
TbbH1F76 GTATGGCGAAGGCATCTGCTGCTCCCAAGAAAGCTGTGGCCAAGAAGGCAGCCCCCAAAA
TbbH1A71 GTATGGCGAAGGCATCTGCTGCTCCCAAGAAAGCTGTGGCCAAGAAGGCAGCCCCCAAAA
TbbH1B71 GTATGGCGAAGGCATCTGCTGCTCCCAAGAAAGCTGTGGCCAAGAAGGCAGCCCCCAAAA
TbbH1C71 GTATGGCGAAGGCATCTGCTGCTCCCAAGAAAGCTGTGGCTAAGAAGGCAGCCCCCAAAA
TbbH1A61 GTATGGCGAAGGCATCTGCTGCTCCCAAGAAAGCTGTGGCCAAGAAGGCAGCCCCCAAAA
TbbH1A76 AGGCTGCTCCAAAGAAGGCTGTCGCTAAGAAGGCTGCTCCAAAGAAGGCTGTCGCTAAGA
TbbH1B76 AGGCTGCTCCTAAGAAGGCTGTCGCTAAGAAGGCTGCTCCTAAGAAGGCTGTCGCTAAGA
TbbH1C76 AGGCTGCTCCAAAGAAGGCTGTCGCTAAGAAGGCTGCTCCAAAGAAGGCTGTCGCTAAGA
TbbH1D76 AGGCTGCTCCAAAGAAGGCTGTCGCTAAGAAGGCTGCTCCAAAGAAGGCCGTCGCTAAGA
TbbH1E76 AGGCTGCTCCTAAGAAGGCTGTCGCTAAGAAGGGTGCTCCAAAGAAGGCTGTCGCTAAGA
TbbH1F76 AGGCTGCTCCAAAGAAGGCTGTCGCTAAGAAGGGTGCTCCAAAGAAGGCTGTCGCTAAGA
TbbH1A71 AGGCTGCTCCAAAGAAGGCTGTCGCTAAGAAGGGTGCTCCAAAGAAGGCTGTCGCTAAGA
TbbH1B71 AGGCTGCTCCAAAGAAGGCTGTCGCTAAGAAGGCTGCTCCAAAGAAGGCTGTCGCTAAGA
TbbH1C71 AGGCTGCTCCTAAGAAGGCTGTCGCTAAGAAGGCTGCTCCAAAGAAGGCTGTCGCTAAGA
TbbH1A61 AGGCTGCTCCAAAGAAGGCTGTCGCTAAGAAGGGTGCTCCCAAGAAAGCTGTCGCCAAGA
TbbH1A76 AGGCTGCTCCCAAAAAGGCTGTTGCCAAGAAACCTTTAGCCAAGAAGGCCGCTCCCAAAA
TbbH1B76 AGGCTGCTCCCAAAAAGGCTGTTGCCAAGAAACCTTTAGCCAAGAAGGTTGTCGCCAAAA
TbbH1C76 AGGCTGCTCCCAAAAAGGCTGTTGCCAAGAAACCTTTAGCCAAGAAGGCCGCTCCCAAAA
TbbH1D76 AGGCTGCTCCCAAAAAGGCTGTTGCCAAGAAACCTTTAGCCAAGAAGGCCGCTCCCAAAA
TbbH1E76 AGGCTGCTCCCAAAAAGGCTGTTGCCAAGAAACCTTTAGCCAAGAAGGTTGTCGCCAAAA
TbbH1F76 AGGCTGCTCCCAAAAAGGCTGTTGCCAAGAAACCTTTAGCCAAGAAGGTTGTCGCCAAAA
TbbH1A71 AGGCTGCTCCCAAGAAAGCTG--------------TCGCCAAGAAGGCCGCTCCCAAAA
TbbH1B71 AGGCTGCTCCCAAGAAAGCTG--------------TCGCCAAGAAGGCCGCTCCCAAAA
TbbH1C71 AGGCTGCTCCCAAGAAAGCTG--------------TCGCCAAGAAGGCCGCTCCCAAAA
TbbH1A61 AGGCCGCTCCCAAAAAGG----------
..//..

```
TbbH1A76        AGGTTGCCCCCAAGAAGGTTGCCGGCAAGAAGGCCGCCGCTAAGAAGGCGTGAGCGC-AT
TbbH1B76        AGGTTGCCCCGAAGAAGGTTGCCGGCAAGAAGGCCGCCGCTAAGAAGGCGTGAGCGC-AT
TbbH1C76        AGGTTGCCCCCAAGAAGGTTGCCGGCAAGAAGGCCGCCGCTAAGAAGGCGTGAGCGC-AT
TbbH1D76        AGGTTGCCCCCAAGAAGGTTGCCGGCAAGAAGGCCGCCGCTAAGAAGGCGTGAGCGCGAT
TbbH1E76        AGGTTGCCCCGAAGAAGGTTGCCGGCAAGAAGGCCGCCGCTAAGAAGGCGTGAGCGC-AT
TbbH1F76        AGGTTGCCCCGAAGAAGGTTGCCGGCAAGAAGGCCGCCGCTAAGAAGGCGTGAGCGC-AT
TbbH1A71        AGGTTGCCCCGAAGAAGGTTGCCGGCAAGAAGGCCGCCGCTAAGAAGGCGTGA-------
TbbH1B71        AGGTTGCCCCGAAGAAGGTTGCCGGCAAGAAGGCCGCCGCTAAGAAGGCGTGAGCGC-AT
TbbH1C71        AGGTTGCCCCGAAGAAGGTTGCCGGCAAGAAGGCCGCCGCTAAGAAGGCGTGAGCGC-AT
TbbH1A61        ---TTGCCCCGAAGAAGGTTGCCGGCAAGAAGGCCGCCGCTAAGAAGGCGTGAGCGC-AT
                   ******* ****************************************


TbbH1A76        CCGCTGCTGCCCG-CTATTAGACACG-CTATGAGGTTTACCTGAGTG
TbbH1B76        CCGCTGCTGCCCG-CTATTAGACACG-CTATGAGGTTTACCTGAGTG
TbbH1C76        CCGCTGCTGCCCGGCTATTAGACACGGCTATGAGGTTTACCTGAGTG
TbbH1D76        CCGCTGCTGCCCG-CTATTAGACACG-CTATGAGGTTTACCTGAGTG
TbbH1E76        CCGCTGCTGCCCG-CTATTAGACACG-CTATGAGGTTTACCTGAGTG
TbbH1F76        CCGCTGCTGCCCG-CTATTAGACACG-CTATGAGGTTTACCTGAGTG
TbbH1A71        -------------------------------------------
TbbH1B71        CCGCTGCTGCCCG-CTATTAGACACG-CTATGAGGTTTACCTGAGTG
TbbH1C71        CCGCTGCTGCCCG-CTATTAGACACG-CTATGAGGTTTACCTGAGTG
TbbH1A61        CCGCTGCTGCCCG-CTATTAGACACG-CTATGAGGTTTACCTGAGTG
```

ANNEX D

```
Tandem          caatcttatcaacactcggaagtATGGCGAAGGCATCTGCTGCTCCCAAGAAAGCTGTGG
                                      M  A  K  A  S  A  A  P  K  K  A  V
Tandem          CTAAGAAGGCAGCCCCCAAAAAGGCTGCTCCTAAGAAGGCTGTCGCTAAGAAGGCTGCTC
                A  K  K  A  A  P  K  K  A  A  P  K  K  A  V  A  K  K  A  A
Tandem          CAAAGAAGGCTGTCGCTAAGAAGGCTGCTCCCAAGAAAGCTGTCGCCAAGAAGGCCGCTC
                P  K  K  A  V  A  K  K  A  A  P  K  K  A  V  A  K  K  A  A
Tandem          CCAAAAAGGTTGCCCCGAAGAAGGTTGCCGGCAAGAAGGCCGCCGCTAAGAAGGCGTGAg
                P  K  K  V  A  P  K  K  V  A  G  K  K  A  A  A  K  K  A  Z
Tandem          cgcatccgctgctgcccgctattagacacgctatgaggtttacctgagtgtgggagaaag

Tandem          ctgtcacacgtttcaggacgtcctcgtgcgtctctccaggacggagttagatttttccta

Tandem          tcttacttttttatttctttttatttttttattggatatgtttcgtttgtgggttgcgt

Tandem          cttatgtaccgccatgcggtgttggtgtcgtagcgttgcaaagagcatatcatcctgatg

Tandem          tgtggctatactaactgcctgtgtatggttgtggtccaatagtattactgtaggtctttt

Tandem          tcctttccgcttatttcctccaaactcattgtattctcatgttctttcacattcatttta

Tandem          ttttgctttcatttttctctttctccctgtggaagaagtcgcaatcttatcaacactcgg

Tandem          aagtATGGCGAAGGCATCTGCTGCTCCCAAGAAAGCTGTGGCTAAGAAGGCCGCTCCCAA
                     M  A  K  A  S  A  A  P  K  K  A  V  A  K  K  A  A  P  K
Tandem          AAAGGTTGCCCCGAAGAAGGTTGCCGGCAAGAAGGCCGCCGCTAAGAAGGCGTGAcgcga
                  K  V  A  P  K  K  V  A  G  K  K  A  A  A  K  K  A  Z
Tandem          tccgctgctgcccgctattagacacgctatgaggtttacctgagtg
```

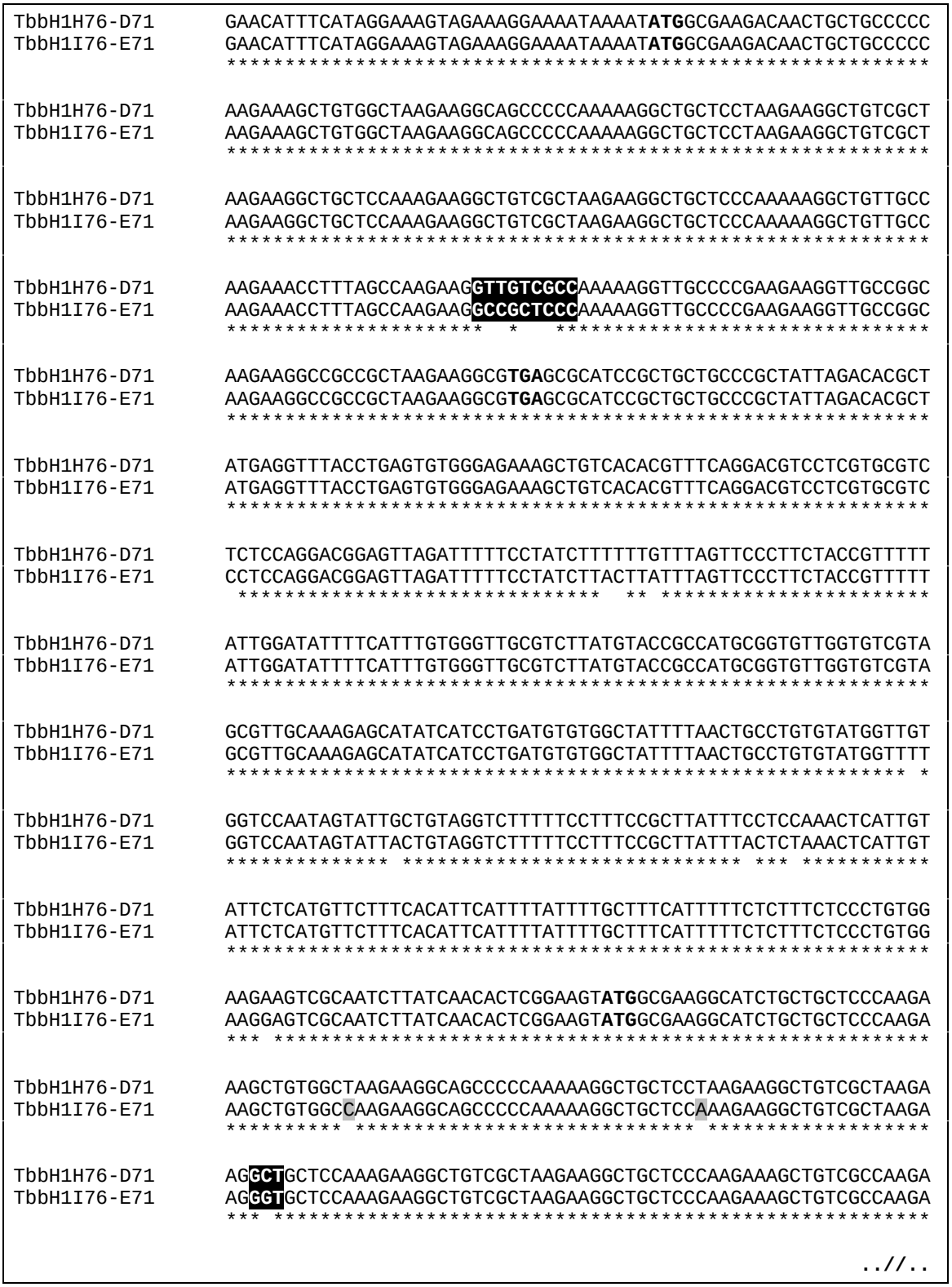
Annex D: Complete nucleotide sequence of the tandem containing a 71 amino acid variant followed by a 36 amino acid variant. Deduced amino acid correspondence is showed in bold letters.

ANNEX E

1.5kb Band	tctttatcgactccccacaaga**ATGAACAACACAACCGCTACTGTGAAGGCCACCCCAAA**
80 aa	**GAAGGTTGCAGCCAAAAAGGCCGCTCCAAAGAAAACTGTGGCCAAAVNGGCTGCTCCAAA**
VARIANT	**GAAGGCTGTTGCAAGCAAACCTTTAGCTAAGAAGGTTGTCGCTAAGAAGGCTGCTCCCAA**
	GAAAGCTGTCGCCAAGAAGGCCGCTCCCAAAAAGGTTGCCCCGAAGAAGGTTGCCGGCAA
	GAAGGCCGCCGCTAAGAAGGCGTGMgcantccgctgctgcccgctattagacacgctatg
	aggtttacctgagtgtgggagaaagctgtcacacgtttcaggacgtcctcgtgcgtccct
	ccaggasgagttagatttttcctatctttttgtwtagttcccawmtaacgttttattg
	gatatgtttcgtttgtgggttgcgtcttatgtaccgccatgbgntntggtgtcgtagcgt
	tgcaaagagcatatcatcctgatgtgtggctattttaactgcctgtgtatggttgtggtc
	caatagtattactgtaggtctttttcctttccgcttatttcctccaaactcattgtattc
	tcatgttctttcacattcattttattttgctttcatttttctctttctccctgtggaaga
	agtcgcaatcctatcaacactcggaagt**ATGGCGAAGSATCTGCTGCTCCCAAGAAAGCT**
76 aa	**GTGGCCAAGAAGGCAGCCCCCAAAAAGGCTGCTCCAAAGAAGGCTGTCGCTAAGAAGGCT**
VARIANT	**GCTCCAAAGAAGGCTGTCGCTAAGAAGGCTGCTCCCAAAAAGGCTGTTGCAAAGAAACCT**
	TTAGCCAAGAAGGCCGCTCCCAAAAAGGTTGCCCCCAAGAAGGTTGCCGGCAAGAAGGCG
	CCGCTAAGAAGGCGTGAgcagtccgctgctgcccgctattagacacgctatgaggtttac
	ctgagtgtgggagaaagctgtcacacgtttcaggacgtcctcgtgcgtcctccaggacgg
	agttagattttcctatctttttgtttagttcccttctaccgttttattggatatgtt
	tcgtttgtgggttgcgtcttatgtacgccatgcggtgttggtgtcgtagcgttgcaaaga
	gcatatcatcctgatgtgtggctattttaactgcctgtgtatggttgtggtccaatagta
	ttactgtaggtctttttcctttccgcttatttcctccaaactcattgtattctcatgttc
	tctcacattcattttattttgctttcatttttctctttctccctgtggaagaagtcgcaa
	tcttatcaacactcggaagt**ATGGCGAAGGCATCTGCTGCTCCCAAGAAAGCTGTGGCCA**
71 aa	**AGAAGGCAGCCCCCAAAAAGGCTGCTCCAAAGAAGCGTGTCGCTAAGAAGGCTGCTCCAA**
VARIANT	**AGAAGGTTGTCGCTAAGAAGGCTGCTCCCAAGAAACKGTCGCCAAGAAGGCCGCTCCCAA**
	AAAGGTTGCCCCGAAGAAGGTTGCCGGCAAGAAGGCCGCCGCTAAGAAGGCGTGAgccga
	Tccgctstgcccgctattagacacgctatgaggtttacctgagtga

Annex E: Complete nucleotide sequence of a T. b. brucei (TREU 927/4) clone containing 3 tandemly arranged ORFs of 80, 76 and 71 amino acids. ORFs are shown in bold capital letters. Only one round of sequencing was performed, explaining undetermined nucleotides.

ANNEX F


TbbH1H76-D71 GAACATTTCATAGGAAAGTAGAAAGGAAAATAAAATATGGCGAAGACAACTGCTGCCCCC
TbbH1I76-E71 GAACATTTCATAGGAAAGTAGAAAGGAAAATAAAATATGGCGAAGACAACTGCTGCCCCC
TbbH1H76-D71 AAGAAAGCTGTGGCTAAGAAGGCAGCCCCCAAAAAGGCTGCTCCTAAGAAGGCTGTCGCT
TbbH1I76-E71 AAGAAAGCTGTGGCTAAGAAGGCAGCCCCCAAAAAGGCTGCTCCTAAGAAGGCTGTCGCT
TbbH1H76-D71 AAGAAGGCTGCTCCAAAGAAGGCTGTCGCTAAGAAGGCTGCTCCCAAAAAGGCTGTTGCC
TbbH1I76-E71 AAGAAGGCTGCTCCAAAGAAGGCTGTCGCTAAGAAGGCTGCTCCCAAAAAGGCTGTTGCC
TbbH1H76-D71 AAGAAACCTTTAGCCAAGAAGGTTGTCGCCAAAAAGGTTGCCCCGAAGAAGGTTGCCGGC
TbbH1I76-E71 AAGAAACCTTTAGCCAAGAAGGCCGCTCCCAAAAAGGTTGCCCCGAAGAAGGTTGCCGGC
TbbH1H76-D71 AAGAAGGCCGCCGCTAAGAAGGCGTGAGCGCATCCGCTGCTGCCCGCTATTAGACACGCT
TbbH1I76-E71 AAGAAGGCCGCCGCTAAGAAGGCGTGAGCGCATCCGCTGCTGCCCGCTATTAGACACGCT
TbbH1H76-D71 ATGAGGTTTACCTGAGTGTGGGAGAAAGCTGTCACACGTTTCAGGACGTCCTCGTGCGTC
TbbH1I76-E71 ATGAGGTTTACCTGAGTGTGGGAGAAAGCTGTCACACGTTTCAGGACGTCCTCGTGCGTC
TbbH1H76-D71 TCTCCAGGACGGAGTTAGATTTTTCCTATCTTTTTTGTTTAGTTCCCTTCTACCGTTTTT
TbbH1I76-E71 CCTCCAGGACGGAGTTAGATTTTTCCTATCTTACTTATTTAGTTCCCTTCTACCGTTTTT
TbbH1H76-D71 ATTGGATATTTTCATTTGTGGGTTGCGTCTTATGTACCGCCATGCGGTGTTGGTGTCGTA
TbbH1I76-E71 ATTGGATATTTTCATTTGTGGGTTGCGTCTTATGTACCGCCATGCGGTGTTGGTGTCGTA
TbbH1H76-D71 GCGTTGCAAAGAGCATATCATCCTGATGTGTGGCTATTTTAACTGCCTGTGTATGGTTGT
TbbH1I76-E71 GCGTTGCAAAGAGCATATCATCCTGATGTGTGGCTATTTTAACTGCCTGTGTATGGTTTT
TbbH1H76-D71 GGTCCAATAGTATTGCTGTAGGTCTTTTTCCTTTCCGCTTATTTCCTCCAAACTCATTGT
TbbH1I76-E71 GGTCCAATAGTATTACTGTAGGTCTTTTTCCTTTCCGCTTATTTACTCTAAACTCATTGT
TbbH1H76-D71 ATTCTCATGTTCTTTCACATTCATTTTATTTTGCTTTCATTTTTCTCTTTCTCCCTGTGG
TbbH1I76-E71 ATTCTCATGTTCTTTCACATTCATTTTATTTTGCTTTCATTTTTCTCTTTCTCCCTGTGG
TbbH1H76-D71 AAGAAGTCGCAATCTTATCAACACTCGGAAGTATGGCGAAGGCATCTGCTGCTCCCAAGA
TbbH1I76-E71 AAGGAGTCGCAATCTTATCAACACTCGGAAGTATGGCGAAGGCATCTGCTGCTCCCAAGA
TbbH1H76-D71 AAGCTGTGGCTAAGAAGGCAGCCCCCAAAAAGGCTGCTCCTAAGAAGGCTGTCGCTAAGA
TbbH1I76-E71 AAGCTGTGGCCAAGAAGGCAGCCCCCAAAAAGGCTGCTCCAAAGAAGGCTGTCGCTAAGA
TbbH1H76-D71 AGGCTGCTCCAAAGAAGGCTGTCGCTAAGAAGGCTGCTCCCAAGAAAGCTGTCGCCAAGA
TbbH1I76-E71 AGGGTGCTCCAAAGAAGGCTGTCGCTAAGAAGGCTGCTCCCAAGAAAGCTGTCGCCAAGA
..//..

```
TbbH1H76-D71      AGGCCGCTCCCAAAAAGGTTGCCCCGAAGAAGGTTGCCGGCAAGAAGGCCGCCGCTAAGA
TbbH1I76-E71      AGGCCGCTCCCAAAAAGGTTGCCCCGAAGAAGGTTGCCGGCAAGAAGGCCGCCGCTAAGA
                  ************************************************************

TbbH1H76-D71      AGGCGTGAGCGCATCCGCTGCTGCCCGCTATTAGACACGCTATGAGGTTTACCTGAGTG
TbbH1I76-E71      AGGCGTGAGCGCATCCGCTGCTGCCCGCTATTAGACACGCTATGAGGTTTACCTGAGTG
                  ************************************************************
```

Annex F: Complete nucleotide sequence of two H1 tandem of *T. b. brucei* (TREU 927/4) each containing a 5' ORF coding for a 76 amino acid variant with two threonines at positions 4 and 5, followed by an ORF coding for a 71 amino acid variant. Respective start and stop codons are shown in bold letters. Black boxes: substitutions resulting in amino acid changes; grey boxes: substitutions that do not influence the amino acid level; *: matching positions.

ANNEX G

A

>***T.b.brucei*** **(strain 427)**
GAAGTCGCAATCTTATCAACACTCGGAAGT**ATG**GCGAAGGCATCTGCTGCTCCCAAGAAAGCTGTGGCTAAGAAGGC
AGCCCCCAAAAAGGCTGCTCCAAAGAAGGCTGTGGCTAAGAAGGCTGCTCCCAAAAAGGCTGTTGCCAAGAAACCTT
TAGCCAAGAAGGTTGTCGCCAAAAAGGTTGCCCCCAAGAAGGTTGCCGGCAAGAAGGCCGCCGCTAAGAAGGCG**TGA**
GCGCATCCGCTGCTGCCCGCTATTAGACACGCTATGAGGTTTACCTGAGTGTGGGAGAAAGCTGTCACACGTTTCAG
AAGGTACTTGTGCGCCCGTCAAGGACGGAGTTAGATTTTTCTTATCTTACTTGTTTAGTTCCCTTCTACCGTTTGTA
TTGGATATGTTTCGTTTGTGGGTTGCGTCTTATGTACCGCCATGCGGTGTTGGTGTCGTAGCGTTGCAGAGAGCATA
TCGTCCTGAAAAAAAAAAAAA

B

```
TbbH1B71        MAKASAAPKKAVAKKAAPKKAAPKKAVAKKAAPKKAVAKKAAPKKAVAKKAAPKKVAPKK
Tbb-427         MAKASAAPKKAVAKKAAPKKAAPKKAVAKKAAPKKAVAKKPLAKKVVA-----KKVAPKK
TbbH1A61        MAKASAAPKKAVAKKAAPKKAAPKKAVAKKGAPKKAVAKKAAP----------KKVAPKK
                ****************************** *********             *******

TbbH1B71        VAGKKAAAKKA
Tbb-427         VAGKKAAAKKA
TbbH1A61        VAGKKAAAKKA
                ***********
```

Annex G: A cDNA clone coding for a 66 amino acid histone H1 variant of a different *T. b. brucei* strain (Strain 427). **A**: Complete nucleotide sequence. Start and stop codons are shown in bold letters. **B**: Alignment of the deduced amino acid sequence of this clone against two H1 variants of T. b. brucei strain TREU 927/4. (Adapted from M. Hynek, Certificate work, University of Neuchâtel (Switzerland); 1999).

VIII. ACKNOWLEDGEMENTS

I especially wish to thank Carmela Dimito and Marc Zehnder for the good atmosphere in the lab, their support at many levels and for their friendship as well as Alain Hirschy for preliminary H1 production and purification experiments and for his help in confocal imaging. We spent some moment of real fun,…, not always behind the bench!

I also wish to express my gratitude to:

Prof. B. Betschart for giving me the opportunity to work in his laboratory and for the reading of the manuscript followed by very helpful advice.

Olivier Rais for his help in working with mice, for the coffee breaks and the free newspaper.

Yves Balmer, Sophie Marc-Martin, Véronique Douet-Orhant, Albert Spielman and Nadine Paris from the biochemistry lab for their help and technical advice.

Dr. R. Brun and especially Cecile Schmid from the Swiss Tropical Institute for the culturing and propagating of trypanosomes.

Dr. T. Seebeck and Dr. I. Roditi from the University of Bern for providing us with cDNA libraries and for helpful information as well as Dr. G. Corradin from the ISREC, Lausanne for providing us with a synthetic peptide for immunization. The continued interest for the project by Dr. H. Hecker is also acknowledged.