Molecular and Biochemical Characterization of the Hypoxanthine-Guanine Phosphoribosyltransferase Gene from the *Trypanosomatidae* family

by

Thomas Edward Allen

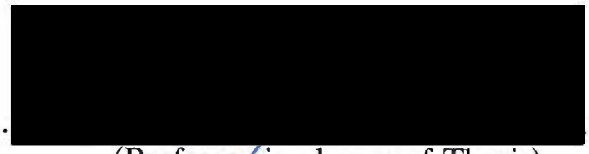
A Dissertation

Presented to the Department of Biochemistry and Molecular Biology
School of Medicine
Oregon Health Sciences University
in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

May 27, 1993

Approved:



(Professor in charge of Thesis)

(Chairman, Graduate Council)

Table of Contents

Appendices

Catalog of Figures

Introduction

Manuscript #1

Manuscript #2

Appendix A

Appendix B

Appendix C

Acknowledgement

Many people have contributed in one way or another to the completion of this thesis but I would like to acknowledge my parents, John and Evelyn Allen, first since they were instrumental in shaping my interests and in developing my love of nature and my curiosity to explore it. In addition, I would like to thank my new family, the Parshleys, who made me welcome in my new home in Portland and who were a constant source of understanding and humor for both my wife and myself over the last several years.

I would also like to thank the members of my committee, Drs, Brennan, Clinton, Riscoe, Heffron, and Landfear, for all of their help and guidance over the course of my research.

Probably the most important decision that a new graduate student makes in his career is in the choice of their advisor. I would like to give special thanks to Dr. Buddy Ullman for all of his help and guidance over the last several years. I feel very fortunate to have had the opportunity to work with a scientist of Buddy's calibre. I would also like to thank him for introducing me to the fascinating world of Kinetoplastida and the diseases they cause. I leave his laboratory with mixed feelings since I am sure that the contributions that he will make to our understanding of these organisms will continue to be both medically and socially significant. I wish him and all the people who have the great fortune of working with him a productive and exciting future.

Finally, I would like to thank all of the people who I have had the great fortune to work with in Dr. Ullmans laboratory:

Sheri Hanson	Joanne Beck
Gene Henchel	Terry Coons
Nancy Hendrickson	Debbie Henderson
Costi Sifri	Ho-Yon Hwang
Keith Wilson	Greg Barton
Dick Brennen	Patrick Jimenez
Lilliana Tolmasky	Sue Bergeson
Maria Schumacher	Jerry Scott

Dr. Buddy Ullman

This thesis is dedicated to my best friend and wife, Dr. Lisa Parshley. Without her love and understanding this thesis would simply not exist.

Summary

Of the major metabolic pathways in parasitic protozoa, the purine pathway is one of the most unique in that each genus of parasite is auxotrophic for purines and has evolved a distinctive series of purine salvage enzymes that enable them to scavenge host purines. Preliminary studies have indicated that the enzyme hypoxanthine-guanine phosphoribosyltransferase (HGPRT) is particularly crucial to the purine salvage process in these organisms. This thesis entailed a thorough examination of HGPRT both at the molecular and biochemical level in members of the *Trypanosomatidae* family.

In order to examine the HGPRT enzyme in detail, the *T. brucei hgprt* was cloned from a cDNA and genomic library using a 174 bp fragment of the *hgprt* that was amplified from genomic *T. brucei* DNA in the polymerase chain reaction (PCR). Nucleotide sequence analysis revealed a 630 bp open reading frame coding for a 210 amino acid protein with a M_r = 23.4 kd. Southern blot analysis suggests that the *hgprt* is found in at least two copies but that the organization of the *hgprt* is not in a tandem repeat. Northern blot analysis revealed two different transcripts, one at 1.4 and one at 1.9 kb. Both transcripts were found in the trypomastigote and procyclic form of the organism suggesting that no differential regulation of the *hgprt* occurs in the different life stages of the parasite. Analysis of the mature message suggests that the heterogeneity in transcript size is found in the 3' end of the message. To confirm that the *hgprt* from *T. brucei* had been isolated and to develop a system for purification of large quantities of recombinant enzyme, the *hgprt* gene was

engineered using PCR and cloned into the prokaryotic expression vector pBAce. Using this expression vector and an *E. coli* strain deficient in *hprt* activity, ≈ 100 mg of protein for each liter of cells was obtained. Utilizing a GTP-agarose affinity column, *T. brucei* recombinant HGPRT was purified to homogeneity. The purified enzyme showed expected phosphoribosylation activity with both hypoxanthine and guanine as substrates. In addition, the *T. brucei* HGPRT enzyme was able to utilize the pyrazolopyrimidine, allopurinol. However the purified enzyme was not able to phosphoribosylate adenine or xanthine, confirming that there are separate phosphoribosyltransferase for these bases. Lineweaver-Burk analysis was used to calculate Km and k_{cat} values for the recombinant enzyme for its substrates. K_m values of 4.8, 2.3 and 243 μM and k_{cat} values of 53, 32, and 4.4 μmol/min/mg protein were found for guanine, hypoxanthine and allopurinol respectively. Several metabolic effectors of HGPRT were also examined for their ability to modulate the ability of HGPRT to phosphoribosylate its substrates. In order of descending effectiveness, GTP, IMP, GMP, ATP, and AMP at a concentration of 1mM were all found to inhibit the HGPRT enzyme. pH optimization analysis revealed that reaction conditions at pH 8.5 showed the greatest amount of phosphoribosylation activity although the HGPRT enzyme was active over a wide pH range. In addition, the quaternary structure of the active enzyme was examined. Experiments involving gel filtration columns and non-denaturing polyacrylamide gels suggest that the active *T. brucei* enzyme is in a dimer configuration.

The *T. cruzi hgprt* was also isolated using a 174 bp fragment that was amplified employing PCR. An open reading frame was identified of 663 nucleotides encoding a 221 amino acid protein with a M_r = 26 kd. When aligned, the T. *brucei* and T. *cruzi* HGPRT

showed a 50 % identity at the amino acid level. Southern blot analysis revealed that the *T. cruzi hgprt* appears to be a single copy gene while northern blot analysis revealed a single transcript at 1.9 kb. The *T. cruzi hgprt* was engineered using PCR and cloned into the expression vector, pBAce. High levels of expression were observed and *T. cruzi* recombinant HGPRT enzyme was purified to homogeneity over a GTP-agarose affinity column. The *T. cruzi* HGPRT demonstrated high levels of enzyme activity with both hypoxanthine and with guanine. Allopurinol also was phosphoribosylated by the recombinant enzyme while adenine and xanthine were not found to be substrates. K_m values of 9.9, 6.4, and 714 μM and K_{cat} values of 55, 97, and 0.8 μmol/min/mg protein were calculated for the recombinant protein using Lineweaver-Burk analysis for guanine, hypoxanthine and allopurinol respectively. As with the *T. brucei* HGPRT, GTP, IMP, and GMP were all found to inhibit the HGPRT enzyme. In contrast however, neither AMP or ATP appeared to effect the activity of the *T. cruzi* HGPRT enzyme. Optimization of Enzyme activity appeared to be highest at pH 8.0 although the enzyme appeared to be very active over a large pH range. Gel filtration experiments with the *T. cruzi* HGPRT enzyme suggest that the active enzyme is also in a dimmer configuration as is the *T. brucei* enzyme.

To further our analysis of the HGPRT enzyme from the *Trypanosomatidae* family, the *Leishmania donovani hgprt* was isolated. In order to clone the *hgprt* from *L. donovani*, it was necessary to isolate native HGPRT protein from the promastigote form of *L. donovani*. Small quantities of HGPRT enzyme was isolated using a modification of the purification scheme used for isolating the recombinant *Trypanosoma* HGPRT. Sequencing of the *L. donovani* HGPRT enzyme proved successful and a 36 b.p. oligonucleotide primer

to the reverse compliment of the sequence was synthesized. Using the primer to the protein sequence as the 3' primer and the splice-leader sequence as the 5' primer, PCR experiments using mRNA from *L. donovani* were performed. A band of ≈ 156 b.p. was isolated and used to screen a *L. donovani* cosmid library. Nucleotide analysis of the positive clones revealed an open reading frame of 632 nucleotides encoding a M_r = 23.6 kd protein. The *L. donovani* gene showed a 50% identity to the *T. cruzi* gene while showing only a 34% identity to the *T. brucei* gene. Southern blot analysis revealed that the *hgprt* gene was found as a single copy. Northern blot analysis disclosed a single transcript of 1.9 Kb. Once again, PCR was used to engineer the *L. donovani hgprt* to allow for subcloning into the expression system pBAce. Expression of HGPRT enzyme has been obtained with this construct with hypoxanthine, guanine and allopurinol being phosphoribosylated by the recombinant enzyme. Neither adenine or xanthine proved to be substrates for the recombinant enzyme.

Comparison of the HGPRTs from members of the *Trypanosomatidae* family with those of other eukaryotic and prokaryotic PRTs revealed that the trypanosomatids exhibited a greater degree of identity to the *V. harveyi* HPRT (32-36%) then to the other eukaryotic HGPRTs (18-25%). This suggests that the HGPRT from the trypanosomatids are the most divergent of the group.

Introduction

A. Parasitic Diseases

With over 4.5 billion people affected by parasitic organisms worldwide, the toll taken in terms of economic and human suffering is staggering. Despite the enormity of the problem, very little is known about the organisms themselves and the treatments prescribed today are not only inadequate but are frequently toxic to the human host. The analysis of unique metabolic pathways in these organisms should enable the design of chemotherapeutic regimens that are selective for the parasite and are nontoxic to the human host. My current project entails both a molecular and biochemical exploration of a unique metabolic pathway in members of the *Trypanosomatidae* family of parasitic protozoa.

A contemplation of the vast variety of parasites is astounding. Defined as any organism that spends part or all its life in contact with one or more hosts from which it derives a benefit at the expense of the host, parasites are found throughout every category of life. When the term is used to include viruses and bacteria that are parasitic, there are far more parasitic organisms then nonparasitic organisms (1). In a phenomenon known as hyperparasitism, parasites themselves are host to a variety of organisms that are parasitic in nature. The parasitic life strategy is so successful that parasites have developed for practically every tissue and organ in the body.

In the categorization of scientific endeavors, parasitology has come to be known as the study of parasitic protozoa and helminths. This still includes the largest and most diverse

group of organisms on earth with over 10,000 different species of parasitic protozoa alone being identified to date. Figure 1 shows the major classifications of parasitic organisms. Most of the burden that parasites inflict can be found in the tropical zones of the earth although parasites are found on every continent and every ocean. Unfortunately, the majority of the more serious parasites are found in the developing countries of the world and so the enormous impact of these diseases are not well known in the developed countries. For example, of the 60 million deaths each year from all causes, 30 million are children under the age of 5 years. Of these 30 million children over 15 million die due to a combination of intestinal parasites and malnutrition (2). Certainly this high incidence of death amongst children is one of the main reasons for the lack of success of population control in developing countries which plays an every spiraling cycle of environmental degradation and human suffering. Today, malaria still receives the dubious distinction of being the single greatest agent of death of mankind. It is estimated that more people have died from malaria then any other single cause including war. One extrapolation of data on the prevalence of infection and subsequent death due to malaria suggests that over 27 billion people have died due to this organism (3). This figure is over 5 times the total population of the planet.

Parasites do not always cause obvious hardships. Many parasites such as *Trypanosoma* which can infect cattle, make large areas of Africa incapable of sustaining livestock. Other parasites such as the causative agent of African river blindness, *Onchocerca*, do not kill their host but debilitate them and make them an unproductive part of society. Other helminths also do not immediately kill their host but create conditions that can bring about malnutrition and, in times of stress on the host, threaten the survival of the host. In addition,

chronic forms of many of the diseases such as visceral leishmaniasis and malaria cause enormous economic hardships on developing countries. It is estimated for example that over 2 billion dollars are lost annually due to chronic malaria (2). More recently with the advent of the AIDS virus, many parasites that were relatively inconsequential to the survival of the host are now proving to be fatal (4,5).

Finally, the sheer poverty, overpopulation and poor sanitary conditions of the majority of people affected by these parasites cannot be understated. In addition, the constant political unrest of these regions and the forced resettlement of the people affected have all had negative impacts on the efforts of world health officials to try and control these diseases. In particular, the lack of commitment on the part of developed nations to eliminate hunger throughout the world assure that without the aid of medical science these diseases will only continue to increase in their affect on the entire human race.

The division of parasitology and bacteriology which occurred earlier in this century has proven to have had a negative impact on the field of parasitology. Indeed until the 1980s, most universities ignored training in parasitology all together. A survey of 141 universities in the United States in 1978 revealed that only 4.7% of the graduate classes in parasitology were devoted to immunology while only 3.6% were devoted to the biochemistry of parasites. In the same study, no classes were found that emphasized the molecular biology of these organisms. Another study completed in 1977 revealed that the global support for research into parasitic diseases affecting over 2.2 billion people including malaria, schistosomiasis, filariasis, amebiasis and ascariasis was a paltry $10 million dollars (6). During this same year the total expenditure in the United States alone for research into

cancer, which affects ≈10 million people, was $815 million dollars. It is shocking to think that organisms that are so prevalent and cause such an incredible amount of harm have been virtually ignored until recent times.

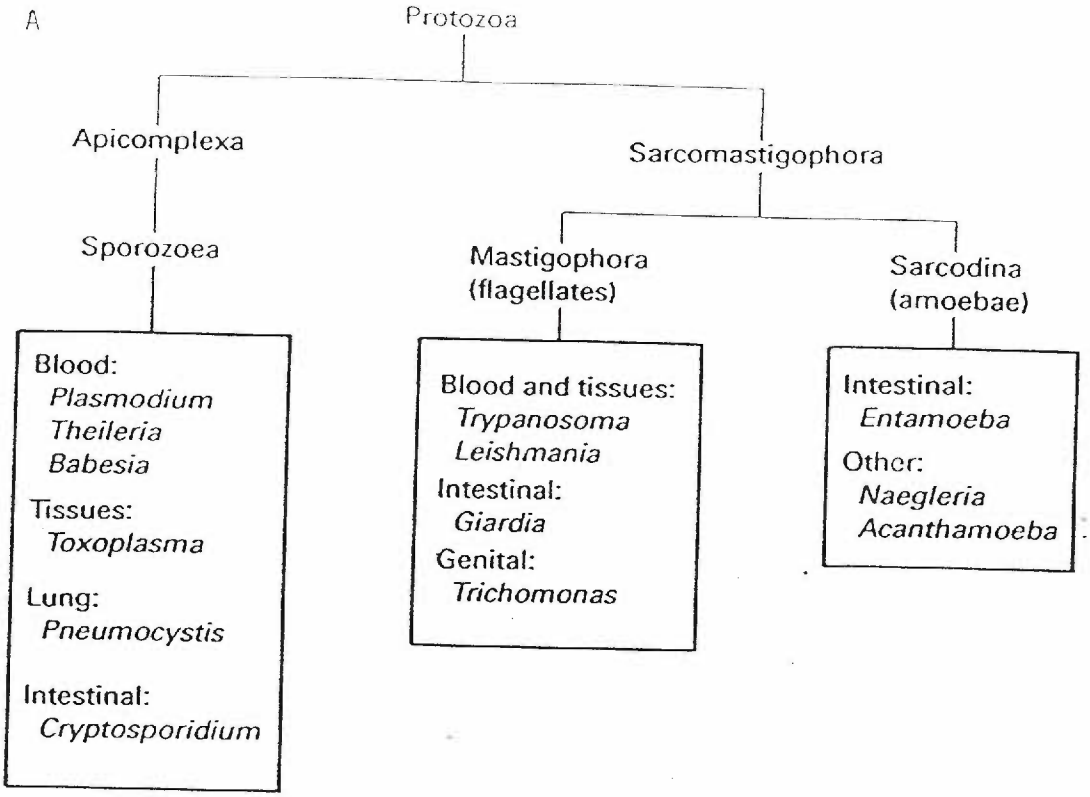
Beginning in the 1980s, new attention has been focused on the study of parasitic organisms. New partnerships between government, universities and industry show promise for the future. In addition, programs such as the Tropical Disease Research Program (TDR) of the World Health Organization is bringing a truly global effort into the control and eradication of parasites. Private foundations such as the John D. and Catherine T. MacArthur foundation have also increased there support for not only basic reaseach but for education as well. An excellent example of these new partnerships is seen in the Biology of Parasitism course that is offered at the Marine Biological Laboratory (MBL) in Woods Hole Massachusetts which has been training young researchers in a variety of topics in parasitology. The course is a collaboration of efforts ranging from the excellent faculty from both industry and academia with funding being provided from private sources such as the Burroughs-Wellcome Fund and industry such as New England Biolabs. These types of collaborative effort hold the key to successful endeavors which will ultimately relieve the burden of parasitic infection on the world population.

Figure 1. Classification of the major parasitic organisms

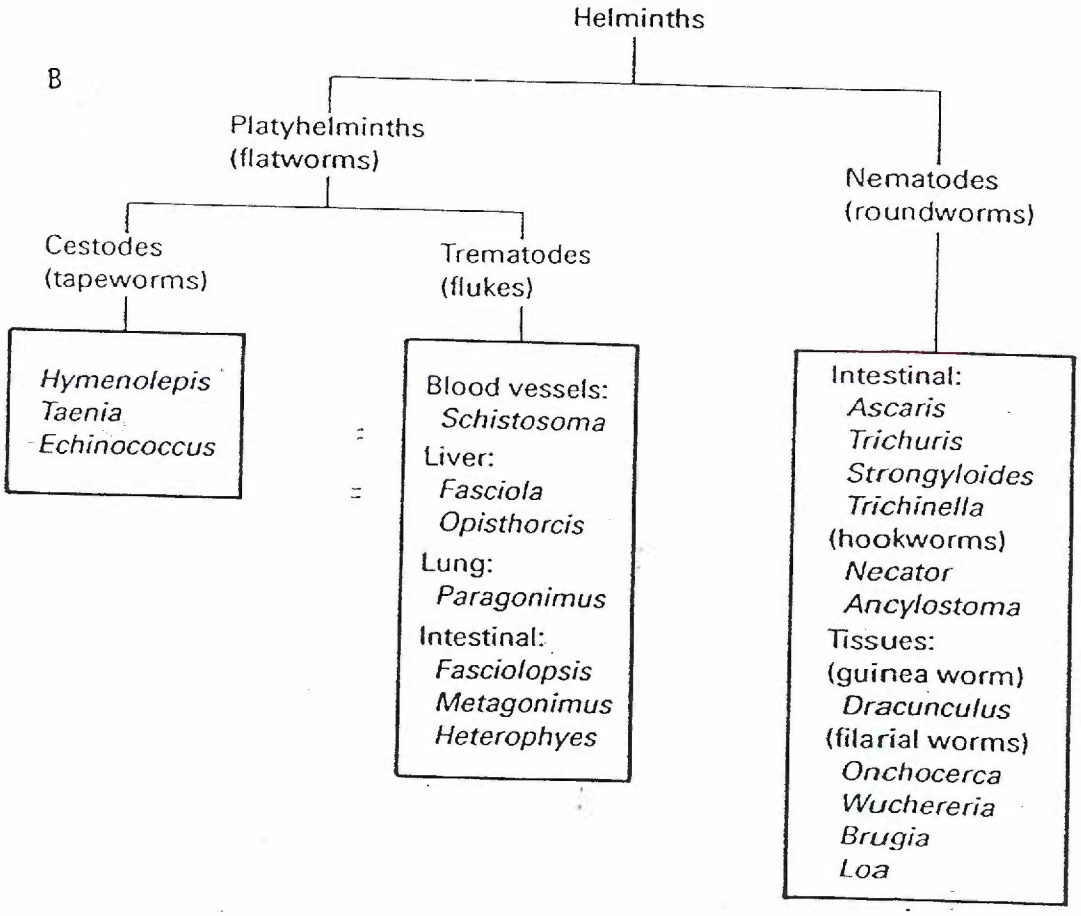
a. Major parasitic protozoa

b. Major parasitic helminths

A


Protozoa
Apicomplexa
Sporozoea
Sarcomastigophora
Mastigophora (flagellates)
Sarcodina (amoebae)
Blood:
Plasmodium
Theileria
Babesia
Tissues:
Toxoplasma
Lung:
Pneumocystis
Intestinal:
Cryptosporidium
Blood and tissues:
Trypanosoma
Leishmania
Intestinal:
Giardia
Genital:
Trichomonas
Intestinal:
Entamoeba
Other:
Naegleria
Acanthamoeba

B


Helminths
Platyhelminths (flatworms)
Nematodes (roundworms)
Cestodes (tapeworms)
Trematodes (flukes)
Hymenolepis
Taenia
Echinococcus
Blood vessels:
Schistosoma
Liver:
Fasciola
Opisthorcis
Lung:
Paragonimus
Intestinal:
Fasciolopsis
Metagonimus
Heterophyes
Intestinal:
Ascaris
Trichuris
Strongyloides
Trichinella
(hookworms)
Necator
Ancylostoma
Tissues:
(guinea worm)
Dracunculus
(filarial worms)
Onchocerca
Wuchereria
Brugia
Loa

B. *Kinetoplastida*

Members of the order *Kinetoplastida* have been called both primitive and highly specialized (7). Figure 2 shows the taxonomic classification of the order and its suborders. The *Kinetoplastida* contains flagellates that can be either "free living" or parasitic. The organisms can have one or two flagellum which are composed of a central axoneme surrounded by an outer sheath that is a continuation of the cell membrane (1). The axoneme arise from a centriole called a basal body or kinetosome. The flagellar pocket is found at the base of the flagellum surrounding the kinetosome. In addition, a network of pellicular microtubules are found beneath the cell membranes of these organisms and are thought to be involved in adding strength and support. The *Kinetoplastida* contain a single mitochondria which often extends the entire length of the body. The mitochondria is tube like in appearance and contains the kinetoplast in close proximity to the flagellar basal bodies. The kinetoplast, for which the order derives it's name, is a organized network of linked DNA that is in the form of maxicircles and minicircles. There are approximately 10,000 minicircles and between 20 and 50 maxicircles in intact kinetoplast DNA (kDNA) (8). It has been known for a long time that the maxicircles encode the genes required for mitochondrial proteins but the function of the minicircles was a perplexing question until recently. It is now known that the maxicircles do not always contain all the information necessary for a functional protein and that the minicircles contain the information to correctly edit (RNA editing) transcripts of the maxicircles to form the mature message (9). In fact it has been observed that RNA editing can even be responsible for creating correct

AUG initiation codons for some messages (10). Another fascinating feature of the order *Kinetoplastida* is that many, if not all, transcription events occur by first producing a polycistronic pre-MRNA precursor which is then cleaved and trans-spliced with a 39 nucleotide sequence that is common to all mature mRNAs. No evidence of cis-splicing has been observed in this order of organisms.

The order *Kinetoplastida* contains two suborders; the free living *Bodonina* and the exclusively parasitic *Trypanosomatina*. This discussion will focus on the later of the two. The suborder *Trypanosomatina* contains a single family; the *Trypanosomatidae*. *Trypanosomatidae* can be either monoxenous or heteroxenous. The organisms undergo dramatic morphological and metabolic changes particularly during the transformation that occurs when the organism changes hosts (invertebrate to vertebrate and vise versa). Before 1980, morphological characteristics were described in terms of the genus in which the form was observed leading to much confusion (7). For example, if you were describing what today is termed the amastigote form of *Trypanosoma cruzi*, you would have referred to it as the leishmania form of the organism. In 1980 Hoare and Wallace described a system based on the morphology of the organism instead of relating to the genus name (7). Figure 3 illustrates the current nomenclature with the root of the classification relating to the flagellum. The suffix of the nomenclature, mastigote, literally means "whip like" while the prefix refers to the position of the flagellum. This nomenclature has been important in simplifying classification and identification of these organisms both in the laboratory and in the field.

Figure 2. **Classification of the order *Kinetoplastida***

Classification of Kinetoplastida

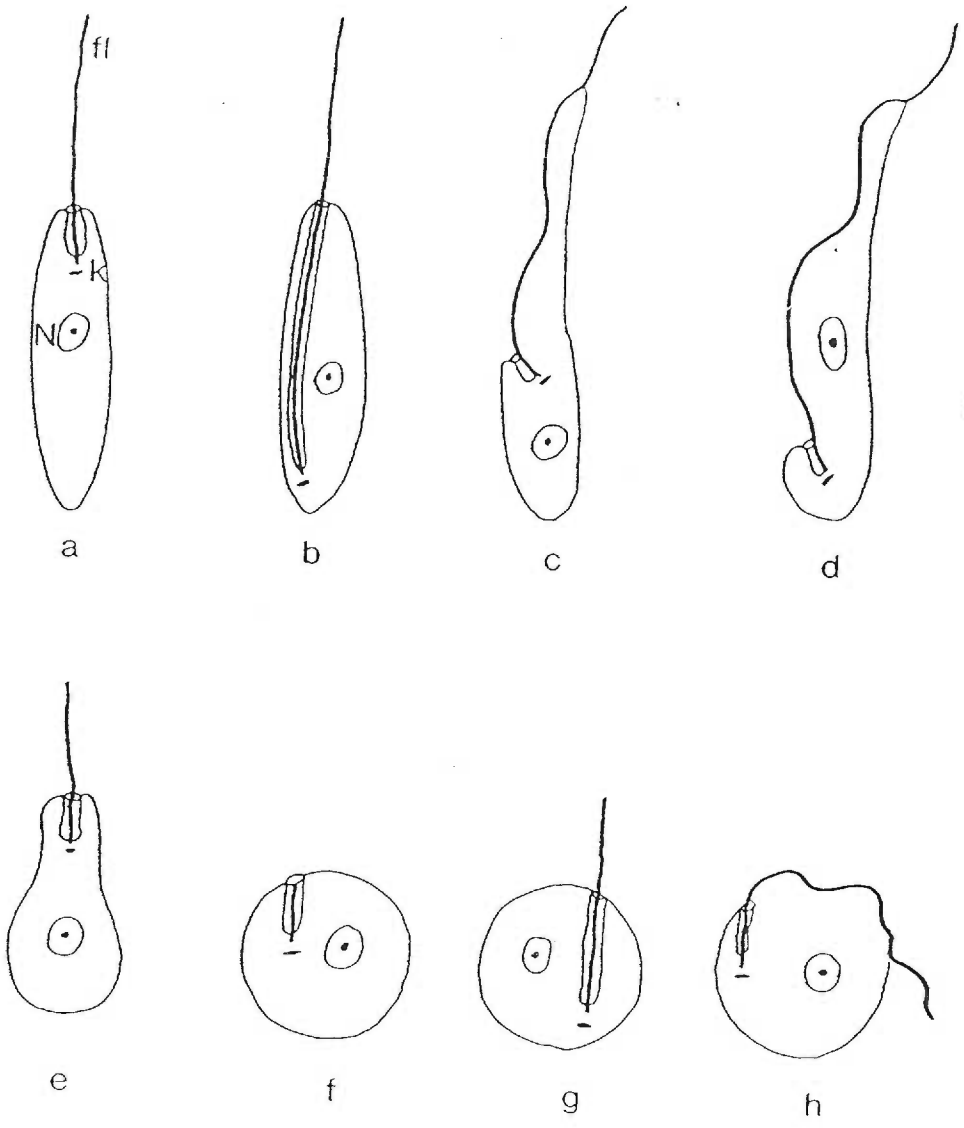
Kingdom	Protista
Subkingdom	Protozoa
phylum	Sarcomastigophora
Subphylum	Mastigophoria
Class	Zoomastigophorea
Order	Kinetoplastida
Suborder	Trypanosomatina
Family	Trypanosomatidae
Genus	

Crithidia Leptomonas Herpetomonas Trypanosoma Phytomonas

Endotrypanum Leishmania Rhynchoidomonas

Figure 3. **Morphological classification of the family *Trypanosomatidae***

a. Promastigote, b. Opisthomastigote, c. Epimastigote, d. Trypomastigote, e. Choanomastigote, f. Amastigote, g. Paramastigote, h. Sphaeromastigote


fl
k
N
a
b
c
d
e
f
g
h

C. *Trypanosoma*

The genus *Trypanosoma* includes both human and animal parasites. Except for *T. equiperdum*, all are heteroxenous and are transmitted by blood-feeding invertebrates. The genus is divided into two broad classes based on its development in the invertebrate vector. The Salivaria species, which include those that cause African sleeping sickness, develop in the front portion of the digestive tract. The stercoraria species develop in the hind gut of the vector and include *T. cruzi* which is the causative agent of Chagas' disease. The important *Trypanosoma* species are listed in figure 4.

1. **African Trypanosomiasis**

The three most medically and economically relevant *Trypanosoma* that are found in Africa include *T.b. rhodesiense* and *T.b. gambiense*, which cause African sleeping sickness in humans, and *T.b. brucei* which infects native and introduced species of animals. The parasites are morphologically indistinguishable but have different hosts and vectors. They also show dramatic dissimilarity in their pathogenesis and physiological characteristics (11). African *Trypanosoma* are found between latitude 15 degrees north and 25 degrees south (figure 5). Their range coincides with that of their vector, the tsetse fly (*Glossina* spp.).

Figure 4. **Members of the genus *Trypanosoma***

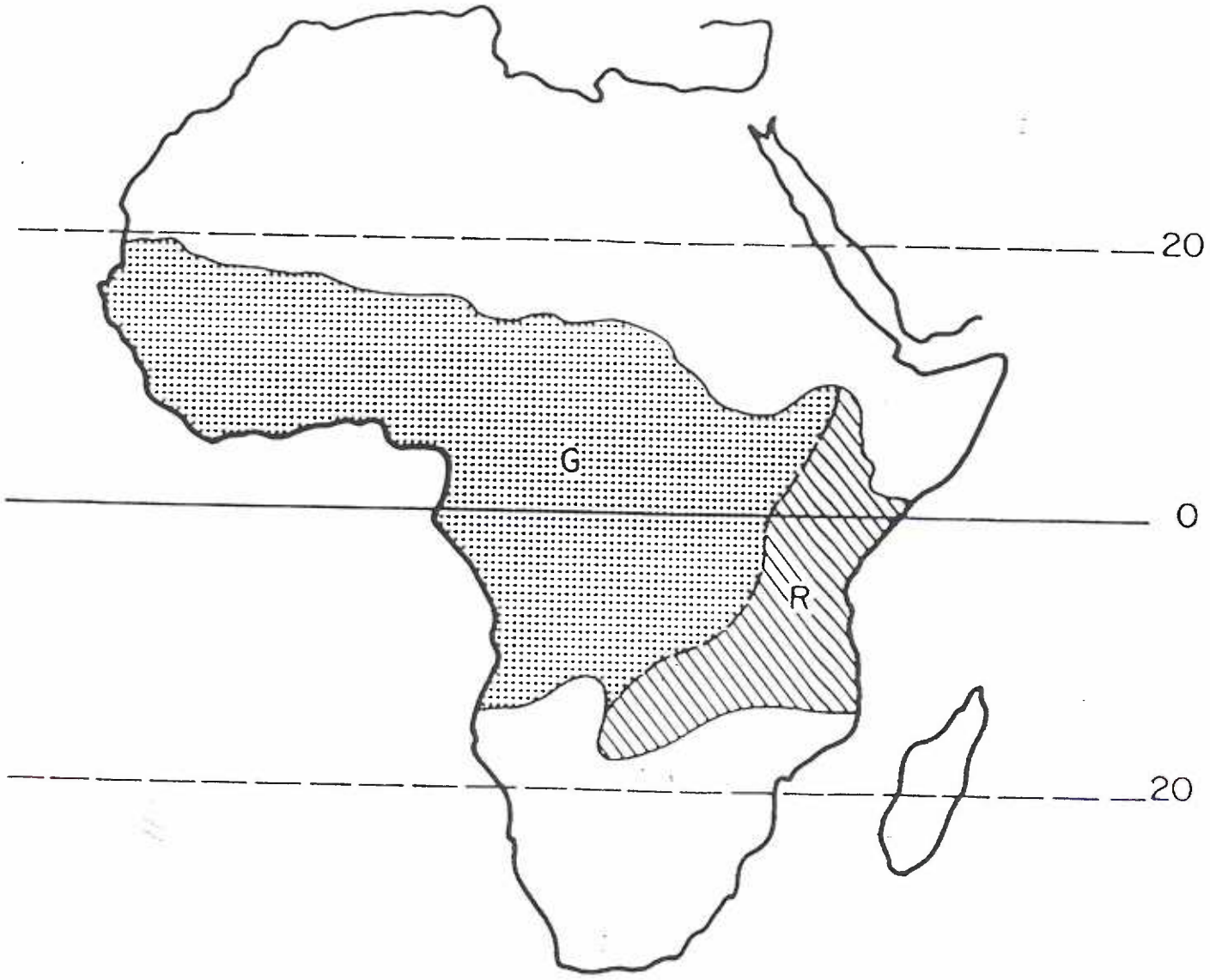
Important Trypanosoma Species

Species	Disease	Geographic Distribution
Salivaria		
T.b.brucei	Nagana	Tropical Africa
T.b.gambiense	Sleeping Sickness	West Africa
T.b.rhodeiense	Sleeping Sickness	East Africa
T.congolense	Bovine Trypanosomiasis	Congo,Zululand
T.simiae	Virulent Trypanosomiasis	East Africa
Stercoraria		
T.cruzi	Chagas' Disease	South America
T.rangeli	Nonpathogenic	South America
T.theileri	Nonpathogenic	Worldwide

Figure 5. **Distribution of African trypanosomiasis**

R - *T.b. rhodesiense*

G - *T.b. gambiense*


20
G
0
R
20

a. **Animal African Trypanosomiasis**

The organism *T.b. brucei* along with *T. vivax* and *T. congolense* are the causative agent of nagana in native ruminants found in Africa. The organisms are also capable of infecting many introduced species such as sheep, goats, oxen, horses, dogs and mules. They do not appear to be infective toward humans but the economic impact on the population of Africa cannot be overstated. It is estimated that 4.5 million square miles (approximately the size of the United States) is incapable of supporting more then 1/5 of the lands calculated carrying capacity for cattle (7). The increase in livestock could be used to feed the ever growing population of the region although the ecological wisdom of increasing cattle production is a matter of dispute (12). Nonetheless, a wise increase in the number of cattle, sheep and goats would help alleviate the malnutrition that is seen in this section of Africa.

The pathology of animal trypanosomiasis varies depending on the species infected. Native wildlife rarely succumb to the parasite and usually great stress is necessary for the parasite burden to become a problem. Indeed, many native species act as "reservoirs" for Trypanosoma and do not appear to be affected at all by the parasite (7,13). The situation is very different for the introduced species and for the native ruminants. The primary characteristic of trypanosomiasis in animals is anaemia and malaise with an enlargement of the spleen and lymph nodes. In susceptible hosts, organ damage can occur with the heart being particularly affected. Immunosuppression has been observed and the reproduction system is also compromised with infertility often the outcome (14). Controversy still surrounds the exact cause for the pathology in affected animals, however it is obvious that

with the wide variety of symptoms and the complex interaction between the vector, reservoir species and host, that the scourge of animal trypanosomiasis will continue to have a negative impact on the population of Africa.

b. **Human African Trypanosomiases**

With the possible exception of malaria, African sleeping sickness has influenced the settlement pattern of more people across the continent of Africa then any other disease. Although few people are infected by the parasite today ($\approx$10-20,000), over 50 million people are at risk of exposure to the parasite thus creating the possibility for a reoccurrence of devastating epidemics such as those that occurred earlier in this century (15,16). This concern has caused the World Health Organization and African health officials to give great priority to the eradication of this disease. Unfortunately, African sleeping sickness has proven to be a difficult disease to stamp out.

The two species which are responsible for creating a disease state in man are *T.b. gambiense* and the more virulent *T.b. rhodesians*. The range in which these organism are found overlap each other with *T.b. rhodesiense* being more confined in it's range. There is some evidence that the two organisms do not so much represent different species as they represent the two extremes of the disease in humans (7,17,18).

The symptoms that are associated with African sleeping sickness include malaise, pain in various joints and what can be a dramatic loss of weight. In the initial phase of the disease an intermittent fever is usually present and a identifiable wound, called a chancre, is seen at the site of the bite by the vector. Symptoms that suggest involvement of the

central nervous system include irrational behavior with epileptic like attacks which can be followed by coma and death. Once the disease progresses to the point of nervous system involvement, treatment is extremely difficult. Autopsies have revealed that infection can damage the heart and enlarged spleens have also been noted. It is obvious that the immune system is compromised which can lead to secondary infections that are often fatal.

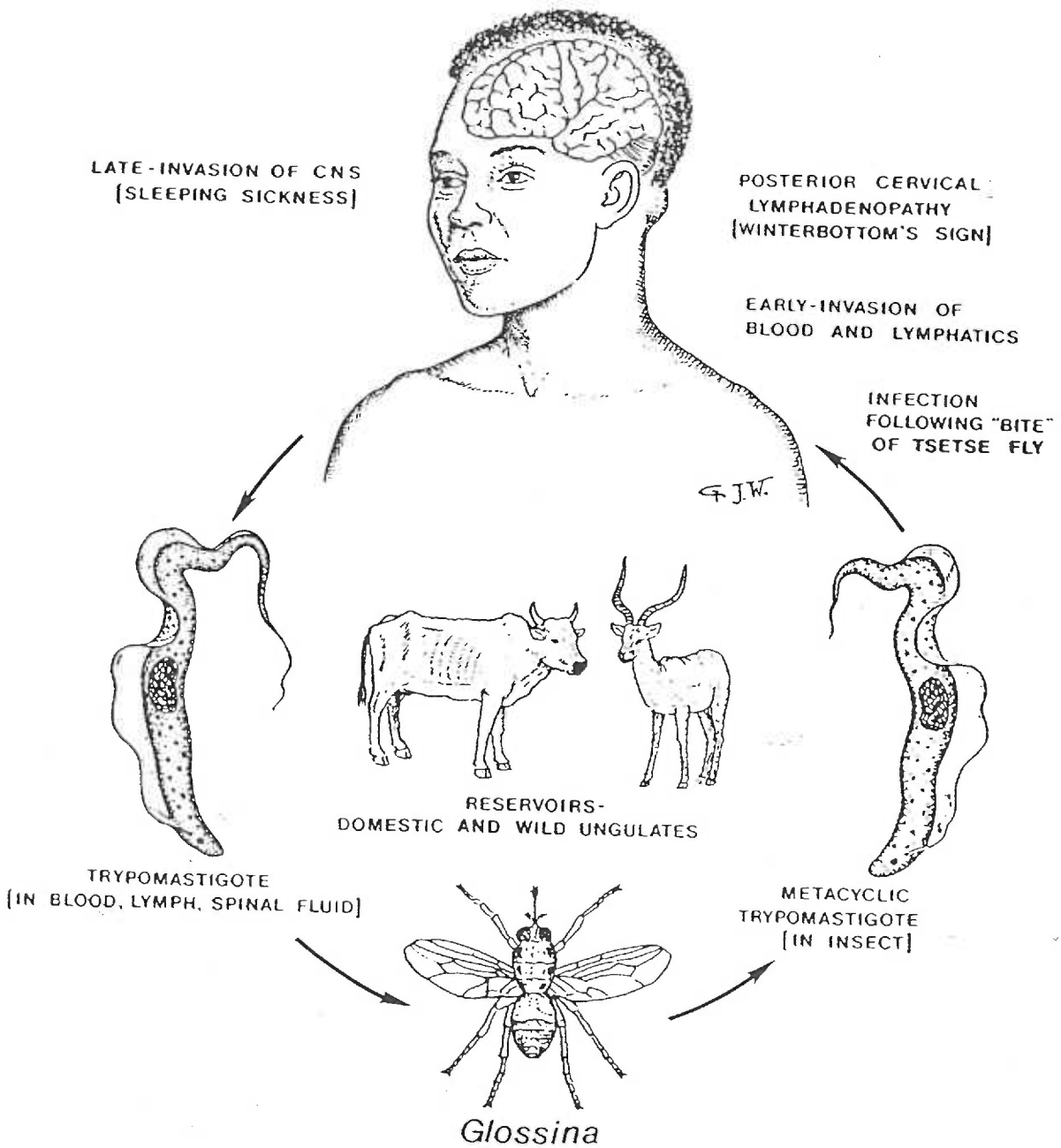
c. Life Cycle

Since the life cycle of the various *Trypanosoma* of Africa are very similar in most respects, this discussion will entail both those species that cause trypanosomiasis in animals and humans. As has been mentioned *Trypanosoma* are both heteroxenous and pleomorphic. In the invertebrate vector, the tsetse fly, the first distinct morphological form that is seen after a blood meal is the Trypomastigote procyclic (Figure 6) form of the parasite which is found in the midgut of the fly where they multiply. The slender procyclics then migrate to the salivary gland where they develop into the epimastigote form of the parasite. Once again the organism reproduces and matures to become the infective metacyclic Trypomastigote. This form is characterized by being short and stumpy and by the absence of an obvious external flagellum. This form appears to be the only life stage in the vector that is capable of infecting vertebrates. The metacyclic Trypomastigotes are injected into the vertebrate host during the blood meal by the vector where they transform into what is described as the slender form of the Trypomastigote. The slender trypomastigote is found in the bloodstream and quickly transforms into the intermediate form of the parasite. It is this form that undergoes reproduction by intensive binary fission in the blood and lymph. The final

bloodstream form, which is termed stumpy, is the form that is infective for the invertebrate vector and when taken up during a blood meal once again become the procyclic trypomastigote form of the organism.

Several important biochemical changes occur between the different life stages of the parasite. One of the most interesting differences that has been observed between the parasite in the invertebrate vector and the parasite in the mammalian host, involves the different means of metabolizing glucose. The slender and intermediate form of the trypomastigotes in the mammalian host do not appear to have a functioning TCA cycle nor are they cyanide sensitive suggesting that they are not undergoing oxidative phosphorylation. This is very different from the parasite in the invertebrate vector where a very active TCA cycle is found and the organisms are sensitive to cyanide (19,20). Another fascinating stage specific feature of the Trypanosoma involves the unique coat of variant-specific surface glycoproteins or VSGs. VSGs are found on the cell surface of the parasite in the vertebrate host but not in the vector (21). It is thought that the VSGs help the parasite evade the hosts immune system and the understanding of the molecular and biochemical events involved in transcription and expression of these genes continues to be a riveting story.

Figure 6. **Life cycle of the African *Trypanosoma***


LATE-INVASION OF CNS
[SLEEPING SICKNESS]
POSTERIOR CERVICAL
LYMPHADENOPATHY
[WINTERBOTTOM'S SIGN]
EARLY-INVASION OF
BLOOD AND LYMPHATICS
INFECTION
FOLLOWING "BITE"
OF TSETSE FLY
RESERVOIRS-
DOMESTIC AND WILD UNGULATES
TRYPOMASTIGOTE
[IN BLOOD, LYMPH, SPINAL FLUID]
METACYCLIC
TRYPOMASTIGOTE
[IN INSECT]
Glossina

d. **Control and Treatment**

The importance of the ability of *Trypanosoma* species to infect the wildlife population of Africa has received much attention in studies of the disease (13). As mentioned above, *Trypanosoma* are capable of using wildlife as a reservoir in which transmission from wildlife stock to domesticated animals and to man is a difficult cycle to break. Early attempts to control the disease focused on simply shooting the infected (and non infected) wild animals to try and interrupt the cycle. This practice was fortunately ended (at least officially) when it was realized the tremendous economic importance of a healthy wildlife population (22). Other dubious but successful programs involved the control of the vector by mass spraying with the chlorinated hydrocarbon insecticide, DDT (23). More recent aerial spraying control programs using endosulfan and pyrethrum proved successful in the short term but were expensive and impossible to continue for indefinite periods of time (23). Additional environmentally unsound control programs included the destruction of the vectors habitat by removing vegetation and by draining infected wetlands. These practices continue to varying degrees today (16,22,23). More encouraging is the recent emphasis on community based, low cost control programs involving tsetse traps and targets in areas endemic for African sleeping sickness. In particular, mono-screen traps impregnated with chemical attractants have proven very successful in areas of Uganda and in northwestern Guinea-Bissau (24,25). Success has also been observed in programs to control animal trypanosomiasis through traps in Zimbabwe (26). In addition, research is underway looking at genetically resistant or trypanotolerant livestock such as the N'dama breed of cattle found

in west Africa (26).

Treatment of human African trypanosomiasis is most effective when started early before central nervous system (CNS) involvement since the most effective and least toxic drugs do not pass the blood brain barrier (27). Currently the two most effective drugs are suramin for the early stages of the disease, and melarsoprol (Mel B) for advance cases.

The pioneering work of Ehrlich and Voegtlin in the early part of this century gave us the first effective drugs for African trypanosomiasis (28). Ehrlich's work on arsenical compounds produced the first useful drug, atoxyl, in 1907. Atoxyl is a pentavalent antimonial which was found to be very active against trypanosomiasis even after CNS involvement. Further work over the next couple of years produced the derivative of atoxyl, tryparsamide in 1912. This was the drug of choice for the next 40 years until Mel B was developed by Freidheim in 1949.

The arsenicals are of two groups, the pentavalent and the trivalent (figure 7). Mel B is an arsenoxide combined with a dimercaprol group which have been used to reverse toxicity of arsenic compounds. In human tissue, Mel B is biotransformed into an arsenoxide which is it's active state. The activity of Mel B and arsenicals in general, is thought to be involved in covalent binding to the parasite through sulfhydryl groups found on the surface of the organism. It has been suggested that the blocking of the sulfhydryl groups may be the cause of the drugs lethal affect on the parasite. Host tissue is also susceptible to binding of Mel B to surface sulfhydryl groups but the selectivity appears to be higher for the parasite then the host. A particularly interesting hypothesis of why this may be the case involves the assumption that the sulphydryl-arsenoxide condensation reaction is reversed at the surface

of the parasite by displacement of the dimercaprol group and the formation of sulfhydryl bonds with the parasite (29). This explanation may also explain why Mel B has proven more effective against the parasite then did earlier arsenical compounds.

Although Mel B is the drug of choice for the treatment of Trypanosomiasis in the later stages, it has some very serious side effects. Encephalopathy, which is thought to be caused by the rapid lethal effect of the drug on the parasite, causes between 1 - 5% mortality in patients treated (11,29). Other possible side effects include damage to the liver, kidney, and more milder symptoms such as skin irritations. One of the greatest benefits of Mel B over the earlier arsenicals is the lack of ocular nerve atrophy which was a common side effect of tryparsamide (11).

The most common drug used today for the treatment of the early stages of trypanosomiasis is suramin which was introduced in 1920. Once again, the work of Ehrlich established the toxic effect that this drug has on Trypanosoma (30). While the drug is much less toxic then the arsenicals, it does not cross the blood brain barrier and is the reason why it is useful only in the earlier stages of the disease. The compound is a non-metallic derivative of urea who's structure is shown in Figure 8. The mechanism of action is still not very well understood however it is believed that glycolytic enzymes may be affected with possible inhibitory effects of the compound on RNA polymerase also playing a role. Suramin is highly effective however. In cases were it is used early enough, a 100% cure rate is seen. Side effects to the drug can be very uncomfortable but for the most part are not life threatening or chronic. Common side effects include vomiting, shock, and loss of consciousness with more serious side effects involving the kidneys (29). The initial health

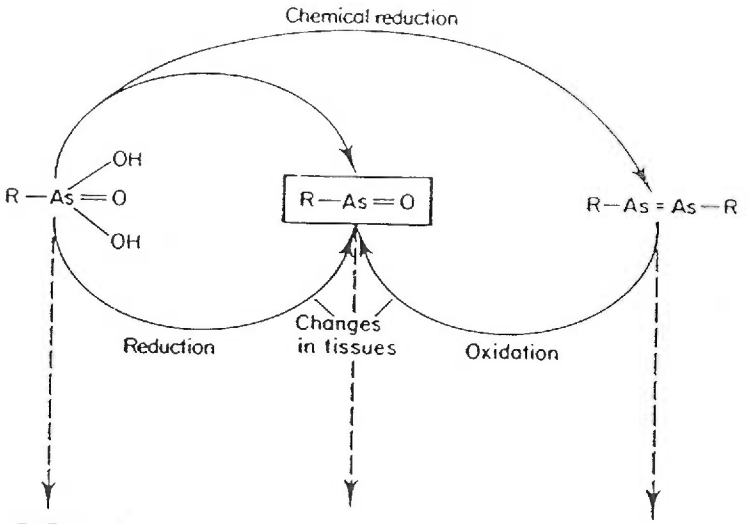
of the patient before treatment correlates well with the severity of the side effects. Malnourished patients in particular show more severe effects to the drug.

Figure 7. **The arsenicals**

a. Comparison of chemical forms as they relate to toxicity and antiparasitic activity

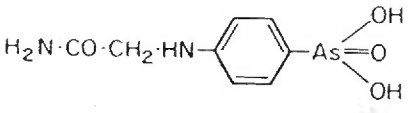
b. Structure of tryparsamide and melarsoprol

(a)

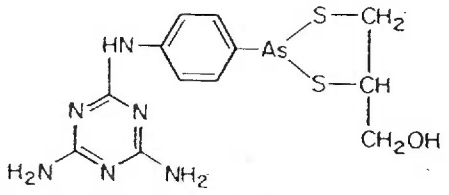


	PHENYLARSONATE (e.g. Atoxyl)	ARSENOXIDE	ARSENOBENZOL
Relative toxicity	+ +	+ + + +	+ + +
Relative *in vitro* activity	0	+ + +	0
Relative *in vivo* activity	+ + Slow	+ + + + Rapid	+ + + Rel. Slow

(b)



Tryparsamide



Melarsoprol

Figure 8. **The structure of suramin**

NaSO3
NaSO3
NaSO3
NH
CO
NH
CH3
CO
HN
CO
NH
CO
NH
H3C
CO
NH
SO3Na
SO3Na
SO3Na

2. **South America Trypanosomiasis**

a. Chagas' Disease

With 35 million people exposed to and 12 to 19 million people infected with, *Trypanosoma cruzi* is one of the great health problems in South America (31,32). Found in nearly every country between latitudes 20° N and 20° S (figure 9), it is a disease of rural populations although with the current trend of migration to urban centers this may change. Reports from Brazil have shown up to 60% of the rural population seropositive with at least 30% of these people exhibiting signs of electro-cardio-graphic abnormalities. In addition to its vector mediated path of infection, *T. cruzi* has also been shown to be infective when found in the blood supply (33). It is not hard to understand why this disease has been given such a high priority by the World Health Organization and the local governments.

First described by Carlos Chagas in 1909, *T. cruzi* is the only pathogenic organism in the Stercoraria group (figure 4). Transmitted by the *Reduviidae* family of arthropods, it is the only known *Trypanosoma* that is transmitted by contamination instead of through the bite of the vector. *T. cruzi* is actually transmitted through the feces of the vector with a fascinating correlation between feeding and defecation timing and the ability of the organism to act as a efficient vector for the infection of mammalian hosts being noted (7). As was mentioned, infection with *T. cruzi* is found most often in rural areas where both domestic and sylvatic reservoirs play a major role in the ability of the parasite to be transmitted to man (1,7). These reservoirs include dogs, cats, rabbits and many species of rodents. Humans are also considered an important reservoir for the parasite. The

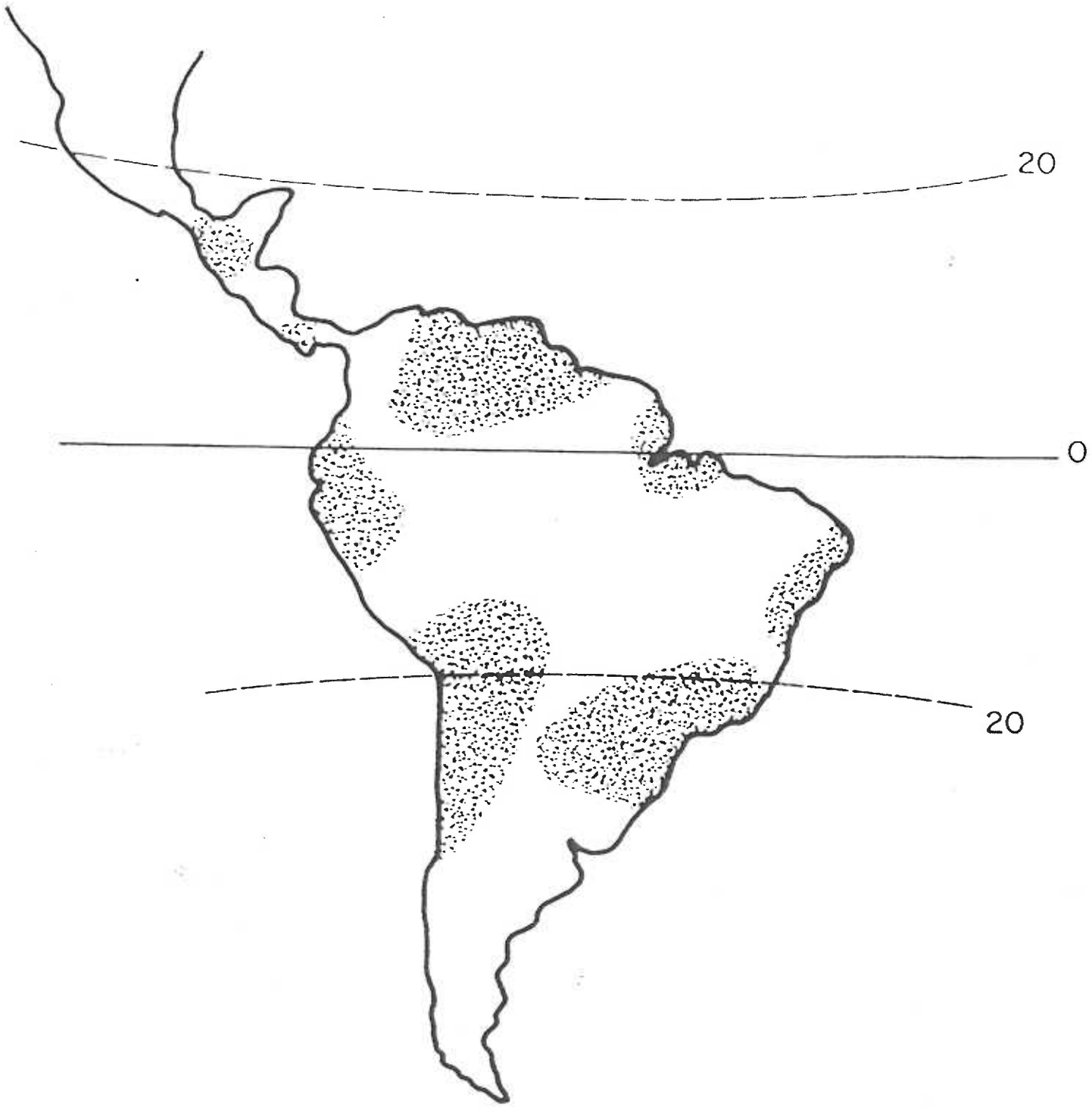
socioeconomic standing of many people in rural areas dramatically increase the ability of the parasite to be transmitted to man. Rural homes very often are constructed of thatch roofs with brick or thatch walls leading to the perfect environment for the vector (34,35). In addition, the *Reduviidae* do not appear to be susceptible to common organochlorine insecticides which make them even more difficult to control (7,36).

Infection usually occurs in early childhood and can persist throughout the life span of the infected individual (11,27,37). The disease, commonly known as Chagas', has an acute and chronic phase. The acute phase is very often asymptomatic particularly in adults. In infants and children however, the acute phase can be very serious (11,38). Infection begins with engorgement of the vector followed by defecation. The feces are then either rubbed directly into the wound by the host or transferred to a mucous membrane, often the conjunctiva of the eye, where the organism is able to enter the body. If the infection occurs at the site of the bite of the vector, a local inflammation is usually seen which is called a chagoma. However well over 50% of all infections appear to occur through the eye leading to edema of the eyelid which is termed Romana's sign. Within 4-12 days after infection, the parasite can be found in any organ of the body with the heart often targeted. Up to 80% of the cardiac ganglion may be destroyed during the acute phase of the disease (1). Symptoms can include anemia, muscle pain and nervous disorders. In serious cases fever may be seen with death caused by encephalitis and myocarditis. Once again, young children and infants show the greatest severity during the acute phases. It should also be noted that *T. cruzi* is capable of crossing the placenta and infecting neonates (34).

The acute phase generally ends between the 2 and 4 months at which time the

infected individual can enter the chronic phase where they may remain asymptomatic for the rest of their lives (38,39). The pathologic manifestations of the disease can occur at any time after entering the chronic phase. The primary cause of death in Chagas' disease is due to heart failure (27,37). Several studies have suggested that 70% of deaths in young adults is due to Chagas'. Other symptoms include central and peripheral nervous system dysfunction and megasyndrome. Megasyndrome results from a destruction of the autonomic ganglia of the esophagus or colon resulting in gross enlargement of the organ which can be fatal.

Figure 9. **Distribution of South American trypanosomiasis**


20
0
20

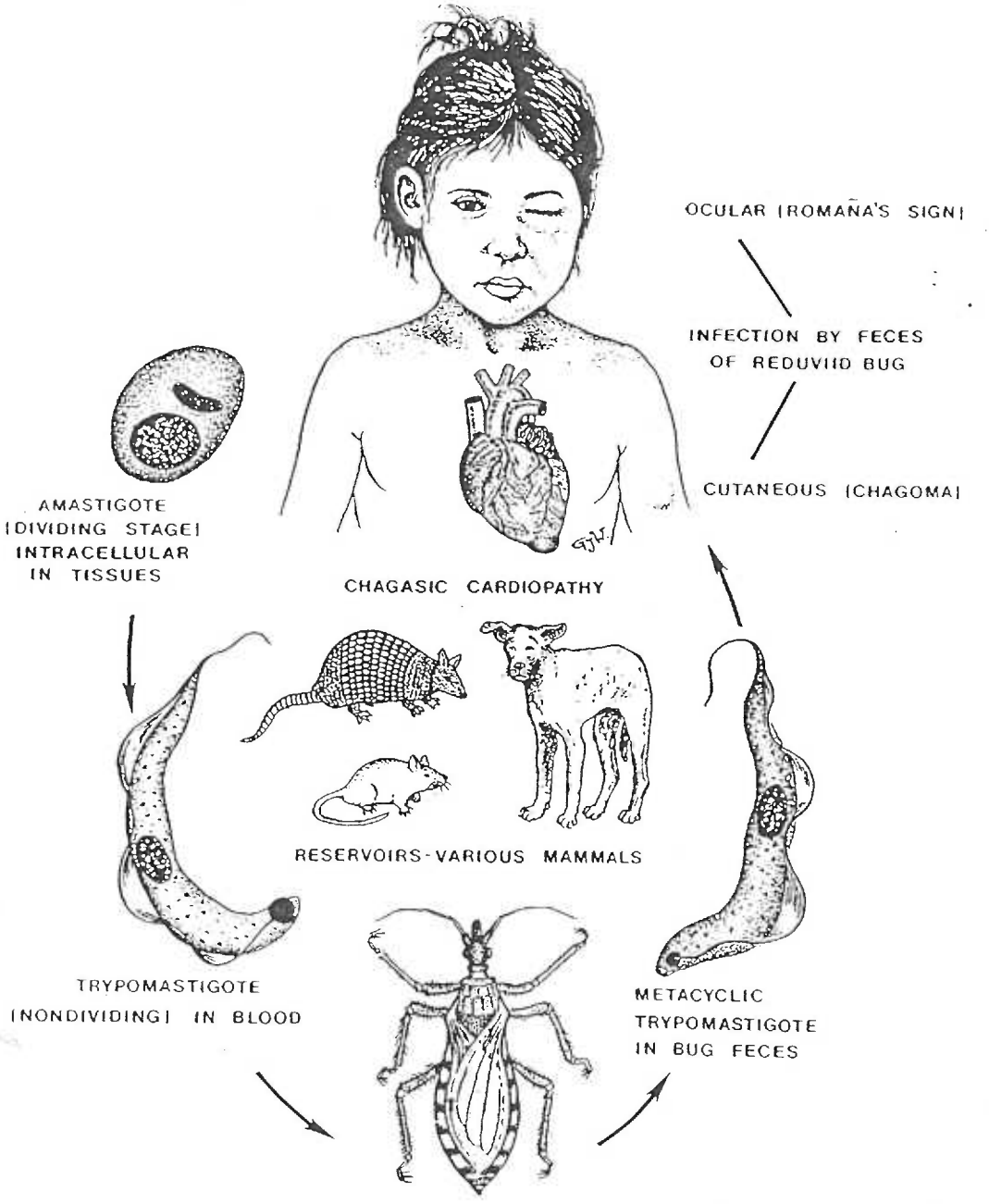
a. **Life Cycle**

As has been eluded to above, *T. cruzi* has a very different life cycle then does its counterparts in Africa (Figure 10). In particular, the route of infection is unique in that *T. cruzi* is passed through the feces of the vector. Once into the host the metacyclic trypomastigotes quickly invade hosts cells where they transform into the amastigote form. It is this form that undergoes rapid binary fission until the cell is lysed releasing the parasite which then can invade other cells and tissue types. There is still debate over the mode of entry of the parasite into hosts cells but it is believed that this occurs either through phagocytosis or active invasion on the part of the parasite (40). The most common cells invaded by the parasite include cells of cardiac and skeletal muscles and reticuloendothelial cells of the liver and spleen. Pseudocysts, which are densely packed regions of amastigotes, are often observed in muscle cells especially in the heart (27). There is a growing body of evidence suggesting that the pathology caused by *T. cruzi* is a result of the immunologic response by the host however more work is necessary to understand the extent of this response (39,41). Intermediate forms such as promastigote and epimastigote can often be found in interstitial spaces. In addition, the trypomastigote form can be found in the blood stream of the mammalian host and it is this form that is taken up by the vector.

Once inside the vector the trypomastigotes migrate to the hindgut and transform into the epimastigote form of the parasite which then undergoes rapid division by binary fission. There is also some suggestion of sexual reproduction at this point in the parasites life cycle, however more experimental evidence is needed to support these reports (42,74). The

epimastigote form is capable of migrating to the rectum of the vector where they attach to the epithelium of the rectal gland. The epimastigote then transform into the infective metacyclic trypomastigote which is found free within the rectal lumen. The life cycle of *T. cruzi* within the vector can be completed in less then 15 days although once the vector is infected with the parasite it remains infected for the rest of it's life. The vector does not seem to be affected by the parasite burden (7).

Figure 10. **The life cycle of *Trypanosoma cruzi***


OCULAR (ROMAÑA'S SIGN)
INFECTION BY FECES OF REDUVIID BUG
CUTANEOUS (CHAGOMA)
AMASTIGOTE (DIVIDING STAGE) INTRACELLULAR IN TISSUES
CHAGASIC CARDIOPATHY
RESERVOIRS - VARIOUS MAMMALS
TRYPOMASTIGOTE (NONDIVIDING) IN BLOOD
METACYCLIC TRYPOMASTIGOTE IN BUG FECES

b. **Control and Treatment**

Since there is currently no acceptable chemoprophylaxes and a staggering 14 million people are infected with *T. cruzi*, control has centered on the vector. This has proven a very difficult task however since there are no adequate pesticides. As was mentioned above, the arthropods of the *Reduviidae* family are very resistant to organochlorine based insecticides such as DDT, while other pesticides including the organophosphorus compounds do not appear to be very effective against the vector and have the additional drawback of being effective for only a short time. Other problems confounding health officials have to do with the rural nature and relatively low economic standing of the people and area's involved. Insecticide spraying needs to be done by knapsack equipment since the interior of the houses need to be sprayed thus making control by this method logistically difficult and expensive. Other methods being explored include the introduction of predators of the vector such as the nematode *Neoaplectana* however many predators being considered are more susceptible to the insecticides currently being used then is the vector making a multiple control approach frustrating. One of the most obvious approaches to controlling the vector would be to improve housing by changing building materials. In areas where the vector is common, change of construction materials for housing have reduced infection rates by *T. cruzi* in the human host from 20% to 50% (36). The major obstacle of course to this approach is cost and if current trends persist this method for control of the vector will continue to be an under used option for the control of Chagas'. Currently there are many programs in progress that are using a variety of combinations of the above approaches and

many of the programs are showing signs of success (43). However, until better ways are found to control the vector and the economic standing of the people infected improves dramatically, Chagas' disease will continue to be a major health concern.

The current treatments for Chagas' disease are even more inadequate then the treatments for African Trypanosomiasis. As with *T. brucei*, clearance from the bloodstream of the trypomastigote form is relatively easy. However with *T. cruzi*, invasion of the cells and tissues of the host can be very rapid with the concurrent transformation of the parasite into the amastigote form which the current battery of drugs do not appear to be effective against. The two compounds currently available for treatment of Chagas' are the nitrofuran, nifurtimox, and the nitroimidazole, benznidazole (figure 11). Both drugs were approved for treatment of Chagas' disease in the late 1970s and have proven successful in clearing the trypomastigote form of the organism from the patient in up to 90% of the cases treated (38).

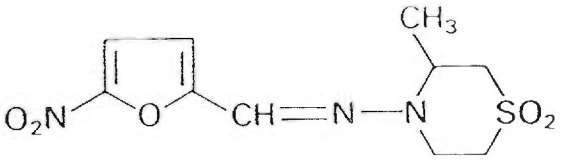
The mechanism of action of nifurtimox is believed to be in the production of highly toxic $-OH^-$ free radicals (29,44). This occurs by the reduction of the nitrofuran to intermediates that can be oxidized to superoxide anions (O^{2-}). This is then followed by the production of hydrogen peroxide (H_2O_2) catalyzed by the action of superoxide dismutase. H_2O_2 can then interact with other superoxide anions to produce the toxic OH^- radicals. Although there are serious side effect with the use of this drug, some specificity is seen. It has been suggested that *T. cruzi* lacks catalase and is not capable of releasing H_2O_2 thus allowing for accumulation of the toxic free radicals in the parasite. It is also believed that the mode of action of benznidazole is also through the production of free radicals but with the finding that both groups of compounds may inhibit protein production and nucleic acid

synthesis, more then one mode of action may be responsible for the toxic effect of these compounds (45,46).

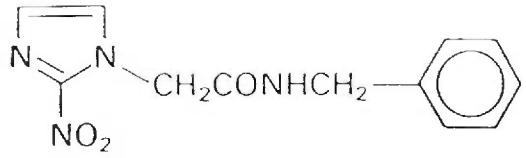
Both benznidazole and nifurtimox are difficult drugs to administer to patients. Long treatment periods (up to 120 days) and the necessity of constant supervision of the patient make these expensive treatments. In addition, serious side effects are seen involving the central nervous system and the gastrointestinal tract in at least 50% of all patients (29). Common side effects include vomiting, anorexia and polyneuritis however these side effects do disappear upon discontinuing treatment. The principal weakness of these drugs is there inability to clear the disease once the organism has entered the host cells. Patients are commonly found seronegative after treatment just to become positive for the parasite after treatment has ceased. Obviously, in the absence of a vaccine, new drugs are desperately needed.

As was mentioned, another important route for transmission of *T. cruzi* is through the blood supply. Currently the treatment of blood with gentian violet (figure 12) has proven successful. Gentian violet however appears to effect the erythrocytes suggesting that more research is needed to remove this pathway of transmission (47,48).

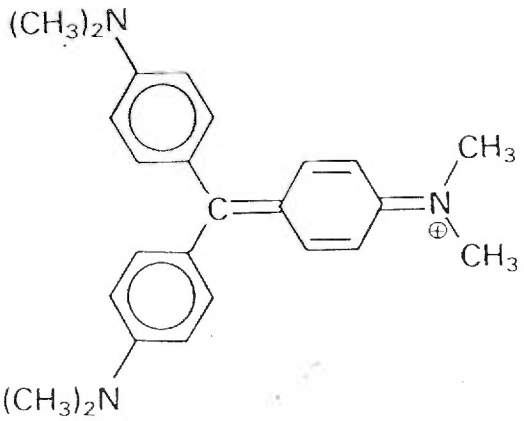
Figure 11. **Compounds used to control *T. cruzi***



Nifurtimox



Benznidazole



Gentian violet

D. ***Leishmania***

The genus *Leishmania* contains six species which are known to cause disease in man (figure 12). Although the individual species produce vastly different disease states, they are morphologically indistinguishable. Indeed there is no strict division between the different organisms and what disease state they cause (7,49). All members of the genus are obligate intracellular parasites who predominantly invade macrophage cells of the host which are the cells intended to rid the body of such invasions. Although not confirmed until 1942, *Leishmania* are all transmitted by the blood feeding subfamily of *Psychodidae*, the *Phlebotominae,* which are more commonly known as the sandfly (50). The disease can be visceral, cutaneous or mucosal in nature. The progression of the infection is still not very well understood but factors that play a role include the parasite strain and the genetic makeup and nutritional state of the host (51,52). As with *Trypanosoma*, species of *Leishmania* are considered zoonotic (6). They infect a wide range of mammalian hosts with the domestic dog and many different species of rodents acting as important reservoirs for the parasite. Although all six species of disease causing *Leishmania* are of importance, I will limit this discussion to the causative agent of visceral leishmaniasis, *Leishmania donovani.*

1. ***Leishmania donovani***

In the early part of the century, visceral leishmaniasis was often misdiagnosed as chronic malaria. It wasn't until 1903 that Drs. Leishman and Donovan described the

causative agent of visceral leishmaniasis, *Leishmania donovani* (7). Found in tropical and subtropical parts of the world, *L. donovani* has a wide distribution pattern (figure 13). No accurate estimation of the prevalence of infection of Leishmania is known but recent reports suggest that approximately 12 million are infected with the parasites and another 350 million people are at risk of infection (50).

The epidemiology of visceral leishmaniasis is very complicated and varies from region to region. In most of its range, the dog is considered to be the most important reservoir for the disease with many studies showing a relationship between canine visceral leishmaniasis and human visceral leishmaniasis (53,54). However in other parts of the world, such as in India, Kenya, and southern China, the dog does not appear to play a major role with humans themselves appearing to be the primary reservoir (50). In the Sudan another pattern emerges suggesting that rodents play a predominant role in acting as a reservoir for the parasite (7). In addition to the above variations, distinct features of the disease have been noted. In particular is the propensity of visceral leishmaniasis in mediterranean countries to be isolated to infants (*L.d. infantum*) (7). The reason for this is still unclear and the significance in doubt but a wide range of clinical differences have been noted.

Figure 12. **Members of the genus *Leishmania***

Important Leishmania Species

Species	Disease	Geographic Distribution
Cutaneous Leishmaniasis		
L.tropica minor	Dry Cutaneous	South Europe, Middle East
L.tropica major	Wet Cutaneous, oriental sore	South Europe, Middle East Africa
L.braziliensis	Espundia, Mucocutaneous	Mexico-Brazil
L.m.mexicana	Chicleros ulcer, cutaneous	Central America
L.m.amazonensis	Diffuse,Cutaneous	Amazonas Region
Visceral Leishmaniasis		
L.d.donovani	Kala-azar	Africa,Asia,USSR,S.America
L.d.infantum	Visceral,infantile	Mediterranean
L.d.chagasi	Visceral	South America

Figure 13. **Distribution of visceral leishmaniasis**

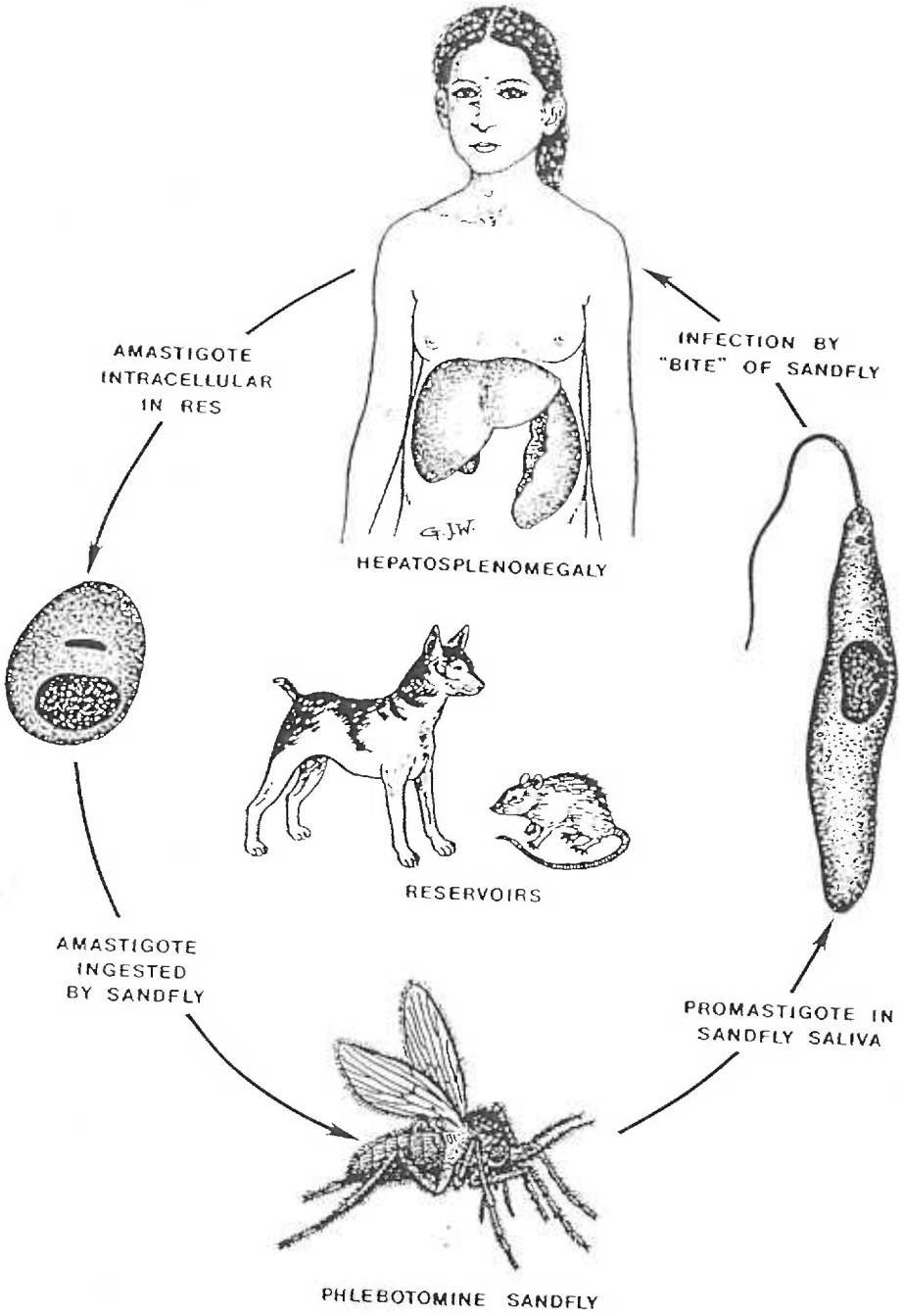
a. **Life cycle**

Considering the fact that the predominate cell type invaded by the parasite in the mammalian host is the macrophage, the life cycle of members of the genus *Leishmania* is fascinating (figure 14). Injected at the bite site of the vector, the parasite is swiftly taken up by macrophage that are attracted to the wound created during the blood meal. The organism is found within the phagocytic vacuoles where they transform into the amastigote form of the organism. It is this form that undergoes rapid division by binary fission until the cell is lysed allowing the parasite to spread to surrounding macrophage. In visceral Leishmaniasis, the parasite is disseminated throughout the body with the spleen, bone marrow and liver most seriously affected (7,11).

When the vector takes a blood meal from an infected host, macrophage containing amastigotes can be ingested where they are released from the macrophage during digestion in the gut of the vector. The parasite transforms back into the promastigote form and migrates to the midgut were it undergoes division. The mature promastigotes then migrate to the anterior of the vectors gut were it can then be transmitted once again to the mammalian host. The entire cycle in the fly can take as little as ten days although very little is actually known about development in the vector. An interesting distinction in the development of *Leishmania* in the vector as compared to the development of *Trypanosoma* is that *Leishmania* appear to be capable of infection at all stages in the vector were trypomastigotes of *Trypanosoma* are only infective when they develop into metacyclic trypomastigotes (7).

The fascinating development of the parasite in the host has given rise to intense studies of the interaction between the host and the parasite. Experiments which exploit the murine model to study the pathogenesis of leishmaniasis have given us insight into the immunological response of the host to invasion. Work is currently being pursued to understand the effect of the various cytokines on the course of the infection (55-57). Of particular importance is to understand how *Leishmania* can survive within the macrophage of the host. Several recent studies have suggested that the parasite may affect the ability of the macrophage to produce lysosomal proteases or the ability of these enzymes to gain entrance to the parasitophorous vacuole (58). However other studies have shown that the parasitophorous vacuole contains all the enzymes normally found in a lysosome and that the survival of the organism is based on proteins found on the cell surface of the parasite (59). In particular, the cell surface metalloproteinase gp63 has received considerable attention recently for its ability to protect proteins in lysosomes that have been coated with this protein (60). This area of research is important from both a basic biology standpoint and also as a possible target for both chemotherapeutic and vaccine intervention in the survival strategy of the parasite.

Figure 14. **Life cycle of *Leishmania* species**


AMASTIGOTE
INTRACELLULAR
IN RES
INFECTION BY
"BITE" OF SANDFLY
HEPATOSPLENOMEGALY
RESERVOIRS
AMASTIGOTE
INGESTED
BY SANDFLY
PROMASTIGOTE IN
SANDFLY SALIVA
PHLEBOTOMINE SANDFLY

b. **Control and Treatment**

Success in the control of visceral leishmaniasis has been more of a side benefit to the large scale control programs for malaria. The use of DDT to control the Tsetse fly which transmits malaria, also helped to control the sandfly. In fact spraying with insecticides helped to virtually eliminate kala-azar from India between 1950 and 1970 (61). However with the scaling back of the use of DDT and with no other substitute as effective, both malaria and leishmaniasis have increased in incidence (62). In China and areas of the Mediterranean, control programs involving the dog helped immensely in reducing the cases of leishmaniasis (7,63). During the cultural revolution in China (1940-1960s), eradication programs involved simply shooting any dog that was found. Many other simple methods of reducing the risk of becoming infected have been used for centuries. These include migration from areas of high incidence and the use of insect repellents and bed netting. Current control programs involve the use of insecticides and treatment of any individual infected with the parasite. This approach, especially in the Bihar region of India, has meet with mixed success (64,65). Once again the economics of developing new treatments and the intense poverty of the people affected has played a predominant role in assuring that the disease continues to be a major health problem.

Visceral leishmaniasis can manifest a considerable variety in symptoms and in the timing of the onset of those symptoms. In general however there is an acute and chronic division in how the disease progresses in man. The acute form of the disease is first noticed as a high fever with generally two peaks each day. The appearance of fever with a double

spike is probably the reason why leishmaniasis was often misdiagnosed as chronic malaria. Vomiting and diarrhea is often associated with the acute form of the disease with the formation of hepatosplenomegaly followed by death if the disease is not treated. The acute form of the disease can often kill its victim in 6 to 12 months whereas the chronic form of the disease can take up to 3 years (7,11). The chronic form is generally not diagnosed until the disease is relatively advanced. Fever, diarrhea and vomiting are also noticed with the chronic form but often more intermittent. In fact the fever can disappear for long periods but then can reappear at any time. Anemia along with the gradual wasting of the individual is seen in both acute and chronic cases. In the chronic form of the disease this often allows for secondary infections which can kill the patient long before they succumb to the parasite (1,27). Other symptoms commonly seen include leucopenia and high levels of protein in the blood with the vast majority of this protein being gamma globulins. Whether the patient has acute or chronic manifestations of the disease, if left untreated, visceral leishmaniasis is fatal in 90% of all cases (7).

Kala-azar also shows dramatic clinical variations between different regions where it is found. One of the more striking examples is seen in India where up to 20% of all recovered patients exhibit post-kala-azar dermal leishmaniasis (PKDL) (7,66,67). PKDL usually occurs within 1 to 2 years after the patient is successfully treated or in rare cases of self recovery. It manifests itself by the appearance of hypo or hyperpigmentation with the formation of popular and nodular lesions which can become ulcerated. Although lesions can form on any part of the body, the face seems to be particularly affected causing gross disfigurement. The nodules are packed with macrophage containing large numbers of

parasites although the spleen, liver and bone marrow can remain completely free. The large quantity of parasites can act as a rich reservoir of parasites for the continuation of the organisms life cycle. Interestingly however, kala-azar patients in other parts of the parasites range show very low incidence of PKDL with patients in Africa showing no higher incidence rates then 5% (7). Currently there is very little information to explain the disparity between different regions but variations need to be taken into consideration when decisions on treatment are made. One suggestion for the high rate of PKDL may be due to incomplete clearance of the parasite from the patient due to inadequate length of treatment or suboptimal dosage of chemotherapy (68).

Current treatment for visceral leishmaniasis is still based on antimonials that were first introduced in 1912 by Dr. Viannas. The use of antimonials can be documented at least as far back as the 16th century when it was used for its supposed effectiveness as a emetic (29). Use of the compounds did not continue for very long due to the extreme toxicity that was noted. However antimonials were reintroduced because of their effectiveness in treating leishmaniasis although the problem of toxicity of these compounds remained. The first compound used was potassium antimonyl tartrate which proved to be extremely toxic. Due to the work that was on going at the time on arsenical compounds, it was known that the pentavalent forms of the arsenicals were less toxic and this proved to be true for the antimonials as well. The derivatives of pentavalent antimonials that have proven most useful are the antimonic acid derivatives (figure 15). Although the exact structure remains unknown, it is believed that in these compounds the antimony group is attached through a oxygen molecule to a carbon chain that consists of many hydroxyl groups. The extreme

complexity of these compounds have defeated researchers attempts at ascertaining the mode of action of these drugs. It is not clear what form is actually the active form of the compounds and little evidence is available that suggests that they are biotransformed to the more toxic trivalent form as with the arsenicals. It is known that the decrease in toxicity of the pentavalent antimonials is due to the rapid clearance of the compounds from the patient (29). This has it drawbacks however since treatment is required more often and for a longer period of time.

The two pentavalent antimonials currently used are sodium stibogluconate (pentostam) and N-methylglutamine antimonate (glucontime). The difference in use is more a regional preference than any specific attribute of the two compounds with pentostam being the most commonly used. These compounds are given intravenously and hospitalization is necessary. It is interesting to note that variations are once again seen between different regions involving the effectiveness of these compounds with a much higher dose required in the Sudan for effecting a cure than in India (7). In addition, children often need much larger doses relative to body weight than do adults (69).

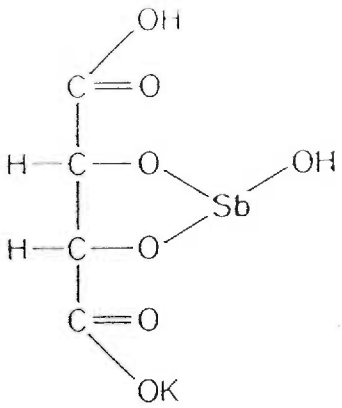
The pentavalent antimonials remain the most effective treatment for visceral leishmaniasis, however several other compounds are now used for cases when these compounds are not effective. The most useful alternative is the compound pentamidine which is an aromatic derivative of synthalin, a diamidine (figure 16). Pentamidine is used most often in cases of antimonial resistance (29). The mechanism of action is still unclear but it is thought to involve the ability of the compound to bind to DNA. Indeed, studies have suggested that pentamidine may bind to kDNA however this does not appear to affect

reproduction of the organism in the mammalian host and so it is believed that other mechanisms must be involved (70). It is well established that the activity of pentamidine is related to the spatial positioning of the two pairs of amidine groups which depends on the bridge between the two aromatic rings. Toxicity of this compound is relatively low especially when given intramuscularly where serious nephrotoxic side effects that the compound can create do not appear. Other compounds that are used include amphotericin B which was isolated from the bacteria *Streptomyces* (figure 16). Used for its antifungal properties, amphotericin B is extremely toxic and is only used as a last resort particularly in cases of advance mucosal leishmaniasis and rarely in advance cases of visceral leishmaniasis.

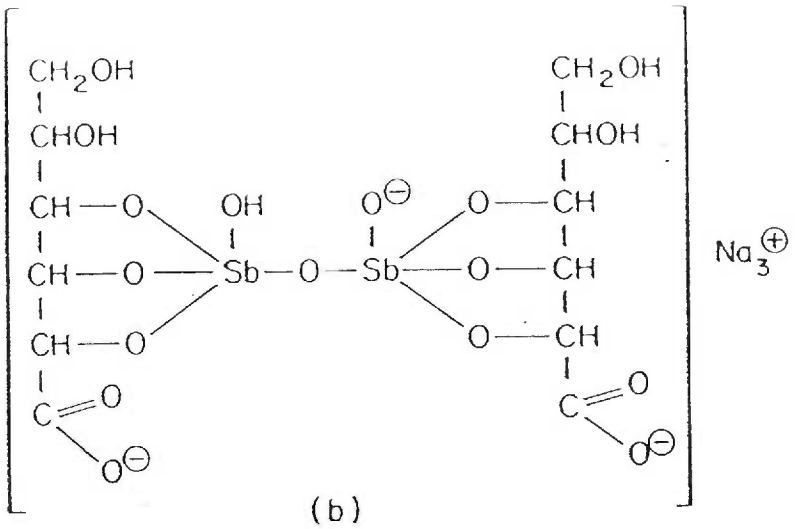
Figure 15. **The antimonials**

a. Tartar emetic

b. Sodium stibogluconate

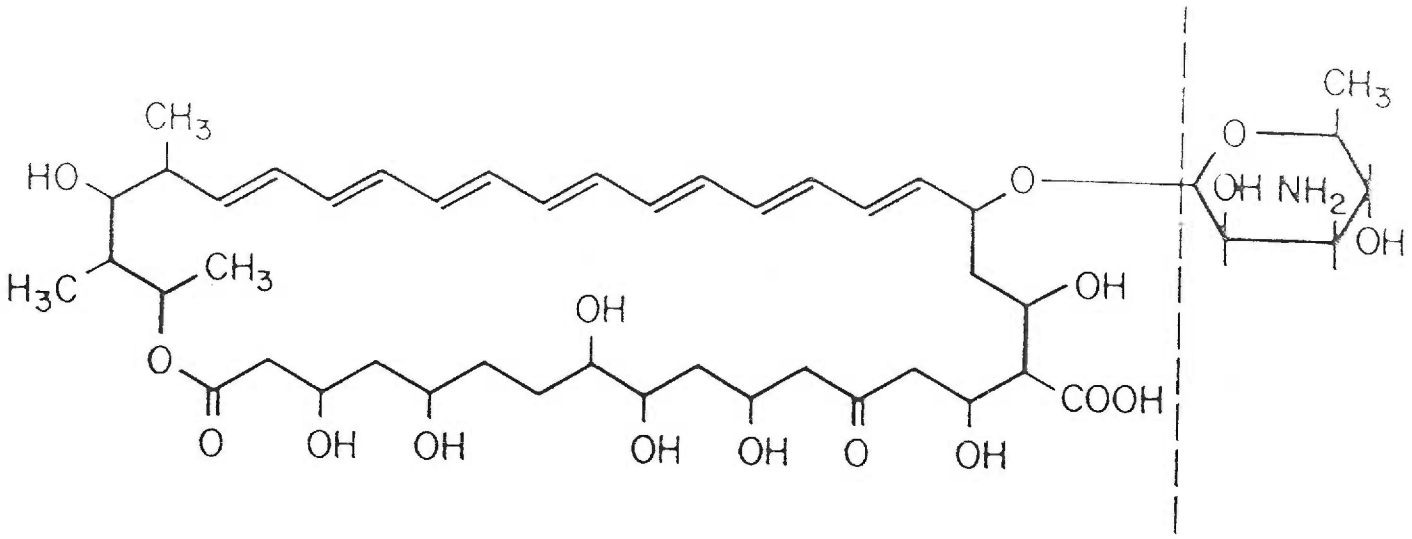


(a)

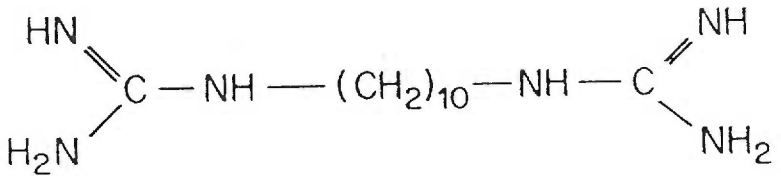


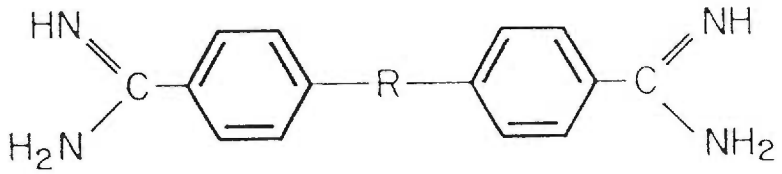
(b)

Figure 16. **The structures of pentamidine, diminazene and amphotericin B**



Amphotericin B





Pentamidine R = —O·$(CH_2)_5$·O—

Diminazene R = —NH—N=N—

E. **Approaches to Drug Development**

The search for effective drugs is as old as mankind has been on this earth. Early hunter-gatherers looked for naturally occurring compounds that would relieve pain, that could be used in religious rituals and that would help in the hunting of food. These early discoveries were based on observations of their surroundings and random testing. This has continued to be the pattern for drug discovery until the second half of this century although as will be discussed, random testing is still an important technique.

The advent of modern drug development can be traced back to the introduction of organic chemistry over 170 years ago (71). This discipline and the vast number of unique compounds that were produced rapidly displaced the screening of natural compounds. In addition, the increase in our understanding of basic biochemistry along with insights into how drugs are metabolized and excreted (pharmacokinetics/pharmacodynamics) from the human body has lead to dramatic increases in the number of useful drugs available. However with all the advances science had made, the discovery of "lead" compounds still remained the most time consuming and frustrating step in the discovery of new drugs (72).

In the early part of the 20th century, Paul Ehrlich helped to revolutionize the way that we thought about the treatment of diseases and how we viewed drugs (73). His insight into drug/target interactions helped to revolutionize the way scientists looked at how drugs affect the body. In addition, his emphasis on selectivity of drug action became the basis for all future work on chemotherapeutic agents. His contribution to the treatment of parasitic

diseases, with the introduction of arsenical for trypanosomiasis and salvarsan for the treatment of syphilis, laid the foundation for much of the advances in chemotherapy over the next 50 years. Indeed, the most important advance in the treatment of disease so far this century was the development of sulphonamides (1938) which was a direct extension of Ehrlicks earlier work.

But it was the discovery by Dr. Woods and coworkers of the mechanism of action for the sulfonamides (1962) that firmly established the ability to study drug action at the molecular level (75). It is to this date that the modern science of drug design can be traced. The discovery that the sulfonamides exert their action by selectively interfering with bacteria reaffirmed Ehrlicks earlier emphasis on exploiting differences in the biology between the disease causing agent and the patient. This discovery also helped pave the way for a more rational approach to the development of drugs. This included a thorough investigation of the pharmacological parameters of the chemotherapeutic agent and detailed examination of the biology of the disease causing agent.

Drug design based on modern understandings of pharmacology, biochemistry and molecular biology have made great gains since 1963. The ability to modify current lead compounds through molecular or pharmaceutical modification has proven very successful (28,72). In addition, understanding the basic biochemistry of the disease has allowed for the development of drugs based on analogs of naturally occurring metabolites. One of the best known examples of the development of effective chemotherapeutic agents by metabolic analogy is the use of allopurinol in the treatment of hyperuricemia and gout (76).

During the last several decades numerous advances in our ability to identify specific

genes and to clone, express, and purify the proteins they encode has allowed for highly specific in vitro assays to be developed permitting detailed analysis of the protein products (77). Our ability to ascertain protein structure and to determine other physical parameters has also progressed significantly (78,79). Finally, the advances made in computer technology has been astounding. It is the last of these advances which has made the idea of a multi-discipline approach to a rational design of new chemotherapeutic agents a possibility. In particular, the combination of protein structure determination and structure-based inhibitor design promises great advances in our ability to tailor specific chemotherapeutic agents to combat disease (80). Indeed, several examples now exists for the design of inhibitors to targets for which the structure is not even available (81). The power of employing computers is demonstrated with programs such as DOCK, which allows for the screening of over 500,000 compounds in a single day. When these type of computer generated simulations are then combined with advances in automated in vitro activity assays, the time from determination of a possible target to lead drug generation can be drastically reduced.

As has been mentioned, the first step in a rational approach to drug design is to determine a significant difference between the disease causing agent and the affected organism. For parasitic diseases, the search for unique attributes of parasite metabolism has been the major emphasis in the exploration of these organisms. Figure 17 lists just a few of the unique aspects of parasites that are being examined. Current interest is focused on pathways that are involved in energy production through ATP synthesis as well as proteins involved in DNA, RNA and protein synthesis. In particular, polyamine biosynthesis and purine metabolism have been identified as two exploitable metabolic pathways. The

development and current clinical trials of difluoromethylornithine (DFMO), which inhibits the enzyme ornithine decarboxylase in the polyamine pathway, is one of the few new drugs that has been introduced in decades (82). Promising results have also recently been obtained involving novel inhibitors for thymidylate synthetase, an enzyme involved in DNA production, which were developed using the DOCK program mentioned above (83). As our understanding of the basic biology of parasites continues to grow, the list of potential sites for chemotherapeutic intervention is certain to increase.

Figure 17. **Possible targets for chemotherapeutic intervention**

Potential Sites for Cheomotherapuetic Intervention

Glycolysis

- compartmentallization: Glycosomes
- Stage specific expression

Electron Transport

- Compartmentiziation: Mitochndria
- Kinetoplast RNA editing process

Pyrimidine Metabolism

- Biosynthesis

Purine Metabolism

- Transport and metabolism: HGPRTase, 5'and 3' nucleotidase

Folate Metabolism

- DHFR-TS

Polyamine Metabolism

- Ornithine decarboxylase and DFMO

Thiol Metabolism

- Trypanothione reductase

Proteinases

- Protection, Invasion and Metabolism: gp63, Serine and cystein proteiases

Glycoproteins

- GPI Anchors

F. Purine Metabolism

As was mentioned earlier, the first requirement in a rational approach to the control of parasitic infections is to ascertain a unique feature about the parasites biology and then design strategies to exploit that difference. Of the metabolic pathways that have been studied in parasitic protozoa, the purine pathway is one of the most unique in that all parasitic protozoa investigated to date lack the ability to produce purines through the *de novo* pathway (84-92). To meet their purine requirement, each genus of parasite has developed a unique series of enzymes to salvage purines from their environment. This fundamental difference between the mammalian host and the parasite has created an interest in the enzymes involved in purine metabolism in *Kinetoplast*.

1. Purine Salvage in Trypanosmatides

Little variation is seen between the *Trypanosoma* and *Leishmania* in the way that they salvage purines although some differences are seen between the different life stages (93). Figure 18 shows an outline of the relevant enzymes involved in purine salvage including the interconverting enzymes (94). Several studies have suggested that the major pool of purines for the parasite in the mammalian host is probably the adenylate pool (95,96). Both 5' and 3' nucleotidase activity have been found on the cell surface of *Leishmania* and *Trypanosoma* along with acid phosphatase activity thus allowing for the transport of nucleosides into the cell (97,98,99). Two separate nucleoside transporters have been characterized in *L. donovani* (100). The first is capable of transporting inosine and guanosine and the second transports adenosine and pyrimidines. Two separate transporters have also been reported in *T. cruzi*

(101). Once the nucleoside is within the parasite several pathways are available for its conversion into the nucleotide pool.

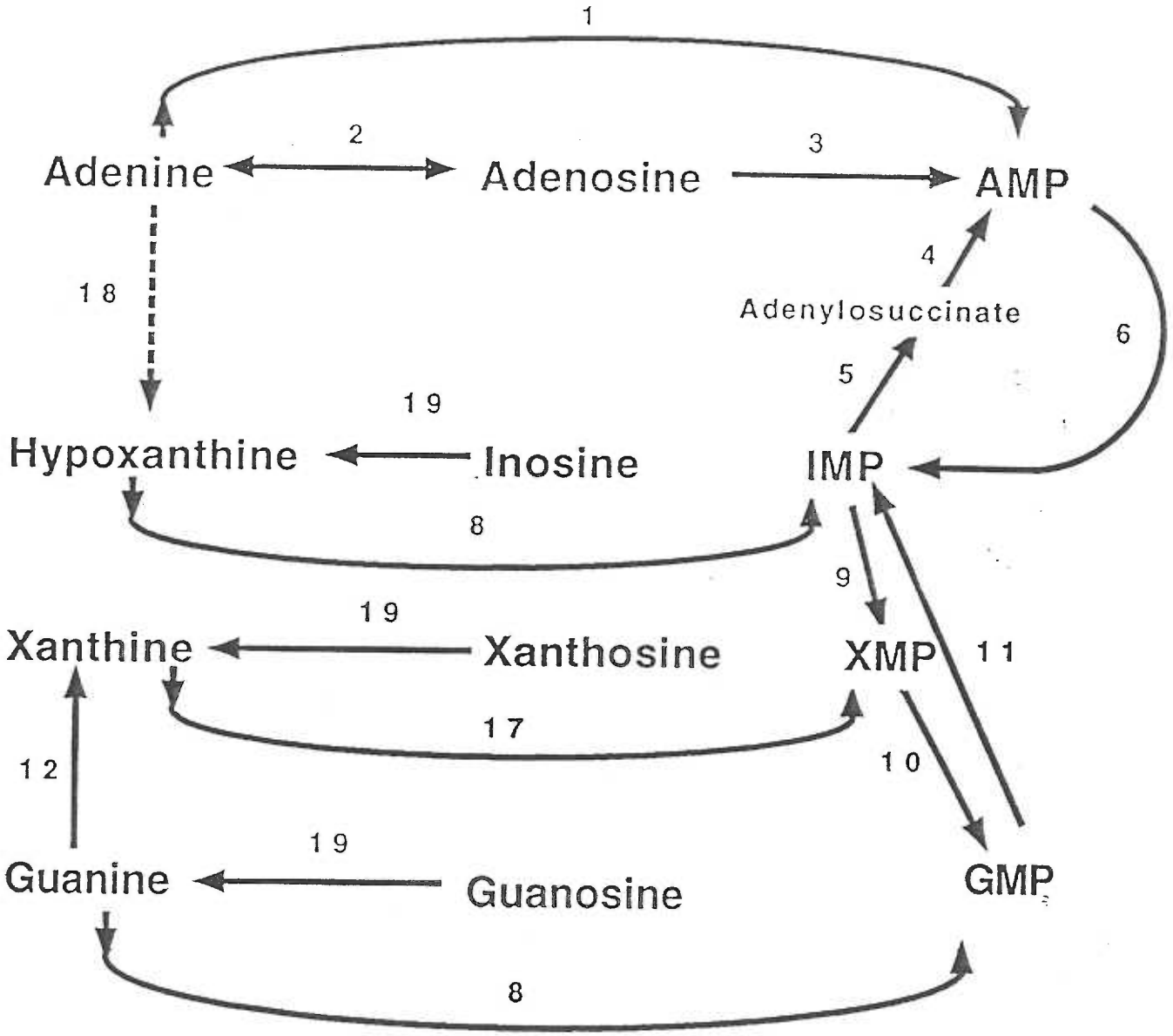
In *Leishmania* promastigotes, adenosine can be phosphorylated directly to adenosine 5'-monophosphate (AMP) by the activity of adenosine kinase or it can be cleaved to adenine through nucleoside phosphorylase activity. Adenine can then be deaminated to hypoxanthine through the activity of adenine deaminase or phosphoribosylated to AMP through adenine phosphoribosyltransferase (APRT). Studies using radiolabeled substrates have suggested that the major pathway for adenosine metabolism in this form of the parasite is through adenine deaminase to hypoxanthine (84,98). However, several studies have suggested that in the *Trypanosoma* and the amastigote form of *Leishmania*, adenine deaminase may be absent (88,90,93,102). In addition, adenosine kinase activity appears to be much higher in the amastigote form of *Leishmania* and the presence of adenosine deaminase activity, which deaminates adenosine to inosine, has also been detected (94,103). Inosine can then be converted to hypoxanthine through nucleoside hydrolase activity which has been detected in all three parasites (104,105,106). Hypoxanthine is then converted to the nucleotide level through the enzyme hypoxanthine-guanine phosphoribosyltransferase. It should be noted that conflicting studies have been published on the presences or absences of adenine deaminase in all three organisms. The metabolic fate of adenosine is probably the most important dissimilarity between the *Trypanosoma* and *Leishmania* however further studies are needed to clarify this question.

Leishmania and *Trypanosoma* are also capable of salvaging inosine, guanosine, and xanthosine although the efficiency does vary between species. In addition to hypoxanthine

and adenine mentioned above, guanine and xanthine are also salvaged. In the salvage of each of these ribonucleosides and/or purine bases, the enzymes involved in phosphoribosylation play a central role in the efficient salvage of purines for these parasites.

Figure 18. **Purine metabolism in Trypanosomatids**

1. APRT, 2. nucleoside phosphorylase, 3. adenosine kinase, 4. adenylosuccinate lyase, 5. adenylosuccinate synthetase, 6. AMP deaminase, 7. adenosine deaminase, 8. HGPRT, 9. IMP dehydrogenase, 10. GMP synthetase, 11. GMP reductase, 12. guanine deaminase, 17. XPRT, 18. adenine deaminase, 19. purine nucleoside hydrolase


1
2
3
Adenine
Adenosine
AMP
18
4
Adenylosuccinate
6
5
19
Hypoxanthine
Inosine
IMP
8
9
19
Xanthine
Xanthosine
XMP
11
17
12
10
19
Guanine
Guanosine
GMP
8

2. **The Phosphoribosyltransferase**

Because of the prominent role that the phosphoribosyltransferases play in purine metabolism, they have been the object of considerable study. All members of the *Trypanosomatidae* family contain a phosphoribosyltransferase (PRTase) for adenine, hypoxanthine/guanine and xanthine. The presence of a separate PRTase for xanthine is one of the significant differences between the mammalian host and the parasite in that the mammalian cell does not appear to have a PRTase for xanthine. The PRTase's catalyzes a condensation reaction of a purine base with a 5-phosphoribosyl-1-pyrophosphate group (PRPP) (figure 19).

The observation that the enzyme HGPRT is capable of metabolizing the purine analog allopurinol (HPP) to allopurinol ribonucleotide (HPPR-MP) has lead to speculation that the enzyme HGPRT may be a possible site for chemotherapeutic intervention (107). This metabolic action of the enzyme HGPRT is lethal to the parasite since HPPR-MP is capable of being incorporated as the adenylate analog into RNA where it is believed to cause blockage of protein synthesis (108). Mammalian cells are incapable of this metabolic conversion (109,110).

Allopurinol is an isomer of hypoxanthine which has an inversion of the nitrogen in position 7 to position 8 producing 4-hydroxypyrazolo[3,4-d]pyrimidine. It is a member of a group of purine analogs called the pyrazolopyrimidines (fig 20). Allopurinol has been used for many years in the treatment of hyperuricemia and gout and has been found to be a

suicide inhibitor of xanthine oxidase an enzyme not found in parasitic protozoa (111,112). In the mammalian cell, the majority of allopurinol is converted to oxipurinol by xanthine oxidase (figure 21). Oxipurinol is a much more potent inhibitor of xanthine oxidase. A small amount of allopurinol is excreted in the urine or is converted to HPPR-MP which has been detected in the tissue of rats however no evidence of the incorporation of allopurinol into the nucleotide pool exists (110).

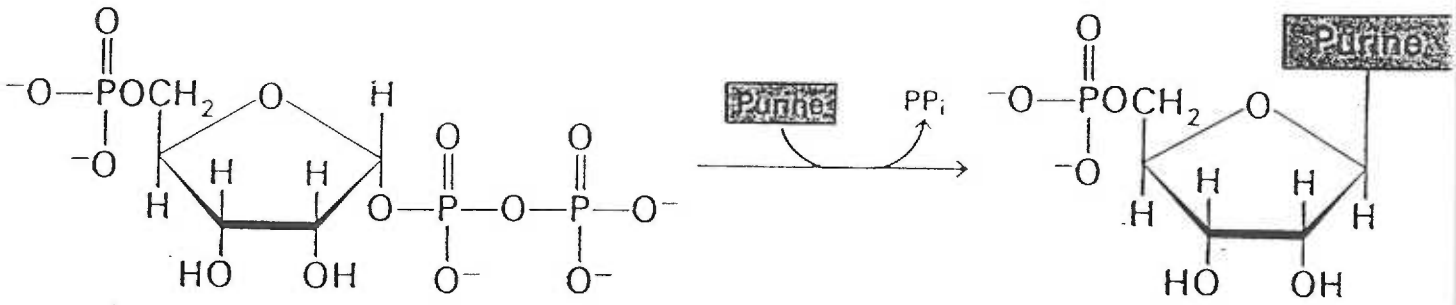
Several studies have shown that all three parasites are capable of concentrating HPP against a gradient (113-116). It has also been shown that HPP is converted to HPPR-MP where it is then aminated to 4-amino-pyrazolopyrimidine riboside 5'-monophosphate (APPR-MP) followed by phosphorylation and incorporation into the parasites RNA but not into the DNA. Although no definitive experiment has been performed to show that the cytotoxic affect of HPP is related to its incorporation into RNA, several studies have given strong indication that this is probably true (106,113,117-119). The addition of adenine to parasites grown in culture with the addition of HPP has been shown to decrease the toxicity of HPP. However, when the nucleotide pools of these parasites is examined there does not appear to be a decrease in purines due to HPP addition and so it is unlikely that HPP is cytotoxic due to purine starvation. In addition, if adenine is given to the parasites 12 hours after exposure to HPP, the cytotoxic affect is not reversed. This suggests that the addition of adenine is not circumventing a block but may be competing with HPP for conversion to the nucleotide level. It has also been shown that the addition of hypoxanthine to HPP treated parasites can also block the transformation of HPP to HPPR-MP. This along with kinetic studies on partially purified HGPRT from L. *donovani* and purified recombinant HGPRT

benznidazole. Even more encouraging is a recent report on the effect of allopurinol on cutaneous leishmaniasis. This study reported a 70% cure rate in those patients receiving allopurinol as compared to a 36% cure rate for patients who received treatments of meglumine antimonate (Glucantime). In addition, patients receiving allopurinol exhibited a very low rate of adverse reactions.

Several of the other pyrazolopyrimidines have also been found to be cytotoxic to trypanosomatids. Allopurinol ribonucleoside (HPPR) has been shown to be very effective in inhibiting the growth of all three parasites although the degree of effect varies (108,116,127,128). Of particular interest in the metabolism of HPPR is the possible role of a nucleoside phosphorylase activity since HPPR does not appear to be cleaved to HPP unlike its isomer, inosine. Formycin A, formycin B and thio derivatives of allopurinol and allopurinol riboside have all shown cytotoxic effects on the trypanosomatids but much less information is available as to their effect in the mammalian host although investigations are in progress (98,108,129,130).

Clearly, the enzyme HGPRT is worthy of further study. Its prominent role in purine salvage and its unique substrate specificity fulfill Ehrlicks requirements for choosing a target that can selectively be exploited. In addition, the availability of a lead compound, allopurinol, should shorten the time necessary to design new and more effective drugs. It is my hope that the development of both biochemical and molecular tools for the study of the HGPRT enzyme and the purine salvage pathway in general, will allow for the development of new and better drugs to control parasitic disease.

Figure 19. **Phosphoribosylation reaction**

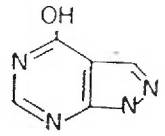


PRPP

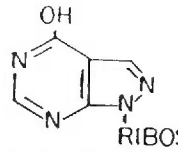
Purine ribonucleotide

Figure 20. **The pyrazolopyrimidines**

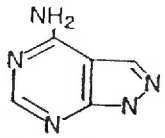
PYRAZOLOPYRIMIDINES



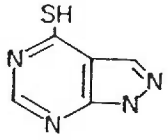
Allopurinol (HPP)



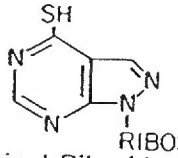
Allopurinol Riboside (HPPR)



Aminopurinol (APP)

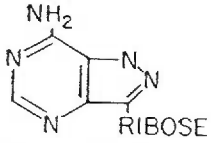


Thiopurinol (TPP)



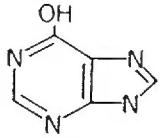
Thiopurinol Riboside (TPPR)

Formycin B

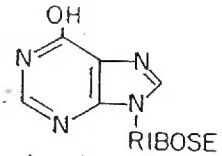


Formycin A

PURINES

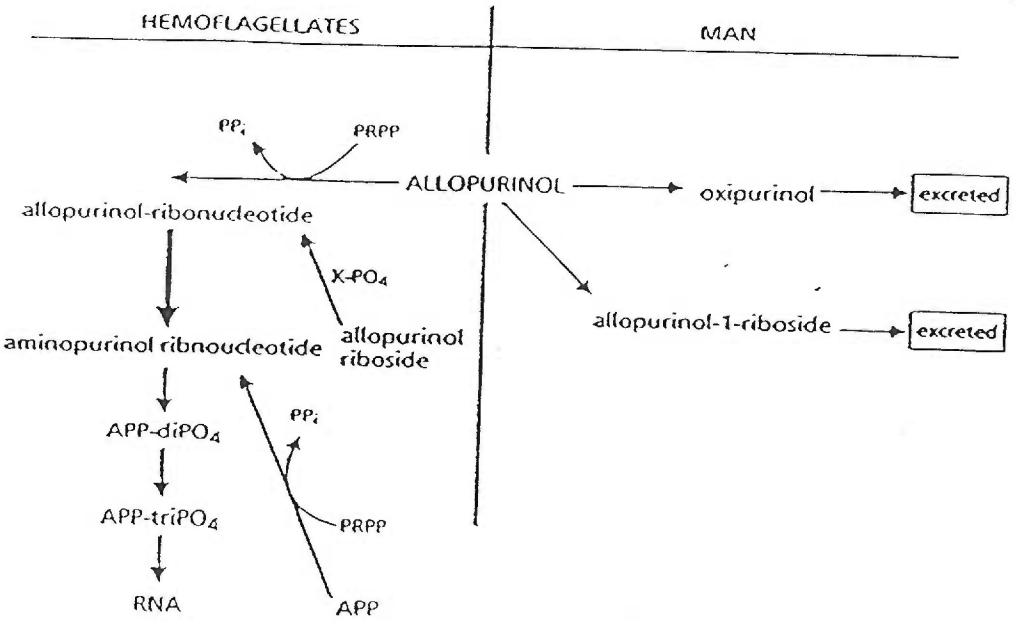


Hypoxanthine



Inosine

Figure 21. **Comparison of allopurinol metabolism in the trypanosomatids and their mammalian hosts**


HEMOFLAGELLATES
MAN
PP$_i$
PRPP
ALLOPURINOL
oxipurinol
excreted
allopurinol-ribonucleotide
X-PO$_4$
allopurinol-1-riboside
excreted
aminopurinol ribnoucleotide
allopurinol riboside
APP-diPO$_4$
PP$_i$
APP-triPO$_4$
PRPP
RNA
APP

G. Specific Aims

1) **To isolate and sequence the gene *hgprt* from two members of the *Trypanosomatidae* family**

a)*Trypanosoma brucei* - Oligonucleotide primers where designed to two conserved regions of the *hgprt* gene as determined by optimum alignment between human, *Schistosoma mansoni* and *Plasmodium falciparum hgprt* genes. Polymerase chain reaction (PCR) technology was used to amplify a 174 bp fragment from *Trypanosoma brucei* genomic DNA which was cloned and sequenced to confirm correct amplification. This fragment was then used to isolate a full length copy of the *hgprt* gene from a EMBLE3 genomic library which was sequenced in its entirety.

b)*Trypanosoma cruzi* - After alignment of the *Trypanosoma brucei* gene with the three genes listed above, new primers where designed and PCR was again used to amplify a fragment from genomic *T. cruzi* DNA. This fragment was used to isolate a full length genomic clone from a Lambda gt-11 *T.cruzi* library which was sequenced in its entirety.

2) **To characterize both the gene and the mature transcripts for the *Trypanosoma hgprt***

a)Southern analysis - Genomic DNA from both organisms was digested with several enzymes. The DNA was separated by agarose gel electrophoresis and the separated DNA was transferred to a nylon membrane. The DNA was probed with the appropriate full length clone at high stringency. Analysis of the hybridization signals suggest that the *T. brucei hgprt*

gene is not in a tandem repeat but the pattern suggests that there may be more than one copy of the gene. The hybridization signal for the *T. cruzi* gene suggests that there probably is just one copy of the *hgprt* gene.

b)Northern analysis - RNA was isolated and separated on a agarose/formaldehyde gel. The separated RNA was then transferred to a nylon membrane and hybridized with the appropriate probe. The hybridization pattern for *T. cruzi* suggested a single transcript of 1.8 kb while the *T. brucei* showed two transcripts at 1.4 kb and 1.9 kb. Additional analysis of the *T. brucei* mature transcript revealed that the heterogeneity of the size of the transcript was due to differences in the 3' end of the transcript.

c)Analysis of 5' end of mature message - A fascinating feature of the order *Kinetoplastida* is that many, if not all, transcription events occur by first producing a polycistronic pre-mRNA precursor which is then cleaved and trans-spliced with a 39 nucleotide sequence that is common to all mature mRNAs. We can take advantage of this common sequence or "splice leader" region in PCR experiments to map the limits of the 5' end of the mature message. Reverse transcription experiments were performed with RNA from both organisms to produce cDNA which was then followed by PCR amplification using a primer complimentary to the splice leader sequence and an antisense primer 3' of possible start sites of translation. The PCR product was then cloned and sequenced. This information was valuable in deciding the first methionine in the protein sequence and allowed us to make decisions involving the construction of our expression vector.

3) **To express *hgprt* in an expression system and purify the recombinant protein**

Both the *T. cruzi* and *T. brucei hgprt* genes have been engineered using PCR to allow for subcloning into the prokaryotic expression vector pBAce. This expression vector uses the bacterial alkaline phosphatase promoter which is derepressed by low phosphate growth conditions and has been used successfully to express several other proteins at high levels. The recombinant construct was then transformed into *E.coli* strain S0606. This cell line is genetically deficient in both hypoxanthine phosphoribosyltransferase (HPRT) and xanthine-guanine phosphoribosyltransferase (XGPRT) activity, thus greatly simplifying our protein purification process. A GTP affinity column was used to purify the recombinant protein and it has been our experience to routinely obtain 100 to 200 mg of purified recombinant protein for each liter of *E. coli* cells grown.

4) **To analyze the kinetic parameters of the recombinant proteins**

a)Km and K_{cat} Determinations - Using a sensitive radioenzymatic assay employing C_{14} labeled substrates, the Km and K_{cat} values for several of the possible substrates for HGPRT were determined. Possible substrates included hypoxanthine, guanine, adenine, xanthine, and allopurinol an isomer of hypoxanthine. Values for PRPP were not obtained due to the rapid loss of activity of the pure enzyme in the absence of PRPP. The results from these experiments confirmed the unique substrate specificity of the *Trypanosoma* enzymes. Unlike the human, *P. falciparum* and *S. mansoni* HGPRT enzymes, the

trypanosomas HGPRT is capable of conversion of the purine analog allopurinol to the nucleotide level. This unique action of the trypanosomas has been shown to be toxic to this group of organisms. These experiments also show that unlike the *P. falciparum* HGXPRT enzyme which metabolizes xanthine, the trypanosomas HGPRT does not which supports experimental data that there is a separate xanthine phosphoribosyltransferase in these organisms.

c)Metabolic inhibition studies - Several metabolites are known to affect the activity of HGPRT in both a negative and positive regulatory nature. Using ^{14}C hypoxanthine the extent to which a variety of anionic metabolites affect the activity of HGPRT was determined.

d)Effects of pH on activity - In order to determine at what pH the recombinant enzyme exhibits the greatest amount of activity, 20uM ^{14}C hypoxanthine in TMD (20mM Tris, 5mM $MgCl_2$, 1mM DTT) buffer at various pH levels were incubated with freshly purified recombinant protein and phosphoribosylation activity monitored.

e)Stability studies - An important parameter to establish for the recombinant protein is its stability over time, at different temperatures and at various glycerol concentrations. These experiments are particularly important since the stability of the protein will determine feasibility of using stored protein in crystallization experiments. These studies suggest that the recombinant protein is very stable at temperatures between 4° and -70° C.

e)Quaternary structure analysis - Much confusion exists in the literature on the possibility that the enzyme HGPRT exits as a multimeric homodimer when in it's active form. To determine the multimeric configuration of the trypanosomas HGPRT protein,

freshly purified recombinant HGPRT was analyzed over a size exclusion column using an HPLC. Our results suggest that the native state of the enzyme is a dimer.

5) To isolate the native HGPRT protein from *Leishmania donovani* promastigotes

a) Protein purification and sequencing of amino terminus -Originally it was our hope that with isolating the *T. brucei* and *T. cruzi* genes for *hgprt* we would be able to use these as probes for isolating the *L. donovani* gene. This however has not proven to be as easy as we hoped. In order to clone the *L. donovani* gene, the native HGPRT protein was partially purified from *L. donovani* promastigotes using a slight variation on the purification scheme that was used for the recombinant proteins discussed above. This purification scheme allowed us to purify enough protein for amino terminus sequencing.

b)Isolation of gene for HGPRT - cDNA was produced using purified total nucleic acids from the promastigote form of *L. donovani*. Using a degenerate primer designed to the protein sequence (part a) and a primer to the SLS of the mature message (part 2-C), the 5' end of the mature message for the hgprt gene was amplified using PCR. The amplified product was then used to probe a L.donovani cosmid library at high stringency. The full length genomic clone for *L. donovani hgprt* was isolated and sequenced in it's entirety.

c)Expression and purification of recombinant protein - The *hgprt* gene from *L. donovani* was engineered using PCR and subcloned into the expression vector pBAce. Enzymatically active HGPRT was obtained using this system.

7) Phylogenetic Analysis Using HGPRT as a Pattern Sequence

Much interest has been given to the evolutionary relationship between organisms using DNA, RNA and protein sequences as an indicator of association. With the isolation of the *hgprt* gene, it is possible to use this enzyme as a model for constructing a phylogeny. Several different algorithms were used to determine the relationship of the trypanosomatids between each other and between other eukaryotic and prokaryotic organisms for which *hgprt* gene has been isolated.

References

1. Schmidt, G.D. & Roberts, L.S. (1989) Foundations of Parasitology, Times Mirror-Mosby, Toronto

2. Howard, L.M. (1971) Journal of Parasitol. **57**, 143-147

3. Hyde, J.E. (1990) Molecular Parasitology, Van Nostrand Reinhold, New York

4. Mills, J. & Masur, H. (1990) Scientific America **263**, 50-57

5. Porter, S.B. & Sande, M.A. (1992) N.Eng.J.Med. **327**, 1643-1648

6. Englund, P.T. & Sher, A. (1988) The Biology of Parasitism, Alan R. Liss, New York

7. Molyneux, D.H. & Ashford, R.W. (1983) The Biology of Trypanosoma and Leishmania, Parasites of Man and Domestic Animals, Taylor and Francis, london

8. Borst, P. (1991) TIG **7**,

9. Blum, B., Bakalara, N. & Simpson, L. (1990) Cell **60**, 189-198

10. Feagin, J.E., Shaw, J.M., Simpson, L. & Stuart, K. (1988) Proc.Natl.Acad.Sci.USA **85**, 539-543

11. Markell, E.K., Voge, M. & John, D.T. (1992) Medical parasitology, W.B.Saunders, Philadelphia

12. Orskov, E.R. (1987) Pro.Nutr.Soc. **46**, 301-308

13. Allosopp, R. (1972) Bulletin of the World Health Organization **47**, 735-746

14. Holmes,P.H. & Torr,S.J. (1988) Outlook on Agriculture, **17**, 54-60

15. Ford, J. (1971) The Role of Trypanosomiasis in Africa Ecology. A Study of the Tsetse fly problem, Clarendon Press, Oxford

16. Wellde, B.T., Chumo, D.W., Reardon, M.J. & Smith, D.H. (1989) Annal.Trop.Med.Parasit. **83,Supplement No.1**, 1-11

17. Cibulskis, R.E. (1992) Parasitology **104**, 99-10

18. Mihok, S., Otieno, L.H. & Darji, N. (1990) Parasitology **100**, 219-233

19. Opperdoes, F.R. (1987) Ann.Rev. Microbiol **41**, 127-151

20. Turrens, J.F., Bickar, D. & Lehninger, A.L. (1986) Molecular and Biochemical Parasitology **19**, 259-264

21. Boothroyd, J.C. (1985) Ann.Rev.Microbiol. **37**, 475-502

22. Olando, J. & Mabus, D. (1989) Annal.Trop.Med.Parasit. 83,Supplement No.1, 13-20

23. Oloo, F., Njogu, A.R., Opiyo, E.A. & Mugutu, S. (1989) Annal.Trop.Med.Parasit. 83,Suppliment No.1, 119-125

24. Okoth, J.O. (1986) Parasitology Today **2**, 88

25. Okoth, J.O. (1991) Annal.Trop.Med.Parasit. **85**, 309-314

26. Holmes, P.H. (1991) Trans.Royal Soc.Trop.Med.Hyg. **85**, 163-167

27. Katz, M., Despommier, D.D. & Gwadz, R. (1989) Parasitic Diseases, Springer-Verlag, New York

28. Foye, W.O. (1981) Principles of Medicinal Chemistry, Lea and Febiger, Philadelphia

29. James, D.M. & Gilles, H.M. (1985) Human Antiparasitic Drugs, John Wiley and Sons, New York

30. Meshnick,S.R. (1984) Parasitic Disease, Marcel Dekker, New York

31. Dias, J.C.P. (1987) Parasitology Today **3**, 336

32. Morbidity and Mortality Weekly Report (1992) **41**, 691

33. Kirchhoff, L.V. (1989) Ann.Intern.Med. **111**, 773

34. Azogue, E. (1985) Trans.Royal Soc.Trop.Med.Hyg. **79**, 176-180

35. Briceno-Leon, R. (1987) Parasitology Today **3**, 384-387

36. Garcia-Zapata, M.T. & Marsden, P.D. (1992) Am.J.Trop.Med.Hyg. **46**, 440-443

37. Andrade, Z.A. (1983) Ciba Found Symp **99**, 214-233

38. Marr, J.J. & Docampo, R. (1986) Reviews of Infectious Diseases **8**, 884-903

39. Macedo, V. (1980) J.Brasilian Med. **38**, 34-40

40. McCabe, R.E. (1984) Infect.Immun. **46**, 372-376

41. Takle, G.B. & Hudson, L. (1989) Current Topics in Microbiology and Immunology **145**, 81-91

42. Gibbons,A. (1992) Science **21**, 919-920

43. Schofield, C.J. & Dias, J.C.P. (1991) Mem.Inst.Oswalda Cruz **86**, 285-295

44. Docampo, R., Moreno, S.N.J., Stoppani, A.O.M., Leon, W., Cruz, F.S., Villalta, F. & Muniz, R.F.A. (1981) Biochem.Pharmacol. **30**, 1947-1951

45. Moreno, S.N.J., Docampo, R., Mason, R.P., Leon, W. & Stoppani, A.O.M. (1982) Ach.Biochem.Biophys. **218**, 585-591

46. Gonzalez, N.S. & Cazzulo, J.J. (1989) Biochem.Pharmacol. **38**, 2873-2877

47. Kerndt,P.R., Waskin, H.A., Kirchoff,L.V., Steurer, F., Waterman, S.H., Nelson, J.M., Gellert, G.A., Shulman, I.A. (1991) Transfusion **31**, 814-818

48. Letelier, M.E., Rodriguez, E., Wallace, A., Lorca, M., Repetto, Y., Morello, A., Aldunate, J. (1990) Exp.parasitol. **71**, 357-363

49. Kreutzer, R.D. (1983) Am.J.Trop.Med.Hyg. **32**, 703-715

50. World Health Organization Technical Report Series (1990) **793**,

51. Cerf, B.J., Jones, T.C., Badaro, R., Sampaio, D., Teixeira, R., Johnson, W.D. (1987) J.Infec.Disease **156**, 1030-1033

52. Jahn,A., Lemett,J.M., Diesfeld, H.J. (1986) J.Trop.Med.Hyg. **89**, 91-104

53. Dye, C. (1992) Parasitology **104**, S7-S18

54. Abranches, P., Siva-Pereira, M.C., Conceicao-Silva, F.M. & Janz, J.G. (1991) Journal of Parasitol. **77**, 557-561

55. Boom, W.H., Liebster, L., Abbas, A.K., Titus, R.G. (1990) Infect.Immun. **58**,3863-3870

56. Kaye, P.M., Curry, A.J. & Blackwell, J.M. (1991) J.Immunology **146**, 2763-2770

57. Ho, J.L., He, S.H., Rios, M.J. & Wick, E.A. (1992) J.Infec.Disease **165**, 344-351

58. Chakraborty, P. & Das, P.K. (1989) Biochemical Medicine and Metabolic Biology **41**, 46-55

59. Prina, E., Antoine, J.C., Wiederanders, B. & Kirschke, H. (1990) Infect.Immun. **58**, 1730-1737

60. Chaudhuri, G., Chaudhuri, M., Pan, A. & Chang, K.P. (1989) J.B.C. **264**, 7483-7489

61. Elias, M., Mizanur Rahman, A.J.M. & Khan, N.I. (1989) Bulletin of the World Health Organization **67**, 43-49

62. Rahman, K.M. & Islam, N. (1983) Bulletin of the World Health Organization **61**, 113-116

63. Guan, L.R. (1991) Bulletin of the World Health Organization **69**, 595-601

64. TDR (1990) TDR news **34**, 1-7(Abstract)

65. Marinkelle, C.J. (1980) Bulletin of the World Health Orginization **58**, 807-818

66. Thakur, C.P., Kumar, K. (1992) Annal.Trop.Med.Parasit. **86**, 355-359

67. Thakur, C.P. (1984) Trans.Royal Soc.Trop.Med.Hyg. **78**, 391-398

68. Thakur,C.P. & Kumar, K. (1990) Indian J.Med.Res. **91**, 144-148

69. Chulay, J.D., Anzeze, E.M., Koech, D.K. & Bryceson, A.D. (1983) Trans.R.Soc.Trop.Med.Hyg. **77**, 717-721

70. Looker,D.L., Bernens, R.L. & Marr, J.J. (1984) J.Cell.Biochem. **7A**, 16

71. Burger, A. (1970) Medicinal Cemistry, Wiley-Interscience, New York

72. Hamner, C.E. (1990) Drug Development, CRC Press, Boston

73. Ehrlich, P. (1909) Ber.Ddtsch.Chem.Ges. **42**, 17

74. Cibulskis, R.E. (1988) Parasitology **96**, 303-322

75. Woods, D.D. (1962) J.Gen.Microbiol. **29**, 687-702

76. Hitchings, G.H. (1975) Arthr.Rheum. **18**, 863-870

77. Sarubbi, E. (1991) Febs Lett. **279**, 265

78. Stout, G.H., Jensen, L.H. (1989) X-Ray Structure Determination, John Wiley and Son, New York

79. Johnson, W.C. (1990) Proteins **7**, 205-214

80. Appelt, K. & et al. (1991) J.Medic.Chem. **34**, 1925-1934

81. Cohen, F.E., Gregoret, L.M., Amiri, P., Aldape, K., Railey, J. & McKerrow, J.H. (1991) Biochemistry **30**, 11221-11229

82. Bacchi, C.J., Nathan, H.C., Clarkson, A.B., Bienen, E.J., Biotonti, A.J., McCann, P.P. & Sjoerdsma, A. (1987) Am.J.Trop.Med.Hyg. **36**, 46-52

83. Shoichet, B.K., Stroud, R.M., Santi, D.V., Kuntz, I.D. & Perry, K.M. (1993) Science **259**, 1445-1450

84. Marr, J.J., Berens, R.L. & Nelson, D.J. (1978) Biochim.Biophys. Acta **544**, 360-371

85. Heyworth, P.G., Gutteridge, W.E. & Ginger, C.D. (1978) J. Protozool. **25**,

86. Sherman, I.W. (1979) Microbiol.Rev. **43**, 453-495

87. Boonlayangoor, P., Albach, R.A. & Booden, T. (1980) Arch.Invest.Med. (Mex) **11**, 83-88

88. Berens, R.L., Marr, J.J., LaFon, S.W. & Nelson, D.J. (1981) Molecular and Biochemical Parasitology **3**, 187-196

89. Wang, C.C. & Simashkevich, P.M. (1981) Proc.Natl.Acad.Sci.USA **78**, 6618-6622

90. Fish, W.R., Marr, J.J. & Berens, R.L. (1982) Biochim.Biophys. Acta **714**, 422-428

91. Schwartzman, J.D. & Pfefferkorn, E.R. (1982) Exp.Parasitol. **53**, 77-86

92. Wang, C.C. & Aldritt, S. (1983) J.Exp.Med. **1158**, 1703-1712

93. Looker, D.L., Berens, R.L. & Marr, J.J. (1983) Mol.Biochem.Parasitol. **9**, 15-23

94. Berns, R.L., Krug, E.C. & Marr, J.J. (1992) UnPublished

95. Allen, T., Henschel, E.V., Coons, T., Cross, L., Conley, J. & Ullman, B. (1989) Mol.Biochem.Parasitol. **33**, 273-281

96. Cohen, M.B., Maybaum, J. & Sadee, W. (1980) J.Chrom. **198**, 435-441

97. Dweyer, D.M. & Gottlieb, M. (1984) Mol.Biochem.Parasitol. **10**, 139-150

98. Campbell, T.A., Zlotnick, G.W., Neubert, T.A., Sacci, J.B. & Gottlieb, M. (1991) Mol.Biochem.Parasitol. **47**, 109-118

99. Gottlieb, M. & Dwyer, D.M. (1982) Mol.Cell.Biol. **2**, 76-81

100. Aronow, B., Kaur, K., McCartan, K. & Ullman, B. (1987) Mol.Biochem.Parasitol. **22**, 29-37

101. Finley, R.W., Cooney, D.A. & Dvorak, J.A. (1988) Mol.Biochem.Parasitol. **31**, 133-140

102. Fish, W.R., Looker, D.L., Marr, J.J. & Berens, R.L. (1982) Biochim.Biophys. Acta **719**, 223-231

103. Datta, A.K., Bhaumik, D. & Chatterjee, R. (1987) J.Biol.Chem. **262**, 5515-5521

104. Miller, R.L., Sabourin, C.L.K., Krenitsky, T.A., Berens, R.L. & Marr, J.J. (1984) J.Biol.Chem. **259**, 5073-5077

105. Schmidt, G., Walter, R.D. & Konigk, E. (1975) Trans.Royal Soc.Trop.Med.Hyg. **26**, 19-26

106. Koszalka, G.W. & Krenitsky, T.A. (1979) J.Biol.Chem. **254**, 8185-8193

107. Frank, O., Baker, H. & Hutner, S.H. (1970) J.Protozool. **17**

108. Marr, J.J. & Berens, R.L. (1983) Mol.Biochem.Parasitol. **7**, 339-356

109. Nelson, D.J. & Elion, G.B. (1975) Biochem.Pharmacol. **24**, 1235-1237

110. Nelson, D.J., Bugge, C.J.L., Krasny, H.C. & Elion, G.B. (1973) Biochem.Pharmacol. **22**, 2003-2022

111. Pfaller, M.A. & Marr, J.J. (1974) Antimicrob.Agents Chemother. **5**, 469-472

112. Dewey, V.C. & Kidder, G.W. (1973) J.Protozool. **20**, 678-682

113. Nelson, D.J., Bugge, C.J.L., Elion, G.B., Berns, R.L. & Marr, J.J. (1979) J.B.C. **254** 3959-3964

114. Marr, J.J., Berns, R.L. & Nelson, D.J. (1978) Science **201**, 1018-1020

115. Spector, T. (1977) Biochem.Pharmacol. **26**, 355-358

116. Berens, R.L., Marr, J.J. & Brun, R. (1980) Mol.Biochem.Parasitol. **1**, 69-73

117. Marr, J.J., Berens, R. (1983) Molec.Biochem.Paraasitol. **7**, 339-356

118. Tuttle, J.V. & Krenitsky, T.A. (1980) J.B.C. 255, No3, 909-916

119. Marr, J.J. & Berens, R.L. (1973) J. Protozool. **20**, 678-682

120. Marr, J.J. & Berens, R.L. (1977) J.Infec.Disease **136**, 724-732

121. Berens, R.L., Marr, J.J., Nelson, D.L. & LaFon, S.W. (1980) Biochem.Pharmacol. **29**, 2397-2398

122. Berens, R.L., Marr, J.J., Crux, F.S. & Nelson, D.J. (1982) Antimicrob.Agents Chemother. **22**, 657-661

123. Avila, J.L. & Avila, A. (1980) Exp.Parasitol. **51**, 204-208

124. Avila, J.L., Avila, A. & Munoz, E. (1981) Am.J.Trop.Med.Hyg. **30**, 769-774

125. Gallerano, R.H., Marr, J.J. & Sosa, R.R. (1990) Am.J.Trop.Med.Hyg. **43**, 159-166

126. Martinez, S. & Marr, J.J. (1992) New Engl.J.Med. **326**, 741-744

127. Nelson, D.J., LaFon, S.W., Tuttle, J.V., Miller, W.H., Miller, R.L., Krenitsky, T.A., Elion, G.B., Berens, R.L. & Marr, J.J. (1979) J.Biol.Chem. **254**, 11544-11549

128. Desjardins, R.E., Coser, R.A., Willet, G.P., Childs, G.E. & Confiled, C.J. (1980) Exp.Parasitol. **50**, 260-271

129. Carson, D.A. & Chang, K.P. (1981) Biochem.Biophys.Res.Commun. **100**, 1377-1383

130. Robinson, N., Kaur, K., Emmett, K., Iovannisci, D.M. & Ullman, B. (1984) J.Biol.Chem. **259**, 7637-7643

Cloning and Overexpression of the Hypoxanthine-Guanine Phosphoribosyltransferase Gene from Trypanosoma brucei

by

Thomas E. Allen and Buddy Ullman *

Department of Biochemistry and Molecular Biology

Oregon Health Sciences University

Portland, OR 97201-3098

Telephone (503) 494-8437 - FAX (503) 494-8393

Internet: ullmanb@ohsu.edu

Running Title: *hgprt* from *Trypanosoma brucei*

* To whom correspondence should be addressed

SUMMARY

The hypoxanthine-guanine phosphoribosyltransferase (HGPRT) enzyme of *Trypanosoma brucei* and related parasites provides a rational target for the treatment of African sleeping sickness and several other parasitic diseases. To characterize the *T. brucei* HGPRT enzyme in detail, the *T. brucei hgprt* was isolated by hybridization to a fragment of the *hgprt* that was amplified in the polymerase chain reaction. Nucleotide sequence analysis of a 4.2 kb *Sal*I-*Kpn*I DNA fragment revealed an open reading frame of 630 bp that encoded a protein product of 210 amino acids with a M_r = 23.4 kd. After gap alignment, the *T. brucei* HGPRT exhibited 21-23 % amino acid sequence identity, mostly in three clustered regions, with the HGPRTs from human, *S. mansoni*, and *P. falciparum*, indicating that the trypanosome enzyme was the most divergent of the group. Surprisingly, the *T. brucei* HGPRT was more homologous to the hypoxanthine phosphoribosyltransferase (HPRT) from the prokaryote *V. harveyi* than to the eukaryotic HGPRTs. Northern blot analysis revealed two trypanosome transcripts of 1.4 and 1.9 kb, each expressed to equivalent degrees in insect vector and mammalian forms of the parasite. Southern blot analysis of genomic DNA suggested that the *T. brucei* genome contains more than one *hgprt* copy but that the *hgprt* locus was not arranged in a tandemly repeated array. After appropriate engineering, the *T. brucei hgprt* was inserted into the prokaryotic expression plasmid pBAce and transformed into Sϕ606 *E. coli* that are deficient in both HPRT and xanthine-guanine phosphoribosyltransferase activities. Soluble, enzymatically active recombinant *T. brucei* HGPRT protein was expressed in this system to a level approaching 100 mg per liter of bacterial culture, and ~200 mg of trypanosome HGPRT have now been purified to homogeneity by GTP-agarose affinity

chromatography. The purified recombinant enzyme recognized hypoxanthine, guanine, and allopurinol, but not xanthine or adenine, as substrates and was inhibited by a variety of nucleotide effectors. The availability of a molecular clone encoding the *T. brucei hgprt* and the purification of large quantities of the recombinant HGPRT enzyme to homogeneity provides an experimentally manipulable molecular and biochemical system for the rational design of novel therapeutic agents for the treatment of African sleeping sickness and other diseases of parasitic origin.

INTRODUCTION

Trypanosoma brucei is a protozoan parasite that is the etiologic agent of African sleeping sickness, a devastating and often fatal disease in humans, and nagana, a deadly disease of domestic animals. Transmitted to humans by the bite of an infected tsetse fly, *T. brucei* inhabit and replicate exclusively within extracellular milieus of the host. The conventional therapeutic agents for the treatment of African sleeping sickness are far from ideal since these drugs are ineffective after central nervous system involvement and have numerous undesirable side effects due to their lack of selectivity for the metabolic machinery of the parasite. Furthermore, treatment of African sleeping sickness is often complicated by chemotherapeutic failure, a predicament that is often attributed to drug resistance (1-3).

Selective and rational therapies for parasitic diseases have been generally difficult to devise since many of the major metabolic pathways in these lower eukaryotes are similar, if not identical, to those in their vertebrate host. One metabolic pathway that is conspicuously different between parasites and humans is that for the production of purine nucleotides. Whereas mammalian cells synthesize purine nucleotides *de novo*, all genera of parasitic protozoa that have been examined are auxotrophic for purines (4-12). In order to satisfy their purine requirements, each genus of parasite has evolved a unique complement of purine salvage enzymes for which no identical mammalian counterpart exists. Biochemical and metabolic studies have substantiated the existence of a plethora of purine salvage enzymes in *T. brucei*, including phosphoribosyltransferases (PRTs),[1]

---

[1] The abbreviations used are: PRT, phosphoribosyltransferase; HGPRT, hypoxanthine-guanine phosphoribosyltransferase; HPRT, hypoxanthine phosphoribosyltransferase; XGPRT, xanthine-guanine

nucleoside kinases, and nucleoside cleavage activities (10,13,14). The unique features of the purine salvage system of *T. brucei* and the related parasites *Trypanosoma cruzi* and *Leishmania* cause these pathogenic hemoflagellates to be selectively susceptible to the cytotoxic effects of several pyrazolopyrimidine analogs of hypoxanthine and inosine (15,16). For instance, *T. b. brucei*, *T. b. gambiense*, and *T. b. rhodiensese* bloodstream forms are all sensitive to allopurinol (HPP) (17), a pyrazolopyrimidine isomer of hypoxanthine that is nontoxic to humans and used extensively in the treatment of hyperuricemia and gout (18).

Metabolic activation of HPP in *T. brucei*, as well as in *Leishmania* and *T. cruzi*, is initiated by the enzyme hypoxanthine-guanine phosphoribosyltransferase (HGPRT: IMP:pyrophosphate phosphoribosyltransferase, EC 2.4.2.8). The unique substrate specificity of the HGPRT enzyme of these parasites suggests that the enzyme may serve as a rational target for antiparasitic drugs. To characterize this therapeutically germane enzyme further and to develop molecular reagents for future studies directed toward structure-based drug design, we have cloned and sequenced the *hgprt* from *T. brucei*. Furthermore, the parasite *hgprt* has been inserted into an expression vector, expressed to very high levels in *E. coli*, and large quantities of recombinant HGPRT have been purified to homogeneity.

---

phosphoribosyltransferase; PRPP, phosphoribosylpyrophosphate; HPP, allopurinol; PCR, polymerase chain reaction; RT, reverse transcriptase; ODC, ornithine decarboxylase; DFMO, α-difluoromethylornithine; SDS, sodium dodecyl sulfate.

MATERIALS AND METHODS

Chemicals and Reagents - [α-^{32}P]dCTP (3000 Ci/mmol) and [α-^{35}S]dATP (1320 Ci/mmol) were purchased from New England Nuclear Corp. (Boston, MA). [^{14}C]Hypoxanthine (56 mCi/mmol), [^{14}C]guanine (56 mCi/mmol), and [^{14}C]xanthine (56 mCi/mmol) were bought from Moravek Biochemicals (Brea, CA). [^{14}C]HPP (46 mCi/mmol) was obtained from Research Products International Corp. (Mount Prospect, IL). Synthetic oligonucleotides were prepared by OLIGO'S ETC. (Wilsonville, OR). Thermostable AmpliTaq DNA polymerase was obtained from the Perkin-Elmer Corp. (Norwalk, CT). All restriction and DNA modifying enzymes were acquired from either Bethesda Research Laboratories Life Technologies Inc. (Gaithersburg, MD) or Boehringer Mannheim Biochemicals (Indianapolis, IN). Guanosine 5'-triphosphate covalently attached through its ribose hydroxyl groups to an agarose resin by an adipic acid dihydrazide spacer (catalog number, G 9768), phosphoribosylpyrophosphate (PRPP), and all purine bases were purchased from the Sigma Chemical Co. (St. Louis, MO). All other materials, chemicals, and reagents were of the highest quality commercially available.

Cell Culture - Procyclic forms of *T. b. brucei* strain 1.1a were propagated in SDM-79 medium (19) supplemented with 10 % heat-inactivated fetal calf serum acquired from HyClone Inc. (Logan, UT). The parasites were grown continuously in tightly capped 75 cm^2 tissue culture flasks on a rotary shaker at 28° C.

Nucleic Acid Isolation - Genomic DNA was isolated from 1.0 X 10^{10} *T. brucei* procyclic forms as described (20). Total RNA was prepared by the phenol-chloroform extraction protocol reported by Landfear and Wirth (21). RNA from mammalian forms

of the 1.1a *T. brucei* strain was generously provided by Dr. Marilyn Parsons of the Seattle Biomedical Research Institute (Seattle, WA).

Isolation of a T. brucei hgprt Fragment by the Polymerase Chain Reaction - To generate a homologous DNA probe for the isolation of the *T. brucei hgprt*, a fragment of the gene was amplified from genomic DNA in the polymerase chain reaction (PCR) using degenerate oligonucleotide primers created from completely conserved amino acid sequences on the mammalian (22,23), *Schistosoma mansoni* (24), and *Plasmodium falciparum* (25) HGPRT proteins. A sense primer, 5'-TCTGGATCC-GAG-GA[CT]-AT[CT]-AT[CT]-GA[CT]-AC-3, was constructed with an 9 nucleotide leader encompassing a *Bam*HI site 5' to a degenerate oligonucleotide sequence corresponding to amino acids 134 - 139, EDIIDT, of the human HGPRT. The antisense primer, TCTGAATTC-[AG]TA-[ACG]CC-[ACG]AC-[ACG]AC-[AG]AA-3', was synthesized with an *Eco*RI site preceded by a 3 nucleotide leader 5' to a mixed oligonucleotide generated from residues 187 - 191, FVVGY, of the human HGPRT. Amplification of the specific *hgprt* fragment was accomplished on a Coy Instruments (Ann Arbor, MI) thermocycler using the amplification assay mixture described by Hanson et al. (26). The PCR reaction mixture was overlayed with 50 μl of mineral oil, and the sample manipulated for 30 cycles of denaturation at 94° for 0.5 min, annealing at 40° for 0.5 min, and extension at 72° 1.0 min. The PCR products were separated by gel electrophoresis, digested with *Bam*HI and *Eco*RI, ligated into appropriately digested pBluescript KS^+ vector from Stratagene (San Diego, CA), and transformed into XL-1 Blue *Escherichia coli*. Large scale plasmid preparations were prepared according to the large scale alkaline lysis protocol (27). The 174 bp *hgprt* fragment was sequenced according to the dideoxy chain

termination method (28) with [^{35}S]dATP as the radiolabel and the Sequenase 2.0 sequencing kit from United States Biochemical Corp (Cleveland, OH).

Libraries of T. brucei DNA - The cDNA library in λ-ZAP (Stratagene, La Jolla, Ca.) was constructed from cDNA isolated from the EATRO 164 strain of *T. b. brucei.* The genomic library in EMBL3 was synthesized from genomic DNA prepared from *T. b. brucei* strain IsTat 1.7. Both the cDNA and genomic libraries were generously provided by Drs. Parsons and Peter Myler of the Seattle Biomedical Research Institute.

Isolation of the hgprt from T. brucei Libraries - The *T. brucei hgprt* was isolated from the *T. brucei* genomic library in two steps. First, the amplified *hgprt* PCR product was employed as a probe to screen the *T. brucei* cDNA library, and second, an *hgprt* cDNA that lacked the 5' protein coding portion of the *hgprt* was used to screen the genomic library. The prehybridization, hybridization, and washing conditions employed to isolate the *hgprt* cDNAs and genomic locus were identical to those described previously by this laboratory for the isolation of the *L. donovani* ornithine decarboxylase (ODC) gene (26). Positive bacteriophage from each library were carried through a tertiary plaque purification, and their DNA was isolated by CsCl density gradient centrifugation (29).

Subcloning and Sequencing of T. brucei hgprt - The *hgprt*-containing *Bam*HI inserts from the cDNA library were separated from the phage DNA on 0.8 % agarose gels, and the longest cDNA, a 1.4 kb fragment, was ligated into *Bam*HI-digested pBluescript KS$^+$ and transformed into XL-1 Blue *E. coli.* DNA from the genomic clones was cleaved with a variety of restriction endonucleases, electrophoresed and blotted onto Nytran membranes (Schleicher & Schuell, Keene, NH), and probed with the *hgprt* cDNA under

the same high stringency conditions employed for screening the libraries. A 4.2 kb *Sal*I-*Kpn*I genomic fragment that hybridized to the *hgprt* cDNA was ligated into pBluescript KS^+ and transformed into the XL-1 Blue *E. coli*. Large scale plasmid preparations of pBluescript containing trypanosome DNA and nucleotide sequencing of double stranded DNA were performed as described above for the PCR fragment.

Computer Assisted DNA and Amino Acid Sequence Analysis - Analyses of nucleotide and predicted amino acid sequences were performed on an IBM compatible AT computer using the Sequence Analysis Program from International Biotechnologies, Inc. (New Haven, CT). Protein sequences were aligned using the CLUSTAL V multiple sequence alignment program, a modification of the CLUSTAL package described by Higgins and Sharp (30). Amino acid similarity scores between nonidentical amino acid pairs were obtained from the log-odds amino acid similarity matrix of Dayhoff (31).

Southern and Northern Hybridizations - Southern blots of genomic DNA that had been digested with the appropriate restriction enzymes using the conditions recommended by the supplier were performed as described (29). A Northern blot of RNA isolated from bloodstream and procyclic forms of *T. brucei* strain 1.1a was hybridized to the *hgprt* under the same high stringency conditions described for probing the Southern blots. The blotted RNA was generously provided by Dr. Parsons.

Mapping the 5' Terminus of the Mature hgprt mRNA - The 5' terminus of the mature *T. brucei hgprt* transcript was defined by a modification of the PCR amplification protocol reported by Hanson et al. (26). cDNA synthesis from total RNA was achieved with random hexamer primers using the reverse transciptase (RT) protocol described by Kawasaki (32). The RT mixture was then adjusted to 10 mM Tris-HCl, pH 8.3, 50 mM

KCl, and 1.5 mM $MgCl_2$, and 10 pmol of each oligonucleotide primer and 5 units of *Taq* DNA polymerase were added. The sense primer, 5'-CTCGGGATCCCAACGCTATATAAGTATCAG TTTCTGTACTTTATTG-3', contained an 11 nucleotide leader containing a *Bam*HI restriction site followed by 31 nucleotides matching positions 5 - 35 of the *L. donovani* mini-exon (33). The mini-exon is an RNA that is *trans*-spliced onto the 5' end of all mRNAs from *Trypanosoma* and *Leishmania* species (34,35]). Of the 31 nucleotides of the sense primer, the *T. brucei* and *L. donovani* mini-exon genes differ in only 4 positions, thus, ensuring efficient priming of the cDNA in the PCR. The antisense primer, 5'-CTCGAATTCGCTGCCCTTCAATACAGA-3', was constructed with a 9 nucleotide leader that contained an *Eco*RI recognition site 5' to nucleotides 151 - 168 of the predicted protein coding portion of the *T. brucei hgprt*. The PCR cocktail was overlayed with mineral oil, and the sample incubated at 94, 50, and 72 °C, respectively, for 35 cycles in the thermocycler. A 250 bp PCR product that hybridized to the *T. brucei hgprt* was digested with *Bam*HI and *Eco*RI, subcloned into pBluescript KS^+, and sequenced as described above.

Overexpression of T. brucei hgprt in E. coli - The *T. brucei hgprt* was ligated into the pBAce expression vector constructed by Craig et al. (36). The pBAce vector utilizes the bacterial alkaline phosphatase promoter to direct the high level cytoplasmic expression of foreign proteins in *E. coli*. An *Nde*I restriction site was introduced at the initiation methionine codon and a *Sal*I cleavage site was inserted 3' to the termination signal by PCR mutagenesis in two separate amplification reactions, necessitated by the presence of an internal *Nde*I site within the *T. brucei hgprt*. The 5' end of the *hgprt* was amplified

using 5'-TCTCATATGGAACCAGCTTGCAAATA-3' as the mutagenic primer and 5'-GTAGGAATTCCACACGTGTGG-3'as the antisense primer containing an internal *Eco*RI cleavage site. The 3' end of the gene was amplified using 5'-TGTGGAATTCCTGCGGGCCTC-3' containing the same internal *Eco*RI site as the sense primer and the mutagenic 5'-TCTGTCGACACAAGGCACAAGAGCTACCTG-3' as the antisense primer. The two PCR fragments were digested with the appropriate restriction enzymes and ligated simultaneously into pBAce that had been cleaved with *Nde*I and *Sal*I. The ligation mixture was transformed into the *E. coli* K-12 strain Sϕ606 (Δ*gpt-pro-lac, thi, hpt*) that was initially described (37) and generously provided to this laboratory by Dr. Per Nygaard of the University Institute of Biological Chemistry B (Copenhagen, Denmark) through Dr. Sidney Craig of the University of California (San Francisco, CA). This strain contains a deletion in the gene encoding the bacterial xanthine-guanine phosphoribosyltransferase (XGPRT) activity and is also deficient in hypoxanthine phosphoribosyltransferase (HPRT) activity. DNA was prepared from the transformants, and the 630 bp *Nde*I-*Sal*I fragment containing the entire *T. brucei hgprt* was sequenced.

500 ml cultures of Sϕ606 bacteria transformed with the recombinant expression plasmid were incubated in a low phosphate medium modified from that previously described (36) by reducing the concentration of "vitamin free" Casamino acids (Difco Laboratories, Detroit, MI) to 0.1 % and deleting the equimolar mixture of NaH_2PO_4 and Na_2HPO_4. The adjustments were crucial for maximizing expression of recombinant protein, as the "vitamin free" Casamino acids contained 100 umol/g of inorganic phosphate. The final concentration of phosphate in the modified induction medium was,

therefore, ~100 μM.

Inorganic Phosphate Assay - Inorganic phosphate levels in the *hgprt* expression medium were determined as described (38).

HGPRT Assay - HGPRT activity was assayed as previously reported (39). The rates of [^{14}C]HPP phosphoribosylation were measured by replacing hypoxanthine or guanine in the HGPRT cocktail with HPP.

HGPRT Purification - Recombinant *T. brucei* HGPRT was purified by a modification of the method described by Allen et al. (39) for the purification of native HGPRT from *L. donovani*. 500 ml cultures of Sϕ606 *E. coli* were grown for 12 hr at 37° to an OD_{600} = 0.9 - 1.2. The cells were washed several times with phosphate buffered saline (137 mM NaCl, 2.7 mM KCl, 8.1 mM Na_2HPO_4, and 1.0 mM KH_2PO_4) and lysed by French press in 20 mM Tris, pH 7.4, 5 mM $MgCl_2$, and 1mM dithiothreitol (TMD) buffer. After centrifugation at 40,000 X g for 30 min at 4°, the cell supernatant was chromatographed over a 10 ml GTP-agarose affinity column. The column was then washed with 10 column volumes of TMD and the protein eluted in TMD containing 1.0 mM PRPP.

Polyacrylamide Gel Electrophoresis - Sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis on 12 % acrylamide slab gels was performed as described (29).

RESULTS

Isolation and Sequencing of the hgprt from T. brucei - Preliminary attempts to isolate the *hgprt* from *T. brucei*, as well as from other parasites of the Trypanosomatidae family, by cross hybridization under low stringency conditions to either the mouse (23), *S. mansoni* (24), or *P. falciparum* (25) *hgprt* cDNAs were unsuccessful. Consequently, the PCR was exploited to amplify fragments of the *T. brucei hpgrt* from genomic DNA using primers to conserved amino acid sequences among other eukaryotic HGPRT proteins (22-25). A 174 bp fragment that encoded a fragment of the *T. brucei hgprt* was amplified and used as a probe for the isolation of *hgprt* cDNAs from a cDNA library. Nucleotide sequence analysis of the longest cDNA, a 1.4 kb fragment, confirmed that it lacked the 5' end of the protein coding segment of the *T. brucei hgprt*. Therefore, the 1.4 kb cDNA was employed in turn to isolate the *hgprt* from a genomic library. A 4.2 kb *Sal*I-*Kpn*I fragment was subcloned into pBluescript, and the protein coding segment of *hgprt* was sequenced in both directions. The nucleotide and predicted amino acid sequence are indicated in Fig. 1. This sequence contains a single open reading frame of 630 nucleotides that encodes a protein of 210 amino acids with a M_r = 23.4 kd and a calculated isoelectric point of 6.38. The open reading frame is preceded by an in-frame termination codon beginning at position -30 from the adenylate residue of the predicted initiating methionine codon (underlined in Fig. 1). Moreover, PCR of reverse transcribed RNA indicated that the first nucleotide of the mature *hgprt* transcript is located 59 nucleotides 5' to the start codon (marked by an asterisk in Fig. 1).

The predicted amino acid sequence of the *T. brucei* HGPRT was aligned with the primary structures of the human (22), *S. mansoni* (24), *P. falciparum* (25), and *Vibrio*

harveyi (40) counterparts (Fig. 2). The *T. brucei* HGPRT is between 8 and 21 amino acids shorter than its eukaryotic equivalents but 34 amino acids longer than the prokaryotic HPRT. Examination of all the PRT sequences after alignment also disclosed a variety of amino acid deletions and insertions throughout their primary structures. When each of the PRTs was examined for amino acid identities in a pairwise fashion to the other enzymes, the *T. brucei* HGPRT appeared to be the most divergent and exhibited greater homology toward the *V. harveyi* HPRT than to the eukaryotic HGPRTs (Table I). The degree of similarity among the eukaryotic HGPRT sequences conforms to the observed phylogenetic relationships among the organisms inferred from analysis of their *rRNA* sequences (41).

Among all the aligned PRT sequences, there are few regions of extensive homology. Unsurprisingly, the stretches of amino acids corresponding to the putative substrate binding domains (24,42,43) are among the most conserved regions of the PRT proteins. In a 16 amino acid region encompassing the PRPP binding domain (24,42,43), corresponding to residues 107 to 122 of the *T. brucei* HGPRT, there are 9 completely conserved amino acids and 6 other positions which can be accounted for by conservative substitutions. A single Lys to Leu substitution found at position 120 of the *T. brucei* HGPRT is the only nonconservative substitution within this stretch of amino acids. Similarly, the putative purine nucleobase binding domain (marked in Fig. 2), coinciding with residues 51 to 63 of the *T. brucei* HGPRT (24,42), is also fairly well conserved among the various PRTs. Within this purine binding motif, there are 3 amino acid identities and 5 conservative changes among the mammalian and parasite HGPRTs and the prokaryotic HPRT. Finally, it is worth noting that a region located proximal to the

COOH termini, a stretch of amino acids for which no function has been ascribed, is also highly conserved among the aligned PRT sequences. Between residues 162 and 181 of the *T. brucei* HGPRT, there are 7 amino acid identities and 9 similarities within this region. Finally, it is worth noting that the amino acid regions outside the 3 conserved areas are highly divergent among the PRT proteins with only an occasional single amino acid identical among the 5 aligned PRT sequences. Among just the eukaryotic HGPRT sequences, the longest sequence of amino acid identity outside the 3 conserved regions is 2 amino acids, a Ser-Tyr dipeptide corresponding to residues 83 and 84 of the *T. brucei* HGPRT.

Size and Quantitation of T. brucei hgprt Transcripts - A Northern blot of total RNA isolated from trypomastigote, stumpy induced, and procyclic forms of the 1.1 *T. brucei* strain was probed with the 774 bp *Nde*I-*Sal*I fragment that contains the protein coding portion of the *hgprt*. Two transcripts of 1.4 and 1.9 kb, each large enough to encode the entire trypanosome HGPRT, were recognized by this probe in all 3 developmental stages of the parasite (Fig. 3). The hybridization signals to both bands were equivalent in the 3 life forms, and the level of *hgprt* expression did not appear to be developmentally regulated.

Structure of the hgprt Locus in T. brucei - As a number of genes in *T. brucei* are arranged in tandemly repeated arrays (44,45), the *hgprt* locus in the 1.1a strain of *T. brucei* was analyzed for repeated sequences by restriction enzyme analysis and Southern blotting (Fig. 4). *T. brucei* DNA was digested with a variety of restriction enzymes, blotted onto Nytran filters, and probed with the 774 bp *Nde*I-*Sal*I fragment. Restriction enzymes *Nde*I, *Eco*RI, and *Sac*II, each of which cleaves the *hgprt* once, excised 3 or more

fragments of dissimilar sizes that hybridized to the *hgprt* probe (Fig. 4). If *hgprt* was organized as a tandem repeat within the *T. brucei* genome, this strategy would yield a prominent band of identical size in all lanes of a Southern blot in which the genomic DNA had been digested with a single cutter. Moreover, digestion of *T. brucei* DNA with either *Ava*I, *Hind*III, or *Kpn*I, which do not cleave within *hgprt*, generates two DNA fragments that hybridize to the *hgprt*. These data suggest that *hgprt* is present in at least two copies within the *T. brucei* genome, but that the intergenic region between the two loci is very large.

Overexpression of hgprt in E. coli - After appropriate PCR mutagenesis, the *T. brucei hgprt* was ligated into the *E. coli* expression plasmid pBAce and transformed into Sϕ606 *E. coli* that are genetically deficient in both the HPRT and XGPRT activities. Nucleotide sequencing demonstrated that the primary structures of the recombinant and native HGPRT enzymes were identical. After inoculation into low phosphate medium, the transformed *E. coli* were grown for ~12 hr to an OD_{600} = 0.9 - 1.2. Further growth of cells in this induction medium to stationary phase densities severely diminished *hgprt* expression. As indicated in Fig. 5, Sϕ606 *E. coli* transformed with the *hgprt* construct expressed very high levels of HGPRT activity, whereas the same cells transformed with a control pBAce plasmid lacking *hgprt* expressed no HGPRT activity. Proteins from the Sϕ606 *E. coli* were then fractionated by SDS polyacrylamide gel electrophoresis and stained with Coomassie Blue. As demonstrated in Fig. 6, the recombinant *T. brucei* HGPRT is the predominant protein in *E. coli* lysates transformed with the *hgprt*-containing plasmid and is not detected in cells transformed with the control plasmid. The recombinant HGPRT migrates slightly more slowly than that expected from its

predicted M_r.

Purification and Substrate Specificity of T. brucei HGPRT - Recombinant *T. brucei* HGPRT protein was purified by GTP-agarose affinity chromatography followed by elution with 1.0 mM PRPP (Fig. 7). No HGPRT activity was present either in the void volume or in washes prior to the addition of the PRPP. As demonstrated in Fig. 6, passage over the affinity column resulted in a one step purification of the protein to homogeneity. ~100 mg of pure *T. brucei* HGPRT could be obtained by this strategy per liter of bacterial culture. The substrate specificity of the pure recombinant trypanosome HGPRT was also evaluated. Clearly, *T. brucei* HGPRT prefers the naturally occurring nucleobases, hypoxanthine and guanine, as substrates but is also capable of phosphoribosylating HPP. K_m values of 2.3 ± 0.9, 4.8 ± 0.6, and 243 ± 35 μM and k_{cat} values of 32 ± 4.7, 53 ± 2.9, and 4.4 ± 0.6 μmol/min/mg protein were calculated for the recombinant enzyme by Lineweaver-Burk analysis for hypoxanthine, guanine, and HPP, respectively (Fig. 8). Neither xanthine nor adenine were phosphoribosylated by the purified *T. brucei* HGPRT. Inhibition studies also revealed that a variety of nucleotide effectors were inhibitors of the *T. brucei* HGPRT. As demonstrated in Table II, GTP, IMP, GMP, ATP, and AMP at a concentration of 1.0 mM, all reduced, in order of diminishing effectiveness, the capacity of the enzyme to phosphoribosylate 20 μM [^{14}C]hypoxanthine. Moreover, the conversion of [^{14}C]hypoxanthine to the nucleotide level was obliterated by excess hypoxanthine or guanine but unaffected by adenine or xanthine.

DISCUSSION

A full length *hgprt* gene has been isolated from a genomic library of *T. brucei* DNA. Unlike previously isolated parasite *hgprt* cDNAs that were isolated by cross hybridization to the murine *hgprt* cDNA, the nucleotide sequence disparity of the *T. brucei hgprt* gene precluded its recognition by the heterologous *hgprt* probes. Thus, the PCR was exploited to amplify a specific *hgprt* fragment from *T. brucei* using oligonucleotide primers generated to the PRPP binding and COOH terminal sequences that are highly conserved among the mammalian (22,23), *S. mansoni* (24), and *P. falciparum* (25) HGPRTs. Nucleotide sequence analysis revealed that the primary structure of the *T. brucei* HGPRT protein exhibited greater similarity to the *V. harveyi* HPRT than to its counterpart from higher eukaryotes. Moreover, the size and location of predicted gaps in the sequence alignment of the *T. brucei* HGPRT with the other PRTs also showed that the trypanosome enzyme displayed greater similarity to the prokaryotic HPRT than to the eukaryotic HGPRTs. This homology cannot be explained by substrate specificity variations, since the prokaryotic enzyme phosphoribosylates hypoxanthine exclusively, while xanthine and guanine are recognized by a biochemically and genetically distinct protein (37,40,46). The lack of homology between the *T. brucei* and eukaryotic HGPRTs is somewhat surprising in view of the extensive homology observed between the primary structures of the ODC (47) and IMP dehydrogenase (48) proteins from *T. brucei* and mouse. The *T. brucei* and murine ODC enzymes exhibit a 69 % amino acid identity over a core region of 376 amino acids with no gaps in the alignment, whereas the *T. brucei* and human IMP dehydrogenase proteins are identical in 53 % of their residues with only a few small gaps in the computed alignment. The

amino acid sequence divergence among the HGPRT proteins from phylogenetically diverse organisms suggests that conservation of primary structure is not critical for enzyme function. Despite the extensive divergence of primary structures, analyses of the secondary structures of mammalian and parasite HGPRTs using the algorithm of Chou and Fasman (49), predict several common features, including a β-α-β structure characteristic of a mononucleotide fold involved in PRPP binding to PRTs (24,42, and data not shown).

The sequence homology that is observed among the various PRT proteins occurs in several distinct regions separated and flanked by much longer regions of limited homology. These alignments, along with others that compare sequences among the PRT superfamily (42,43), permit evaluation of the structural features that are required for catalytic competence of these enzymes. The longest stretch of amino acid sequence similarity among the eukaryotic HGPRTs occurs close to the COOH terminus in which 10 identical and 8 similar amino acids are found within a 20 amino acid stretch that corresponds to residues 162 and 181 of the *T. brucei* HGPRT. Although no known function has been imputed to this region of the protein, several naturally occurring missense mutations in this regions confer HGPRTase deficiency in humans and elevate apparent K_m values for both substrates dramatically (50), indicating the importance of these sequences in catalytic competence and substrate binding.

The putative PRPP binding domain of HGPRT proteins was inferred from alignments of biochemically distinct members of the PRT superfamily, including HPRTs (22,23), and PRTs for adenine (42,43), xanthine-guanine (45), orotate (51), and glutamine (52). This region is also well conserved in the *T. brucei* HGPRT. Over a 16

amino acid stretch encompassing the PRPP binding motif coinciding with residues 107 to 123 of the *T. brucei* HGPRT, there are 9 identical and 6 similar amino acids among the parasite and mammalian enzymes. Naturally occurring mutations within the PRPP binding site of the human HGPRT also diminish catalytic activity dramatically or completely (50).

The amino acid sequences that control the substrate specificities of PRT enzymes for purine bases have been imputed to a region encompassing amino acids 51 - 63 of the *T. brucei* HGPRT. Within this highly conserved purine binding domain lies a central Arg residue and a Thr-Ala dipeptide in the *P. falciparum* enzyme (25) that distinguishes the malarial purine binding site from those of the human (22), *S. mansoni* (24), *T. brucei*, and *V. harveyi* (40). A conserved aromatic amino acid and an Ala-Asp are found at these positions in the other HPRTs. These striking differences may account for the unique ability of the malarial HGPRT to recognize xanthine as a substrate (53). The other eukaryotic HGPRTs, as well as the prokaryotic HPRT, do not phosphoribosylate xanthine. The Arg residue, as well as the immediately adjacent Ser and Gly residues in the *P. falciparum* HGPRT purine binding site, are also found within the purine base binding motif of the bacterial XGPRT (46).

The Ser and Tyr residues that are located at positions 83 and 84 of the trypanosome protein are completely conserved among the other eukaryotic HGPRTs and are also predicted on the basis of secondary structure considerations to play a role in purine binding to the enzyme (24,42). A Thr replaces the Tyr in the *V. harveyi* HPRT (40), which recognizes only hypoxanthine as a naturally occurring nucleobase substrate, suggesting that the Tyr may be critical for guanine binding to HGPRT enzymes.

Although amino acid distinctions within the longer purine binding domain exist, no single difference in this region can easily account for the selective capacity of the eukaryotic proteins to phosphoribosylate guanine. Finally, among all the PRTs, the *T. brucei* enzyme has the unique ability to phosphoribosylate the pyrazolopyrimidine HPP (15-17). The single amino acid within the 13 residue purine binding domain of the trypanosome HGPRT that is most dissimilar to all its counterparts and, therefore, may contribute to its ability to recognize pyrazolopyrimdines is the Ser at position 51.

The *hgprt* gene is not differentially expressed in *T. brucei*. Northern blot analysis revealed that two transcripts of 1.4 and 1.9 kb, both of which are sufficient to encode the trypanosome protein, are expressed to equivalent extents in the insect vector and mammalian stages of *T. b. brucei*. The lack of differential expression of HGPRT is unsurprising in view of the facts that both the procyclic and bloodstream forms of *T. brucei* are auxotrophic for purines (10,13) and that hypoxanthine is a predominant environmental source of purines for the parasite (54). Whether the two transcripts are derived from the same or different genetic loci is not known in view of the fact that these parasites appear to contain more than one *hgprt* copy. As Southern blot and sequence analysis of cDNA generated and amplified from the 5' portion of *hgprt* mRNA revealed only a single fragment, the data suggest that the *hgprt* transcripts are heterogeneous in their 3' termini.

The ability to overexpress and purify the *T. brucei* HGPRT in an *E. coli* background genetically deficient in both HPRT and XGPRT has established that the enzyme is specific for hypoxanthine and guanine and does not recognize either xanthine or adenine. The existence of separate PRTs for xanthine and adenine is supported by

experimental evidence demonstrating the differential incorporation of radiolabeled purine bases into nucleotide pools in both procyclic and bloodstream forms of *T. brucei* species and by the fact that *T. brucei* lack adenine deaminase activity (10,13), an enzyme that is expressed to high levels in promastigotes of *Leishmania* (55). Biochemically distinct PRTs for adenine, hypoxanthine-guanine, and xanthine have also been chromatographically distinguished in *L. donovani* promastigotes (56). Moreover, expression of the recombinant *T. brucei* HGPRT in the Sϕ606 *E. coli* establishes that the enzyme can recognize HPP as a substrate. The procyclic forms of *T. b. brucei*, *T. b. gambiense*, and *T. b. rhodesiense* can further convert HPP to aminopyrazolpyrimidine nucleotides and incorporates the triphosphate of the latter into RNA (17). This metabolic sequence does not exist in humans or mammals, and the procyclic forms of the parasite are selectively sensitive to HPP. HPP exhibits no toxicity toward humans and is widely used as an hypouricemic agent in the treatment of hyperuricemia and gout (18). Although, the curative effects of HPP against African trypanosomiasis in humans have not been established, the pyrazolopyrimidine exhibits demonstrable therapeutic efficacy in preliminary clinical trials against both cutaneous leishmaniasis (57) and Chagas disease (58). Moreover, as hypoxanthine is the salvageable purine present in the highest concentrations in human blood and cerebrospinal fluid (54), the milieus inhabited by the parasite, HGPRT is likely to provide a critical nutritional function in *T. brucei*. The selective toxicity of HPP toward *T. brucei* and the potentially central role of HGPRT in purine salvage imply that the parasite HGPRT may provide a biochemically rational target for the therapeutic manipulation of African sleeping sickness.

Although virtually all antiparasitic regimens are empirical in design, a precedent

for rationally based treatments of African trypanosomiasis exists. α-Difluoromethylornithine (DFMO), an irreversible inhibitor of ODC, the rate limiting step in polyamine biosynthesis (59), has shown considerable promise in the treatment of some forms of African trypanosomiasis (60). The therapeutic efficacy of DFMO is based on the rapid rate of turnover of the mammalian enzyme in contrast to the *T. brucei* ODC (47,61). The instability of the mammalian ODC is conferred by a COOH terminus rich in PEST sequences, a sequence missing in the *T. brucei* ODC. Although, DFMO treatment does not manifest the toxic side effects of the conventional antitrypanosomal reagents and is effective after central nervous system involvement, its adminsitration is prolonged and expensive, and the drug is not particularly effective toward *T. b. rhodesiense* (60). Moreover, resistance to DFMO has been demonstrated *in vitro* (62,63). Thus, the search for newer and improved therapies for the treatment of African trypanosomiasis are imperative. The isolation of the molecular clone encoding the *T. brucei* HGPRT and the functional overexpression and purification of large quantities of pure recombinant HGPRT enzyme from *E. coli* provide an avenue for a structure-based strategy for the design and discovery of novel HGPRT substrate analogs or HGPRTase inhibitors that can serve as effective agents in the treatment of African trypanosomiasis or of any other parasitic disease.

FIG. 1. **Nucleotide and predicted amino acid sequence of the protein coding segment of the *hgprt* from *T. brucei*.** 1,074 bp of the 4.2 kb *Sal*I-*Kpn*I fragment were sequenced in both directions and the amino acid sequence inferred from its 630 bp open reading frame. Nucleotides within the protein coding region of the *hgprt* are enumerated positively from the first adenylate nucleotide of the predicted initiating Met, and amino acids within the HGPRT protein are numbered positively from the first Met. Nucleotides within the 5' untranslated region are enumerated negatively from the same adenylate moiety. An in-frame termination codon 5' to the predicted start site is underlined by three asterisks. The nucleotide that corresponds to the first nucleotide of the mature *hgprt* transcript to which the mini-exon is *trans*-spliced is underlined with an asterisk.

Nucleotide and Predicted Amino Acid Sequence of *T. brucei hgprt*

```
-252  TGGTTACGTTCCTCTCCCTCCACCCTCCTCTCATTTTTTTTTTCCCTCCGTCTCTGTTTAATCTGGCAATATTGACGTG -174
-173  ACGTGAATGCGGCCGTCCCCACAGGTATCATGCAGTGCCTGGTAGTCCACTACCTTGAACCTAACCCAATTATCGTGCA - 95
 -94  ATGTTAGCATCTTCTTTTCCTTCTGCGGTAACAAGGTTTCACAAGAGATATCATAAAGCTCGCGTGAGATATTCTGGAA - 16
 -15  AGAAAAGCAGGAAGA                    *

   1  MET GLU PRO ALA CYS LYS TYR ASP PHE ALA THR SER VAL LEU PHE THR GLU ALA GLU LEU   20
   1  ATG GAA CCA GCT TGC AAA TAC GAC TTC GCA ACG AGT GTC CTC TTT ACA GAG GCA GAA CTA   60

  21  HIS THR ARG MET ARG GLY VAL ALA GLN ARG ILE ALA ASP ASP TYR SER ASN CYS ASN LEU   40
  61  CAC ACT CGC ATG CGC GGT GTG GCG CAG CGT ATT GCC GAT GAC TAC AGC AAC TGC AAT TTG  120

  41  LYS PRO LEU GLU ASN PRO LEU VAL ILE VAL SER VAL LEU LYS GLY SER PHE VAL PHE THR   60
 121  AAG CCA CTT GAA AAT CCT CTG GTA ATT GTG TCT GTA TTG AAG GGC AGC TTC GTC TTC ACT  180

  61  ALA ASP MET VAL ARG ILE LEU GLY ASP PHE GLY VAL PRO THR ARG VAL GLU PHE LEU ARG   80
 181  GCT GAC ATG GTT CGC ATT CTT GGT GAT TTT GGC GTC CCC ACA CGT GTG GAA TTC CTG CGG  240

  81  ALA SER SER TYR GLY HIS ASP THR LYS SER CYS GLY ARG VAL ASP VAL LYS ALA ASP GLY  100
 241  GCC TCG TCA TAT GGT CAC GAT ACT AAA AGT TGT GGT CGA GTT GAC GTG AAG GCT GAC GGT  300

 101  LEU CYS ASP ILE ARG GLY LYS HIS VAL LEU VAL LEU GLU ASP ILE LEU ASP THR ALA LEU  120
 301  CTT TGT GAC ATC CGC GGA AAA CAT GTC CTT GTT TTG GAG GAT ATA CTT GAT ACT GCG CTG  360

 121  THR LEU ARG GLU VAL VAL ASP SER LEU LYS LYS SER GLU PRO ALA SER ILE LYS THR LEU  140
 361  ACG TTG AGG GAA GTG GTG GAC AGC TTG AAA AAG AGC GAA CCC GCG AGC ATT AAA ACC CTC  420

 141  VAL ALA ILE ASP LYS PRO GLY GLY ARG LYS ILE PRO PHE THR ALA GLU TYR VAL VAL ALA  160
 421  GTG GCT ATC GAC AAA CCC GGT GGG CGT AAA ATA CCT TTC ACT GCG GAA TAC GTT GTG GCG  480

 161  ASP VAL PRO ASN VAL PHE VAL VAL GLY TYR GLY LEU ASP TYR ASP GLN SER TYR ARG GLU  180
 481  GAT GTT CCC AAT GTG TTC GTG GTT GGC TAC GGG TTG GAT TAC GAC CAA TCA TAC CGT GAG  540

 181  VAL ARG ASP VAL VAL ILE LEU LYS PRO SER VAL TYR GLU THR TRP GLY LYS GLU LEU GLU  200
 541  GTG CGT GAT GTT GTA ATC CTA AAA CCG AGC GTG TAT GAA ACA TGG GGA AAG GAA CTT GAG  600

 201  ARG ARG LYS ALA ALA GLY GLU ALA LYS ARG                                         210
 601  CGG AGG AAG GCC GCT GGA GAA GCC AAG CGG                                         630

 680  TAAATGCGGTCGCTGATGTGCTGATGGTGGGAGGTAACGGGGTTCTGTGGTGCGATTACCCGTACGTAAATTTGTTCTT  758
 759  TCCCATGCTTCCTCCCTATTTAATGTCACCTGTGTAAGCAGGTAGCTCTTGTGCCTTGTGCTTGTTTTGTTATTTCCCC  837
 838  TACGCTTGTTTCACAATGATCATACTTGGGGAAG                                              873
```

FIG. 2. **Alignment of the amino acid sequences of the *T. brucei*, human, *S. mansoni*, and *P. falciparum* HGPRTs and the *V. harveyi* HPRT.** The predicted primary structure of the *T. brucei* HGPRT protein was aligned with those deduced from the nucleotide sequences of the human (22), *S. mansoni* (24), and *P. falciparum* (25) *hgprt* cDNAs and the *V. harveyi hprt* (40) according to the CLUSTAL V program, a modification of that described by Higgins and Sharp (30). Amino acids identical among the five aligned PRT sequences are indicated by asterisks below the aligned sequences, whereas nonidentical amino acids with similarity scores >10, as calculated by the log-odds amino acid similarity matix of Dayhoff (31), are denoted with a dot below the amino acid sequences. Amino acid positions of the five proteins are designated numerically on the right. The amino acid sequences to which the PCR primers used in the amplification of the *T. brucei hgprt* fragment from genomic DNA were derived and the putative purine binding site of the PRTs (24,42) are designated.

Alignment of *T. brucei* HGPRT with Eukaryotic HGPRTs and Prokaryotic HPRT

```
Human                1  MAT---RSPG-----VVISDDEPGYDLDLFCIPNHYAEDLERVFIPHGLI   42
S. mansoni           1  MSSNMIKADC-----VVIEDSFRGFPTEYFCTSPRYDECLDYVLIPNGMI   45
P. falciparum        1  MPIPNNPGAGENAFDPVFVKDDDGYDLDSFMIPAHYKKYLTKVLVPNGVI   50
T. brucei            1  M-----EPACKYDFAT------------------------SVLFTEAEL   20
V. harveyi           1  M-----KHT------V------------------------EVMISEQEV   14
                        *                                       .   .
                                                    Purine Site
Human               43  MDRTERLARDVMKEMGGHH-------IVALCVLKGGYKFFADLLDYIKAL   85
S. mansoni          46  KDRLEKMSMDIVDYYEACN----ATSITLMCVLKGGFKFLADLVDGLERT   91
P. falciparum       51  KNRIEKLAYDIKKVY---N----NEEFHILCLLKGSRGFFTALLKHLSRI   93
T. brucei           21  HTRMRGVAQRIADDYSNCNLKPLENPLVIVSVLKGSFVFTADMVRILG--   68
V. harveyi          15  QERIRELGKQITERYQGSE------DLVMVGLLRGSFVFMADLARAI---   55
                          *   ..  .                .  . .*.*.  * .  .

Human               86  NRNSDRSIPMTV---DFIRLKSYCNDQSTGDIKVIGGDDLSTLTGKNVLI  132
S. mansoni          92  VRARGIVLPMSV---EFVRVKSYVNDVSIHEPILTGLGDPSEYKDKNVLV  138
P. falciparum       94  HNYSAVEMSKPLFGEHYVRVKSYCNDQSTGTLEIVS-EDLSCLKGKHVLI  142
T. brucei           69  ----DFGVPTRV---EFLRASSYGHDTKSCGRVDVKADGLCDIRGKHVLV  111
V. harveyi          56  ------ELTHQV---DFMTASSTGNTMESSRDVRILKDLDDDIKGKDVLI   96
                            .   .  ...  *  .                  .*.**.
                        Prim I
Human              133  VEDIIDTGKTMQTLLSLVRQYNPKMVKVASLLVKRTPRSVGYKPDFVGFE  182
S. mansoni         139  VEDIIDTGKTITKLISHLDSLSTKSVKVASLLVKRTSPRNDYRPDFVGFE  188
P. falciparum      143  VEDIIDTGKTLVKFCEYLKKFEIKTVAIACLFIKRTPLWNGFKADFVGFS  192
T. brucei          112  LEDILDTALTLREVVDSLKKSEPASIKTLVAIDKPGGRKIPFTAEYVVAD  161
V. harveyi          97  VEDIIDTGNTLNKIREILSLREPKSIAICTLLDKPSRREVEVPVDYVGFA  146
                        .***.**. *.              .     . *         ..*
                            Prim II
Human              183  IPDKFVVGYALDYNEYFRDLNHVCVI---------SETGKAKYKA-----  218
S. mansoni         189  VPNRFVVGYALDYNDNFRDLHHICVI---------NEVGQKKFSVPCTSK  229
P. falciparum      193  IPDHFVVGYSLDYNEIFRDLDHCCLV---------NDEGKKKYK----AT  229
T. brucei          162  VPNVFVVGYGLDYDQSYREVRDVVILKPSVYETWGKELERRK-------A  204
V. harveyi         147  IPDEFVVGVGIDYAQKYRDLPFIGKVVPQE--------------------  176
                        .*. **.* ..** . .*..     .

Human                   ------------                                    218
S. mansoni         230  P----------V                                    231
P. falciparum      230  S----------L                                    231
T. brucei          205  AGEAKR------                                    210
V. harveyi         177  ------------                                    176
```

FIG. 3. **Northern blot analysis of *hgprt* transcripts.** Nytran filters to which RNA from bloodstream, stumpy induced, and procyclic forms of *T. brucei* strain 1.1a RNA was blotted and with the 774 bp *T. brucei Nde*I-*Sal*I *hgprt* fragment from the pBAce expression vector under conditions of high stringency. This digestion released two fragments, a 248 bp *Nde*I-*Nde*I and a 526 bp *Nde*I-*Sal*I fragment, that together encompassed the entire *T. brucei hgprt* coding region. The RNA standards are from the 1.4 - 9.5 kb RNA ladder supplied by GIBCO BRL (Gaithersburg, MD). The sizes of the *hgprt* transcripts are indicated.

FIG. 4. **Structure of the *hgprt* locus.** 10 μg of genomic DNA from *T. brucei* strain 1.1a were incubated overnight with the indicated restriction enzymes, fractionated on a 0.8 % agarose gel, transferred to a Nytran filter, and probed with the 248 bp *Nde*I-*Nde*I and the 526 bp *Nde*I-*Sal*I *hgprt* fragments as described in the legend to FIG. 3. The protein coding region of the *T. brucei hgprt* contains a single restriction site each for *Nde*I, *Eco*RI, and *Sac*II, while *Bam*HI, *Ava*I, *Hind*III, and *Kpn*I do not cleave within the *hgprt* coding region. DNA markers are products of λ phage DNA digested with *Ava*I and were obtained from United States Biochemical (Cleveland, OH).

FIG. 5. **HGPRT overexpression in *E. coli*.** The *T. brucei hgprt* ligated into the prokaryotic expression vector pBAce was transformed into *E. coli* K-12 strain Sϕ606 (Δ*gpt-pro-lac, thi, hpt*) and expressed in the low phosphate medium described in Materials and Methods. A pBAce plasmid lacking *hgprt* was transformed into Sϕ606 cells as a negative control. After a 12 hr of growth, the transformed *E. coli* were harvested, and soluble recombinant HGPRT enzyme activity was measured.

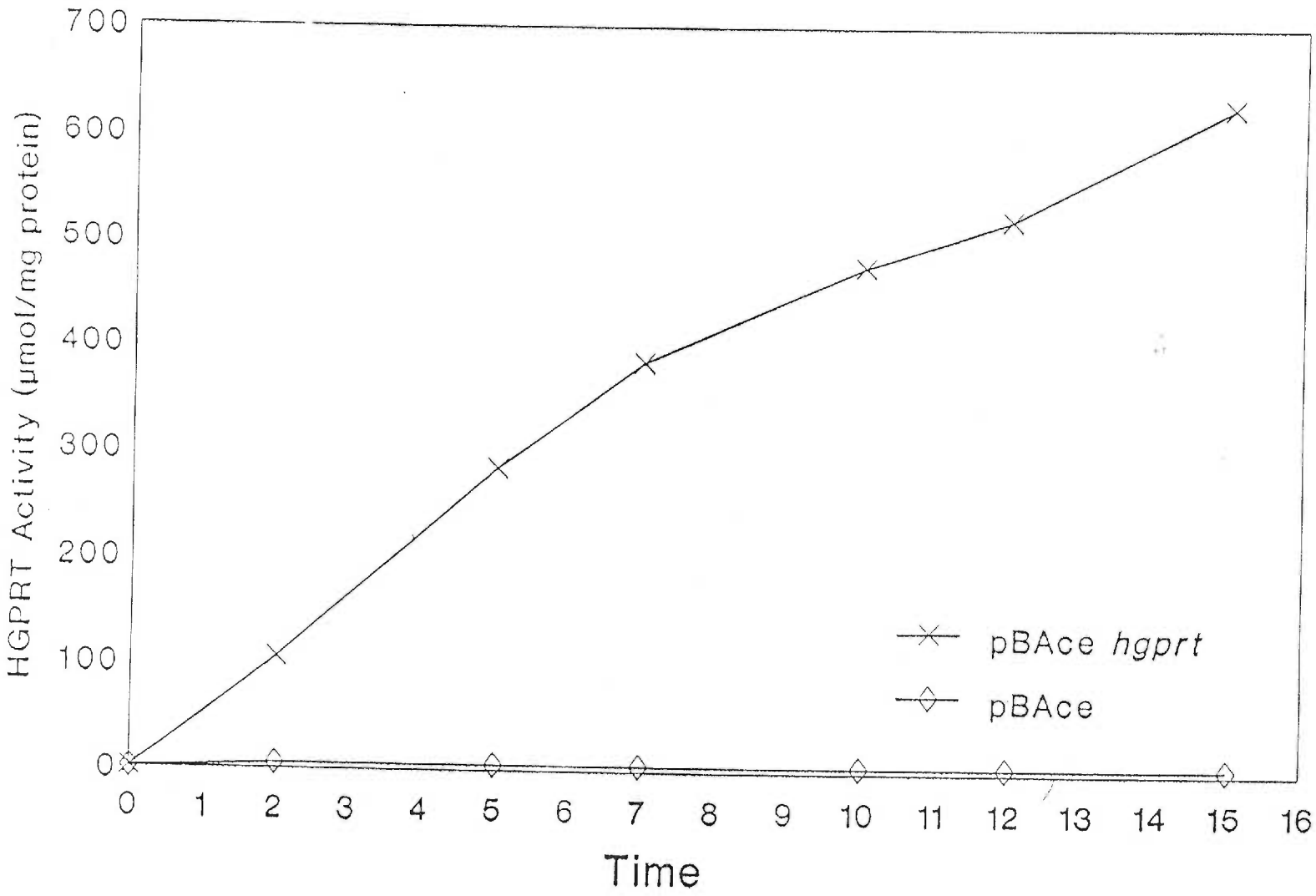

HGPRT Activity (µmol/mg protein)
700
600
500
400
300
200
100
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Time
pBAce hgprt
pBAce

FIG. 6. **SDS polyacrylamide gel electrophoresis of expressed and purified trypanosomal HGPRT.** Sϕ606 *E. coli* containing the control and *hgprt*-containing pBAce plasmids were expressed as described in Materials and Methods and the legend to Fig. 5 and the proteins fractionated by SDS polyacrylamaide gel electrophoresis. Recombinant *T. brucei* HGPRT was purified by GTP-agarose affinity chromatography as described in Materials and Methods. Lanes: A, proteins from Sϕ606 *E. coli* transformed with the control pBAce plasmid; B, proteins from Sϕ606 cells transformed with the *hgprt*-containing pBAce vector; C, PRPP-eluate of *T. brucei* HGPRT chromatographed over GTP-agarose column. ~10 μg of protein was loaded onto each lane of the SDS gel.

FIG. 7. **Purification of Recombinant *T. brucei* HGPRT.** The *T. brucei* HGPRT was purified from Sϕ606 lysates from a GTP-agarose resin by elution in TMD buffer containing 1.0 mM PRPP. Total HGPRT activity in each 5.0 ml fraction is plotted as a function of fraction number. The addition of 1.0 mM PRPP to the TMD buffer is indicated.

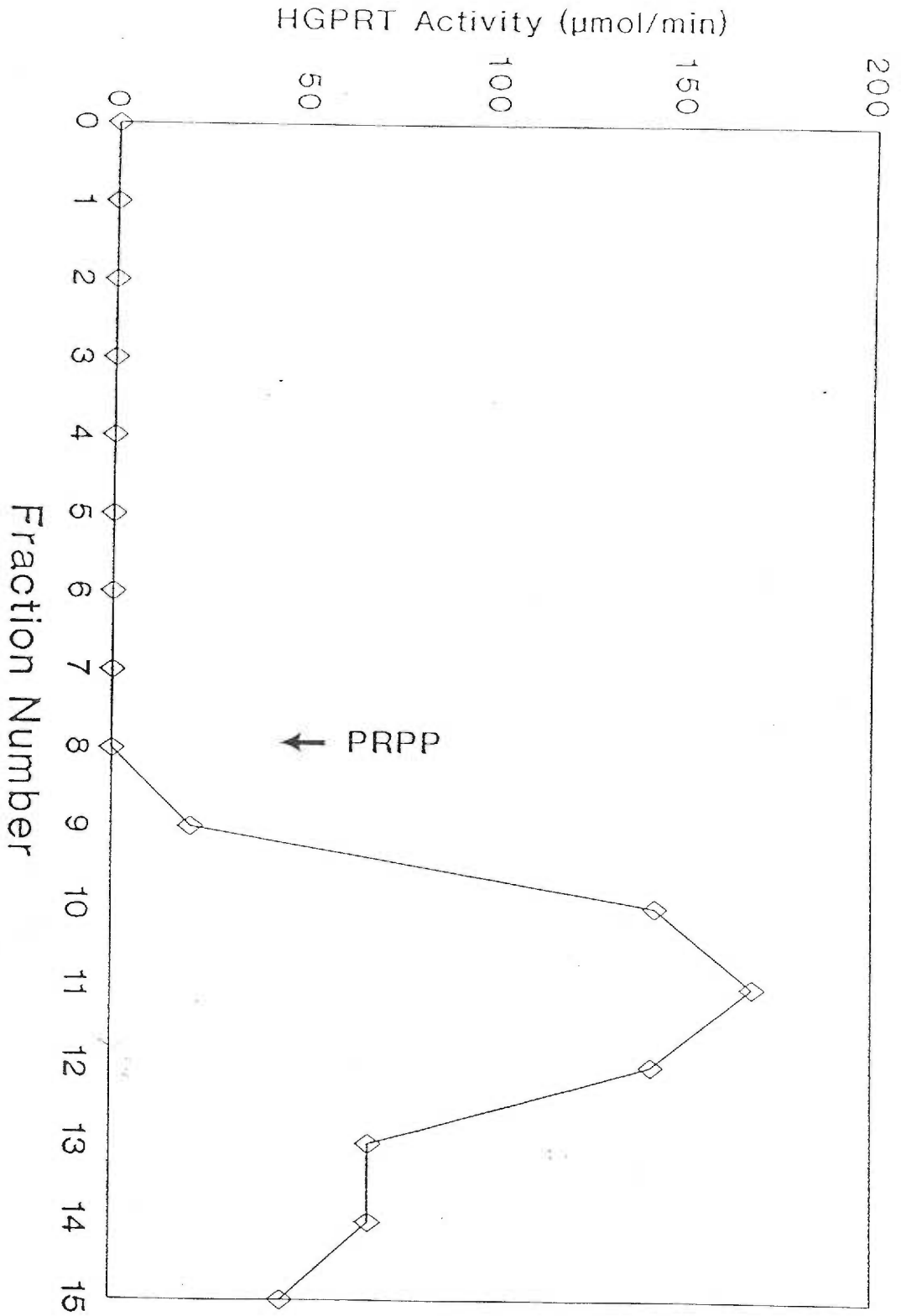
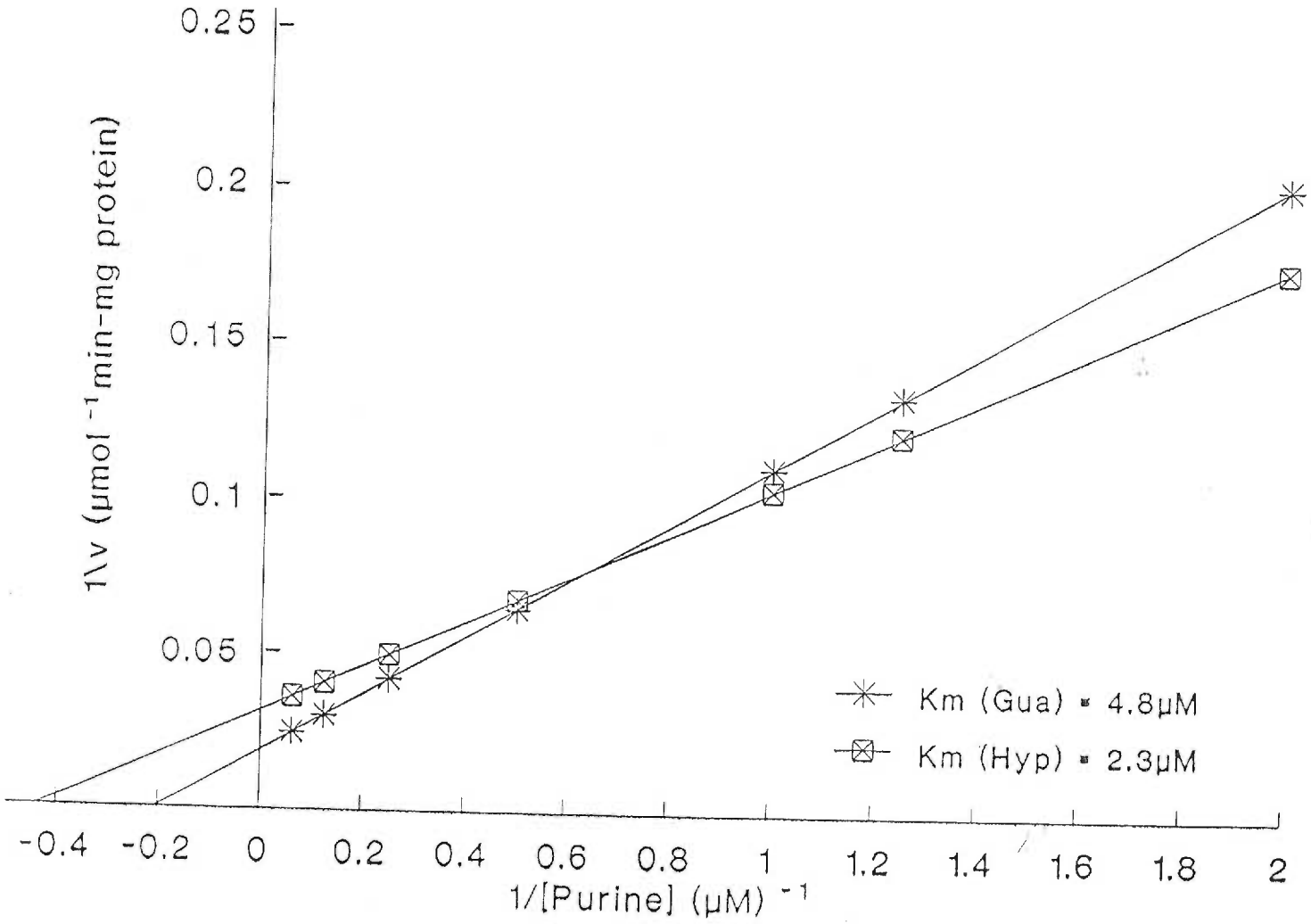

HGPRT Activity (µmol/min)
0
50
100
150
200
Fraction Number
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
PRPP

FIG. 8. **Lineweaver-Burk Analysis of Recombinant *T. brucei* HGPRT.** The ability of the homogeneous *T. brucei* HGPRT to phosphoribosylate hypoxanthine and guanine (Panel A), and HPP (Panel B) was determined at a number of substrate concentrations and the data analyzed by the method of Lineweaver-Burk.


0.25
0.2
0.15
0.1
0.05
1\v (μmol -1 min-mg protein)
Km (Gua) = 4.8μM
Km (Hyp) = 2.3μM
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
1/[Purine] (μM) -1

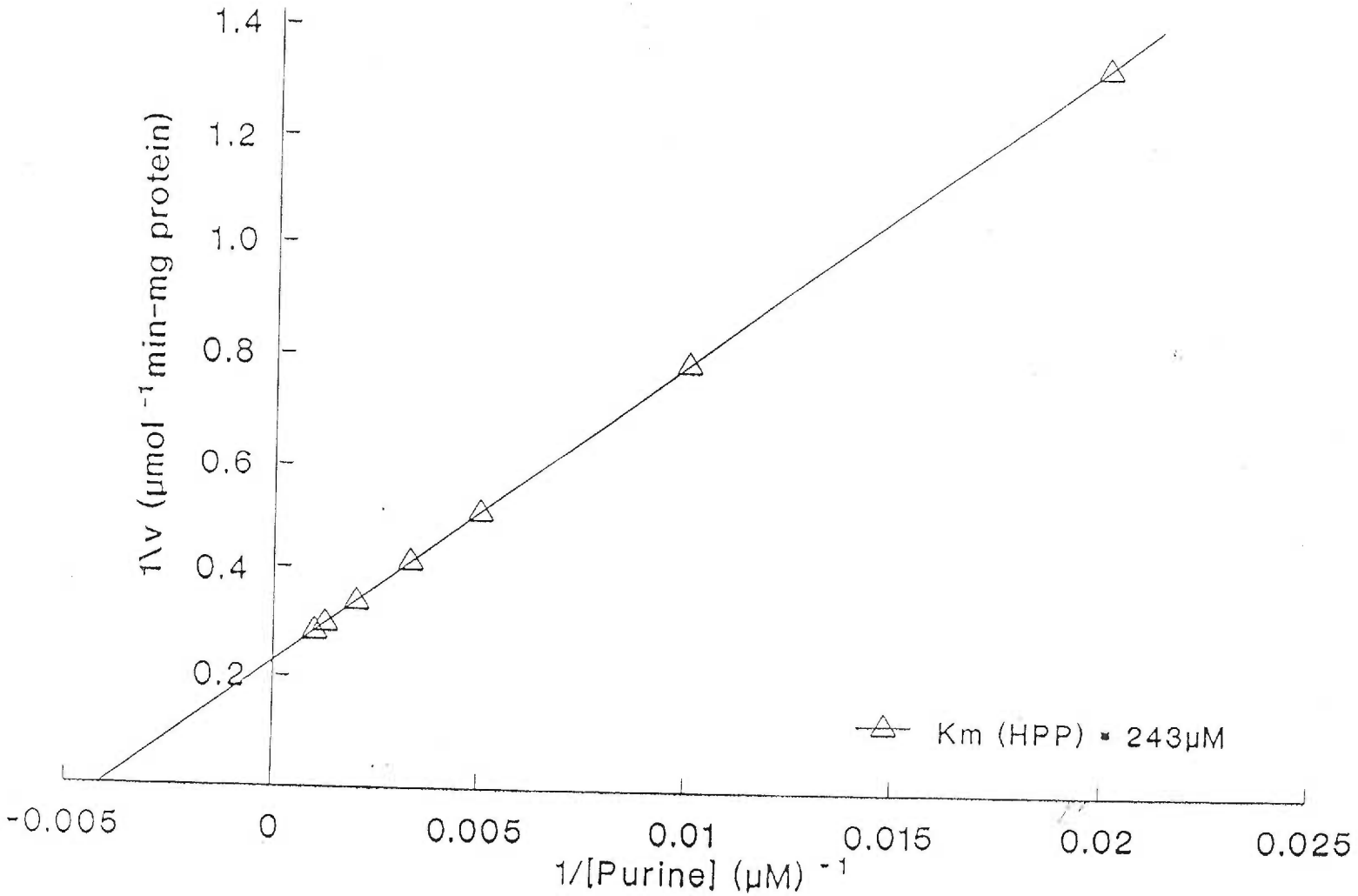

1.4
1.2
1.0
0.8
0.6
0.4
0.2
1\v (µmol -1 min-mg protein)
Km (HPP) = 243µM
-0.005
0
0.005
0.01
0.015
0.02
0.025
1/[Purine] (µM) -1

TABLE I

Comparison of HGPRT Amino Acid Sequences
(% Identity)

	Human	S. mansoni	P. falciparum	V. harveyi
Human				
S. mansoni	46%			
P. falciparum	44%	34%		
V. harveyi	29%	27%	28%	
T. brucei	21%	23%	22%	36%

TABLE II

Inhibitors of [^{14}C]hypoxanthine phosphoribosylation

The effects of 1.0 mM concentrations of various compounds on the ability of the T. brucei HGPRT to phosphoribosylate 20 μM [^{14}C]hypoxanthine were ascertained using the affinity-purified enzyme. The averages and standard deviations of five independent determinations are reported.

Addition	HGPRT Activity (umol/min/mg protein)
None	55.4 ± 1.3
Adenine	53.6 ± 11.8
Hypoxanthine	1.0 ± 1.3
Guanine	2.1 ± 1.3
Xanthine	52 ± 9.4
AMP	43.5 ± 9
IMP	17.2 ± 3
GMP	29 ± 6.8
ATP	30 ± 7.3
GTP	15.7 ± 6.3

ACKNOWLEDGEMENTS

This work was supported by grant AI-23682 from the National Institute of Allergy and Infectious Disease. Thomas Allen was a recipient of an N.L. Tartar Trust Fellowship from the Medical Research Foundation of Oregon.

REFERENCES

1. Williamson, J. (1970) in *The African Trypansomiases* (Mulligan, H.W., ed.) pp. 125-221, Wiley-Interscience, New York, NY
2. Leach, T.M. and Roberts, C.J. (1981) *Pharmacol. Ther.* **13**, 91-147
3. Geary, T.G., Edgar, S.A., and Jensen, J.B. (1986) in *Chemotherapy of Parasitic diseases* (Campbell, W.C., and Rew, R.S., eds.) pp. 209-236, Plenum Press, New York, NY
4. Marr, J.J., Berens, R.L., and Nelson, D.J. (1978) *Biochim. Biophys. Acta* **544**, 360-371
5. Heyworth, P.G., Gutteridge, W.E., and Ginger, C.D. (1978) *J.Protozool.* **25**, 9B
6. Sherman, I.W. (1979) *Microbiol. Rev.* **43**, 453-495.
7. Boonlayangoor, P., Albach, R.A., and Booden, T. (1980) *Arch. Invest. Med. (Mex)* **11**, 83-88
8. Berens, R.L., Marr, J.J., LaFon, S.W., and Nelson, D.J. (1981) *Mol. Biochem. Parasitol.* **3**, 187-196
9. Wang, C.C., and Simashkevich, P.M. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6618-6622
10. Fish, W.R., Marr, J.J., and Berens, R.L. (1982) *Biochim. Biophys. Acta* **714**, 422-428
11. Schwartzman, J.D., and Pfefferkorn, E.R. (1982) *Exp. Parasitol.* **53**, 77-86
12. Wang, C.C., and Aldritt, S. (1983) *J. Exp. Med.* **158**, 1703-1712
13. Fish, W.R., Looker, D.L., Marr, J.J. and Berens, R.L. (1982) *Biochim. Biophys. Acta* **719**, 223-231
14. Schmidt, G., Walter, R.D., and Konigk, E. (1975) *Tropenmed. Parasitol.* **26**, 19-26

15. Marr, J.J.. and Berens, R.L. (1983) *Mol. Biochem. Parasitol.* **7**, 339-356

16. Ullman, B. (1984) *Pharmaceut. Res.* 1, 194-203

17. Fish, W.R., Marr, J.J., Berens, R.L., Looker, D.L., Nelson, D.J., LaFon, S.W., and Balber, A.E. (1985) *Antimicrob. Agents Chemother.* **27**, 33-36

18. Palella, T.D. and Fox, I.H. (1989) in *The Metabolic Basis of Inherited Diseases* (Scriver, C.S., Beaudet, A.L., Sly, W.S. and Valle, D., eds), 6th edition, pp. 965-1006, McGraw Hill Book Co., New York, NY

19. Brun, R., and M. Schonenberger. (1979) *Acta Trop.* **36**, 289-292

20. Wilson, K., Collart, F.R., Huberman, E., Stringer, J.L., and Ullman, B. (1991) *J. Biol. Chem.* **266**, 1665-1671

21. Landfear, S.M., and Wirth, D.F. (1985) *Mol. Biochem. Parasitol.* **15**, 61-82

22. Jolly, D.J., Okayama, H., Berg, P., Esty, A.C., Filpula, D., Bohlen, P., Johnson, G.G., Shively, J.E., Hunkapillar, T., and Friedman, T. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 477-481

23. Konecki, D.S., Brennand, J., Fuscoe, J.C., Caskey, C.T., and Chinault, A.C. (1982) *Nucl. Acids Res.* **10**, 6763-6775

24. Craig, S.P.III, McKerrow, J.H., Newport, G.R., and Wang, C.C. (1988) *Nucl. Acids Res.* **16**, 7087-7101

25. King, A. and Melton, D.W. *Nucl. Acids Res.* **15**, 10469-10481

26. Hanson, S., Adelman, J., and Ullman, B. (1992) *J. Biol. Chem.* **267**, 2350-2359

27. Davis, L.G. Dibner, M.D. and Battey, J.F. (1986) *Basic Methods in Molecular Biology*, Elsevier, New York

28. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467

29. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning.*

A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

30. Higgins, D.G., and Sharp, P.M. (1988) *Gene* **73**, 237-244

31. Dayhoff, M.O. (1978) in *Atlas of Protein Sequence and Structure* (Dayhoff, M.O., ed.) Vol. 5, Suppl. 3., pp. 345-358, National Biomedical Research Foundation, Washington DC

32. Kawasaki, E.S. (1990) in *PCR Protocols* (Inns, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J., eds.) pp. 21-27, Academic Press, Inc., San Diego, CA

33. Wilson, K., Hanson, S., Landfear, S.M., and Ullman, B. (1991) *Nucl. Acids Res.* **19**, 5787

34. Murphy, W.J., Watkins, K.P., and Agabian, N. (1986) *Cell* **47**, 517-525

35. Sutton, R.E., and Boothroyd, J.C. (1986) *Cell* **47**, 527-535

36. Craig, S.P.III, Yuan, L., Kuntz, D.A., McKerrow, J.H., and Wang, C.C. (1988) *Proc. Natl. Acad. Sci. USA* **88**, 2500-2504

37. Jochimsen, B., Nygaard, P., and Vestergaard, T. (1975) *Mol. Gen. Genet.* **143**, 85-91

38. Ames, B.N., and Dubin, D.T. (1960) *J. Biol. Chem.* **235**, 769-775

39. Allen, T., Henschel, E.V., Coons, T., Cross, L., Conley, J.,and Ullman, B. (1989) *Mol. Biochem. Parasitol.* **33**, 273-281

40. Showalter, R.E. and Silverman, M.R. (1990) *Nucl. Acids Res.* **18**, 4621

41. Sogin, M.L., Elwood, H.J., and Gunderson, J.H. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1383-1387

42. Hershey, H.V. and Taylor, M.W. (1986) *Gene* **43**, 287-293

43. Dush, M.K., Sikela, J.M., Khan, S.A., Tischfield, J.A., and Sambrook, P.J. (1987) *Proc. Natl. Acad. Sci. USA* **82**, 2731-2735

44. Tomashow, L.S., Milhausen, M., Rutter, W.J., and Agabian, N. (1983) *Cell*

32, 35-43

45. Tschudi, C., and Ullu, E. (1988) *EMBO J.* **7**, 455-463

46. Richardson, K.K., Fostel, J., and Skopek, T.R. (1983) *Proc. Natl. Acad. Sci. USA* **11**, 8809-8816

47. Phillips, M. A., Coffino, P., and Wang. C.C. (1987) *J. Biol. Chem.* **262**, 8721-8727

48. Wilson, K., Berens, R.L., Sifri, C.D., and Ullman, B., manuscript submitted

49. Chou, P.Y., and Fasman, G.D. (1974) *Biochemistry* **13**, 222-245

50. McKusick, V.A. (1992) *Mendelian Inheritance in Man* pp. 1873-1880, The Johns Hopkins University Press, Baltimore, MD

51. Poulsen, P. Jensen, K.F., Valentin-Hansen, P., Carlsson, P., and Lundberg, L.G. (1983) *Eur. J. Biochem.* **135**, 223-229

52. Tso, J.Y., Zalkin, H., Van Cleemput, M., Yanofsky, C., and Smith, J.M. (1982) *J. Biol. Chem.* **257**, 3525-3531

53. Queen, S.A., Jagt, D.V., and Reyes, P. (1988) *Mol. Biochem. Parasitol.* **30**, 123-134

54. Eells, J., and Spector, R. (1983) *Neurochem. Res.* **8**, 1451-1457

55. Looker, D.L., Berens, R.L., and Marr, J.J. (1983) *Mol. Biochem. Parasitol.* **9**, 15-28

56. Tuttle, J.V., and Krenitsky, T.A. (1980) *J. Biol. Chem.* **255**, 909-916

57. Martinez, S., and Marr, J.J. (1992) *New Engl. J. Med.* **326**, 741-744

58. Gallerano, R.H., Sosa, R.R., and Marr, J.J. (1990) *Am. J. Trop. Med. Hyg.* **43**, 159-166

59. Metcalf, B.W., Bey, P., Danzin, C., Casera, P., and Vevert, J.P. (1978) *J. Am. Chem. Soc.* **100**, 2661-2553

60. Schechter, P.J., Barlow, J.L.R., and Sjoerdsma, A. (1987) in *Inhibition*

of Polyamine Metabolism, Biological Significance and Basis for New Therapies (McCann, P.P., Pegg, A.E., and Sjoerdsma, A., eds.) pp. 345-364, Academic Press, New York

61. Ghoda, L., Phillips, M.A., Bass, K.F., Wang, C.C., and Coffino, P. (1990) *J. Biol. Chem.* **265**, 11823-11826

62. Bellofatto, V., Fairlamb, A.H., Henderson, G.B., and Cross, G.A.M. (1987) *Mol. Biochem. Parasitol.* **25**, 227-238

63. Phillips, M.A., and Wang, C.C. (1987) *Mol. Biochem. Parasitol.* **22**, 9-17

Cloning and Overexpression of the Hypoxanthine-Guanine Phosphoribosyltransferase Gene from *Trypanosoma cruzi*

by

Thomas Allen and Buddy Ullman *

Department of Biochemistry and Molecular Biology

Oregon Health Sciences University

Portland, OR 97201-3098

Telephone (503) 494-8437 - FAX (503) 494-8393

Internet: ullmanb@ohsu.edu

* To whom correspondence should be addressed

The hypoxanthine-guanine phosphoribosyltransferase (HGPRT) enzyme in *Trypanosoma cruzi* serves as a rational target for the selective therapeutic manipulation of Chagas disease, a catastrophic parasitic disease that is intractable to contemporary drug therapy. To establish a structure-based strategy for rational drug design, the *T. cruzi* gene (*hgprt*) encoding HGPRT has now been isolated and sequenced, and its predicted polypeptide compared to other eukaryotic HGPRTs. High level expression of *T. cruzi hgprt* was obtained in *E. coli*, and large amounts of the recombinant protein were purified to homogeneity by GTP-agarose affinity chromatography. Substrates for the pure *T. cruzi* HGPRT include hypoxanthine and guanine, as well as allopurinol, a pyrazolopyrimidine isomer of hypoxanthine that exhibits therapeutic efficacy in patients with Chagas disease and is nontoxic to humans.

Trypanosoma cruzi is the etiologic agent of Chagas disease, a devastating and often fatal parasitic disease in humans. Approximately 16-18 million people are infected with *T. cruzi* worldwide, and >90 million people are at risk for the disease (1). Over 100,000 individuals in the United States are infected with the parasite (2), which poses a threat to the nation's blood supply, and several cases of transfusion-mediated transmission have been documented (3). Chagas disease has two phases, the acute and chronic stages, each of which can be fatal (2,4). Acute Chagas patients have fever, malaise, lymphadenopathy, hepatomegaly, splenomegaly, and myocarditis, which can lead to fatal cardiac arrhythmias. Chronic Chagas disease, which develops several months post infection, is characterized by severe cardiac complications, including atrioventricular blocks, angina, congestive heart failure, cardiac thrombi, embolic strokes, and arrhythmias. Megacolon and megaesophagus are also frequent manifestations of chronic Chagas disease.

The current arsenal of drugs that have been used to treat Chagas disease is far from ideal. These agents include nifurtimox and benznidazole, which are ineffective against chronic Chagas disease, provoke multiple serious adverse side effects, and have proven carcinogenic in animal studies (5). The mechanism(s) of action of these two drugs are unknown. The lack of absolute selectivity of nifurtimox and benznidazole for *T. cruzi* can be ascribed to the lack of target specificity between the metabolic machinery of the parasite and the human host.

The development of a specific and selective drug for the treatment of Chagas or any other parasitic disease depends on fundamental biochemical distinctions between the

parasite and human host. One metabolic pathway that offers an avenue for the selective therapeutic manipulation of Chagas disease is the purine pathway. Whereas mammalian cells can synthesize purine nucleotides *de novo*, *T. cruzi* (6), as well as other parasitic protozoa (7), are auxotrophic for purines. To scavenge host purines, *T. cruzi* possesses a unique series of purine salvage enzymes that enable them to scavenge host purines (8). One purine salvage enzyme, hypoxanthine-guanine phosphoribosyltransferase (HGPRT), initiates the intracellular metabolism of allopurinol (4-hydroxypyrazolo(3,4-*d*)pyrimidine, HPP), a pyrazolopyrimidine isomer of hypoxanthine, to the nucleotide level in the parasite (9). As a consequence of this metabolism, HPP exhibits substantial toxicity toward *T. cruzi* grown *in vitro* (9,10) and in experimental animals (11). Moreover, the drug demonstrates therapeutic efficacy in patients with chronic Chagas disease, although sporadic adverse reactions and occasional therapeutic failure are observed (12). HPP, a drug widely employed in humans for the treatment of hyperuricemia and gout (13), is nontoxic to human cells because it is an inefficient substrate for the human HGPRT (14), and only minute quantities of HPP ribonucleotide are accumulated in mammalian cells (15). Moreover, the parasite HGPRT is likely to fulfill a critical nutritional function by phosphoribosylating hypoxanthine, the predominant salvageable purine within the vertebrate host milieus in which the parasite resides (16). Thus, HGPRT substrate analogs or inhibitors might serve as more effective and selective agents for the treatment and control of Chagas disease.

To characterize the *T. cruzi* HGPRT enzyme in detail and to develop a molecular reagent for exploitation in structure-based strategies for drug design, the gene encoding

HGPRT (*hgprt*) was cloned from a genomic library by hybridization to a 174 bp fragment of the *T. cruzi hgprt* that had been amplified from genomic DNA using the polymerase chain reaction (PCR) (17). Sequence analysis of a genomic clone revealed an open reading frame of 663 nucleotides encoding a polypeptide of 221 amino acids with a M_r=26 kd (Fig. 1). Alignment of the predicted amino acid sequence of the *T. cruzi* HGPRT with the human (18), *Schistosoma mansoni* (19), and *Plasmodium falciparum* (20) HGPRTs revealed 3 regions of significant homology, each separated and flanked by much longer regions without noticable similarity (Fig. 1). The stretches of amino acid similarity correspond to the two putative substrate binding domains (21) and to a region proximal to the COOH terminus of HGPRT to which no function has been ascribed. Alignments of the HGPRTs in a pairwise manner revealed that the *T. cruzi* enzyme exhibited only a 21-25 % identity with the human, *S. mansoni*, and *P. falciparum* HGPRTs. The extent of similitude among the various HGPRTs conformed to the phylogenetic divergence among the organisms expected from *rRNA* alignments (22).

Molecular characterization of the *hgprt* revealed a 1.8 kb transcript in *T. cruzi* epimastigotes (Fig. 2) and indicated that the *hgprt* structural locus was not arranged as a tandem repeat within the *T. cruzi* genome (data not shown) as are many other genes in parasites of the Trypanosomatidae family (23). Digestion of genomic DNA with either *Hind*III, *Nde*I, or *Sal*I, each of which cleaves the *T. cruzi hgprt* once, excised 2 fragments of dissimilar sizes, whereas a tandemly repeated gene would yield a band of identical size in all lanes of a Southern blot in which the genomic DNA had been digested with a single cutter.

After appropriate PCR mutagenesis of the termini of the protein coding portion of the *T. cruzi hgprt*, the gene was ligated into the bacterial expression vector pBAce (24) and the recombinant construct transformed into *E. coli* strain Sϕ606 (*Δgpt-pro-lac, thi, hpt*). Sϕ606 cells are genetically deficient in both xanthine-guanine phosphoribosyltransferase (XGPRT) and hypoxanthine phosphoribosyltransferase (HPRT) activities (25). After induction, the Sϕ606 *E. coli* transformed with the *hgprt*-containing pBAce plasmid expressed high levels of HGPRT activity, whereas no HGPRT was detected in Sϕ606 cells transformed with the pBAce vector alone (Fig. 3). The soluble proteins from the Sϕ606 cells expressing trypanosomal HGPRT were resolved by SDS-PAGE (Fig. 3). The recombinant HGPRT at 26 kd was the predominant protein, accounting for ~70 % of the Coomassie blue stainable protein as evaluated by densitometric analysis.

The recombinant *T. cruzi* HGPRT was purified to homogeneity in a single step by affinity chromatography using a GTP-agarose resin (Fig. 3). The protein was eluted from the column in buffer containing 1.0 mM phosphoribosylpyrophosphate (PRPP), a substrate of the enzyme. The substrate specificity of the *T. cruzi* HGPRT was examined with a variety of purine bases. The purified HGPRT catalyzed the phosphoribosylation of hypoxanthine and guanine but did not recognize adenine or xanthine. The *T. cruzi* HGPRT also phosphoribosylated HPP, although the pyrazolopyrimidine was not as an effective substrate as the naturally occurring purine bases (Table I). Apparent K_m values of 6.4 μM, 9.9 μM, and 714 μM for hypoxanthine, guanine, and HPP, respectively were calculated by Lineweaver-Burk analysis, and the k_{cat} values for hypoxanthine and guanine

were much higher than that for HPP. Apparent K_m values for PRPP were not obtained, since the pure enzyme was extremely labile in the absence of PRPP. In the presence of PRPP, the *T. cruzi* HGPRT was completely stable for >2 months at temperatures ranging from 4° to -80°.

The isolation of the *T. cruzi hgprt* gene, the functional overexpression of the *T. cruzi* HGPRT in *E. coli*, and the purification of large quantities of homogeneous recombinant enzyme provide a molecular foundation for a biochemically rational strategy for therapeutic intervention in an effectively incurable parasitic disease. A thorough molecular genetic dissection of the functional determinants of the *T. cruzi* HGPRT accompanied by the solution of the enzyme's structure by x-ray crystallography will provide an avenue for a structure-based approach for the design and discovery of novel HGPRT substrate analogs or inhibitors that can serve as a selective treatment for Chagas disease.

Fig. 1. Alignment of the predicted primary structure of the *T. cruzi* HGPRT with HGPRT proteins from human (18), *S. mansoni* (19), and *P. falciparum* (20) according to the CLUSTAL V multisequence alignment program, a modification of that previously described (26). Asterisks indicate identical residues and dots indicate conserved amino acids with similarity scores >10, as calculated by the log-odds similarity matrix of Dayhoff (27). The entire nucleotide sequence of the protein coding segment of the *T. cruzi hgprt* and portions of the 5' and 3' untranslated regions have been deposited in GenBank under the accession number L07486. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D. Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

Alignment of *T. cruzi* HGPRT with Eukaryotic HGPRTs

```
Human           1   MAT---RSPG-----VVISDDEPGYDLDLFCIPNHYAEDLERVFIPHGLI   42
S.mansoni       1   MSSNMIKADC-----VVIEDSFRGFPTEYFCTSPRYDECLDYVLIPNGMI   45
P.falciparum    1   MPIPNNPGAGENAFDPVFVKDDDGYDLDSFMIPAHYKKYLTKVLVPNGVI   50
T.cruzi         1   MP--------------------REYEFA-----------EKILFTEEEI   18
                    *.                                    .. . *

Human          43   MDRTERLARDVMKEMGGHH-------IVALCVLKGGYKFFADLLDYIKAL   85
S.mansoni      46   KDRLEKMSMDIVDYYEACN----ATSITLMCVLKGGFKFLADLVDGLERT   91
P.falciparum   51   KNRIEKLAYDIKKVY---N----NEEFHILCLLKGSRGFFTALLKHLSRI   93
T.cruzi        19   RTRIMEVAKRIADDYKGKGLRPYVNPLVLISVLKGSFMFTADLCRAL---   65
                      *   ..  .              . . .***. * . *   .

Human          86   NRNSDRSIPMTV---DFIRLKSYCNDQSTGDIKVIGGDDLSTLTGKNVLI  132
S.mansoni      92   VRARGIVLPMSV---EFVRVKSYVNDVSIHEPILTGLGDPSEYKDKNVLV  138
P.falciparum   94   HNYSAVEMSKPLFGEHYVRVKSYCNDQSTGTLEIVS-EDLSCLKGKHVLI  142
T.cruzi        66   ---SDFNVPVRM---EFICVSSYGEGVTSSGQVRMLLDTRHSIEGHHVLI  109
                         .. .  ... . ** .. .                . .**.
                    Prim I
Human         133   VEDIIDTGKTMQTLLSLVRQYNPKMVKVASLLVKRTPRSVGYKPDFVGFE  182
S.mansoni     139   VEDIIDTGKTITKLISHLDSLSTKSVKVASLLVKRTSPRNDYRPDFVGFE  188
P.falciparum  143   VEDIIDTGKTLVKFCEYLKKFEIKTVAIACLFIKRTPLWNGFKADFVGFS  192
T.cruzi       110   VEDIVDTALTLNYLYHMYFTRRPASLKTVVLLDKREGRRVPFSADYVVAN  159
                    ****.**. *.  .           .    *. **      . .*.*
                        Prim II
Human         183   IPDKFVVGYALDYNEYFRDLNHVCVIS-------ETGKAKYKA-------  218
S.mansoni     189   VPNRFVVGYALDYNDNFRDLHHICVIN-------EVGQKKFSVPCTS---  229
P.falciparum  193   IPDHFVVGYSLDYNEIFRDLDHCCLVN-------DEGKKKYK----A---  229
T.cruzi       160   IPNAFVIGYGLDYDDTYRELRDIVVLRPEVYAEREAARQKKQRAIGSADT  209
                    .*. **.**.***.. .*.* .  ..        . .. *

Human                ------------                                     218
S.mansoni     230    ---------KPV                                     231
P.falciparum  230    ---------TSL                                     231
T.cruzi       210    DRDAKREFHSKY                                     221
```

Fig. 2. Molecular characterization of the *T. cruzi hgprt*. (**A**) Northern blot analysis of RNA prepared from epimastigotes of the Y strain of *T. cruzi*. 10 μg RNA was hybridized to a 663 bp *Nde*I-*Sal*I fragment that contained the entire *T. cruzi hgprt* coding region using conventional protocols (28). The autoradiograph is a 24 hr exposure at -70°. Molecular size standards are indicated on the left.

Fig. 3. Overexpression of *T. cruzi hgprt* in *E. coli* and purification of recombinant HGPRT. (**A**) The termini of the *T. cruzi hgprt* were mutated by PCR mutagenesis, and the gene was ligated into the *Nde*I and *Sal*I sites of the pBAce expression plasmid and transformed into Sϕ606 (Δ*gpt-pro-lac, thi, hpt*) *E. coli*. HGPRT was expressed in a low phosphate medium as described except that the concentration of "vitamin free" Casamino acids was reduced to 0.1%, and the sodium phosphate was not added. Transformed *E. coli* were grown 12 hr at 37° to an OD_{600} = 0.9-1.2, the cells extracted, and 100,000 X g supernatants prepared by centrifugation. HGPRT activity (29) was monitored in 100,000 X g supernatants of cells transformed with the *hgprt* in pBAce (pBAce-*hgprt*) or with the pBAce vector alone. (**B**) SDS-PAGE fractionation of recombinant HGPRT. ~10 μg of protein from the *E. coli* transformed either with pBAce alone (lane A) or with the *hgprt*-containing pBAce (lane B) were fractionated by SDS-PAGE and stained with Coomassie Blue. The HGPRT purified by affinity chromatography was also subjected to SDS-PAGE (lane C).

Table I. Substrate specificity of the *T. cruzi* HGPRT. The ability of pure recombinant HGPRT to phosphoribosylate a variety of purine nucleobases was ascertained in 1.0 mM PRPP and 20 μM nucleobase substrate (29). The results presented are the averages and standard deviations for at least 3 independent determinations. K_m and k_{cat} values for hypoxanthine, guanine, and HPP were determined by Lineweaver-Burk analysis.

Substrate	K_m (μM)	k_{cat} (μmol/min/mg protein)
Hypoxanthine	6.4 ± 1.1	97 ± 6.8
Guanine	9.9 ± 0.3	55 ± 0.8
Adenine	no activity	
Xanthine	no activity	
HPP	714 ± 125	0.8 ± 0.07

REFERENCES AND NOTES

1. J.C.P. Dias, *Parasitol. Today* 3, 336 (1987); World Health Organization, *Wkly. Epidem. Rec.* **65**, 257 (1990).
2. J.M. Hagar, S.H. Rahimtoola, *N. Engl. J. Med.* **325**, 763 (1991).
3. L.V. Kirchhoff, *Ann. Intern. Med.* **111**, 773 (1989).
4. L.V. Kirchhoff and F.A. Neva, *JAMA* **254**, 3058 (1985); M. Katz, D.D. Despommier, R. Gwadz, Eds. *Parasitic Diseases.* (Springer-Verlag, New York, 1989), pp. 170-176.
5. W.C. Campbell and R.S. Rew, *Chemotherapy of Parasitic Diseases.* (Plenum Press, New York, 1986) chap 4, pp. 115-127; L.V. Kirchhoff, A.A. Gam., F.C. Gilliam, *Am. J. Med.* **82**, 915 (1987); J.J. Marr, *J. Lab. Clin. Med.* **118**, 111 (1991).
6. W.E. Gutteridge and M. Gaborak, *Int. J. Biochem.* **10**, 415 (1979); J.J. Marr, R.L. , Berens, and D.J. Nelson, *Mol. Biochem. Parasitol.* **3**, 187 (1981).
7. C.C. Wang, *J. Med. Chem.* **27**, 1 (1984).
8. W.E. Gutteridge and M.J. Davies, *FEBS Lett.* **127**, 211 (1981); W.E. Gutteridge and M.J. Davies, *FEMS Microbiol. Lett.* **13**, 207 (1982).
9. J.J. Marr, R.L. Berens, D.J. Nelson, *Science* **201** 1018 (1978).
10. R.L. Berens, J.J. Marr, F.S. da Cruz, F.S., D.J. Nelson, *Antimicrob. Agents Chemother.* **22**, 657 (1982).
11. J.L. Avila and A. Avila, *Exp. Parasitol.* **51**, 204 (1981).
12. R.H. Gallerano, J.J. Marr, R.R. Sosa, *Am. J. Trop. Med. Hyg.* **43**, 159 (1990); J.J. Marr, *J. Lab. Clin. Med.* **118**, 111 (1991).
13. T.D. Palella and I.H. Fox, (1989) in *The Metabolic Basis of Inherited Diseases*, C.S. Scriver, A.L. Beaudet, W.S. Sly, D. Valle, Eds, (McGraw Hill Book Co., New York 1989), pp. 965-1006.

14. J.B. Tuttle and T.A. Krenitsky, *J. Biol. Chem.* **255**, 909 (1980).

15. D.J. Nelson, C.J.L. Bugge, H.C. Krasny, G.B. Elion, *Biochem. Pharmacol.* **22**, 2003 (1973).

16. J. Eells and R. Spector, *Neurochem. Res.* **8**, 1451R. (1983).

17. The *T. cruzi hgprt* was isolated from a genomic library by hybridization to a 174 bp fragment of the *T. cruzi hgprt* that had been amplified from genomic DNA using the polymerase chain reaction (PCR). The sense primer, 5'-TCTGGATCC-GAG-GA[CT]-AT[CT]-AT[CT]-GA[CT]-AC-3, corresponded to amino acids 134-139, EDIIDT, of the human HGPRT and contained a 9 nucleotide leader encompassing a *Bam*HI site. The antisense primer, TCTGAATTC-[AG]TA-[ACG]CC-[ACG]AC-[ACG]AC-[AG]AA-3', fit residues 187-191, FVVGY, of the human HGPRT and was preceded by an *Eco*RI site. These amino acid stretches are highly conserved among other eukaryotic HGPRTs (18-21). The PCR was performed for 30 cycles of 0.5 min at 94°, 0.5 min at 40°, 1.0 min at 72° using 25 pmol of each degenerate primer and 100 ng *T. cruzi* DNA. The amplified 174 kb fragm0ent was subcloned into pBluescript KS^+ (Stratagene), sequenced using [^{35}S]dATP by the dideoxy method with Sequenase 2.0 (U.S. Biochemicals), and used to probe a genomic library of *T. cruzi* DNA under high stringency conditions. A 3.2 kb *EcoRI* fragment was subcloned from a plaque-purified phage and the entire *hgprt* coding sequence obtained in both directions as described above for the PCR fragment using custom-made primers.

18. D.J. Jolly *et al., Proc. Natl. Acad. Sci. U.S.A.* **80**, 477 (1983); D.S. Konecki, J. Brennand, J.C. Fuscoe, C.T. Caskey, A.C. Chinault, *Nucl. Acids Res.* **10**, 6763 (1982).

19. S.P. Craig III, J.H. McKerrow, G.R. Newport, C.C. Wang, *Nucl. Acids Res.*

16, 7087 (1988).

20. A. King and D.W. Melton, *Nucl. Acids Res.* 15, 10469 (1987).

21. H.V. Hershey and M.W. Taylor, *Gene* 43, 287 (1986).

22. M.L. Sogin, H.J. Elwood, J.H. Gunderson, *Proc. Natl. Acad. Sci. U.S.A.* 83, 1383 (1986); J.A. Lake, V.F. de la Cruz, P.C.G. Ferreira, C. Morel, L. Simpson, *Proc. Natl. Acad. Sci. U.S.A.* 85, 4779 (1988).

23. S.M. Landfear, D.M. Pratt, D.F. Wirth, *Mol. Cell. Biol.* 3, 1070 (1983); R.G. Nelson *et al.*, *Cell* 34, 901 (1983); C. Tschudi, A.S. Young, L. Ruben, C.L. Patton, F.F. Richards, *Proc. Natl. Acad. Sci. U.S.A.* 82, 3998 (1985).

24. S.P. Craig III, L. Yuan, D.A. Kuntz, J.H. McKerrow, C.C. Wang, *Proc. Natl. Acad. Sci. USA* 88, 2500 (1988).

25. B. Jochimsen, P. Nygaard, T. Vestergaard, *Mol. Gen. Genet.* 143, 85 (1975).

26. D.G. Higgins and P.M. Sharp, *Gene* 73, 237 (1988).

27. M.O. Dayhoff, in *Atlas of Protein Sequence and Structure*, M.O. Dayhoff, Ed., (National Biomedical Research Foundation, Washington DC, 1978), vol. 5, suppl. 3., pp. 345-358.

28. J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning. A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989).

29. T. Allen *et al.*, *Mol. Biochem. Parasitol.* 33, 273 (1989).

30. This research was supported by NIH grant AI-23682 to B.U. We thank Dr. Wesley van Voorhuis for making available a genomic library of *T. cruzi* DNA.

Discussion

The work presented in this thesis involved the cloning, expression and characterization of the HGPRT enzyme from three members of the *Trypanosomatidae* family. The three parasitic protozoa studied were *Leishmania donovani*, *Trypanosoma brucei* and *Trypanosoma cruzi*. The rational for this work is based on the hypothesis that the examination of a unique enzyme involved in the basic metabolic machinery of these organism would lead to the development of chemotherapeutic regimes that are selectively toxic to the parasite. The HGPRT enzyme is unique for several reasons. The most important is that parasitic protozoa are auxotrophic for purines and so require a functioning purine salvage pathway. In addition, the substrate specificity of the HGPRT enzyme in members of the *Trypanosomatidae* family is distinctive in that the enzyme is capable of metabolizing the purine analog, allopurinol, to the nucleotide level were it is then incorporated into the RNA of the parasite. Mammalian cells do not incorporate allopurinol into their RNA. For these reasons, we believe that the HGPRT enzyme is an ideal target for a multidisciplinary approach to structure-based drug design.

Several different strategies were employed to isolate the genes encoding the HGPRT enzymes. For the trypanosome, this included the use of degenerate oligonucleotides designed to conserved regions of the HGPRT enzyme from other eukaryotic counterparts for use in PCR amplification. For the *L. donovani hgprt*, it was necessary to isolate protein from the promastigote form of the organism for protein sequencing. A degenerate oligonucleotide designed to the protein sequence and a oligonucleotide designed to the splice leader

sequence were then used to amplify the 5' end of the *hgprt* gene which was in turn used to isolate a full length gene from a genomic library. The predicted molecular weight for the HGPRT enzymes was M_r = 23.4, 26, and 23.6 kd for the *T. brucei, T. cruzi* and *L. donovani* respectively.

Multiple sequence alignment analysis of the various HGPRTs reveled some striking features. The trypanosomatids exhibited a 18-25% identity with their eukaryotic counterparts while displaying a 32-36% identity to the prokaryotic HPT from V. *harveyi*. The considerable divergence between the trypanosomatids and their eukaryotic counterparts is surprisingly high. Indeed, the prokaryotic HPT actually exhibits a higher degree of identity to the other eukaryotes then to the trypanosomatids (28-30%). This suggests that the *hgprt* from the trypanosomatids are the most divergent of the group. The comparison of sequence between the trypanosomatids was also surprising. While the *L. donovani* and *T. cruzi* HGPRT exhibited a 50% identity to each other, the *T. brucei* HGPRT only displayed a 37% identity to the *L. donovani* HGPRT while having a 50% identity to *T. cruzi*.

Comparison of the amino acid sequence of the various HGPRTs displayed several regions were the conservation of residues was much higher then throughout the rest of the protein. A region corresponding to the human HGPRT encompassing the amino acids 66 - 78 are believed to be involved in the purine binding domain for this enzyme. This regions contains a completely conserved central region of a Leu-Lys-Gly tripeptide which is found in all eukaryotic HGPRTs. In addition, of the remaining ten amino acids, six show conserved substitutions making this a very evolutionary conserved region of the HGPRT enzyme. A second region which is also thought to be involved in purine binding is found at position

corresponding to amino acids 103 to 105 in the human sequence. This region has a tripeptide, Lys-Ser-Tyr, that is found in all eukaryotic HGPRTs except for the trypanosomatids. In the parasites the first amino acid shows a Lys for Ser substitution suggesting possible involvement in the unique substrate specificity of the trypanosomatids enzyme.

A second region corresponding to residues 127 to 144 in the human HGPRT, has been suggested as the PRPP binding domain. This region contains 6 identical and 8 similar amino acid residues among the parasites and the mammalian enzymes. The reason why this region has been suggested to encompass the PRPP binding domain comes not only from alignments of the HGPRTs with each other but also when the HGPRTs are aligned with other PRPP binding enzymes.

Finally, a region toward the carboxyl terminus of the HGPRT enzyme shows the longest stretch of similarity among the HGPRTs. This region, corresponding to residues 183 to 202 in the human HGPRT, contains 7 identical and 9 similar amino acid residues. The function of this region is still in doubt but secondary structure predictions suggest that this region may be important in the proper folding of the protein. It is known that several naturally occurring mutations in this region of the human HGPRT can eliminate or greatly reduce the ability of the enzyme to phosphoribosylate its substrates.

Although the HGPRTs from the trypanosomatids display extensive divergence from their eukaryotic counterparts at the primary sequence level, analysis of secondary structure using the algorithm of Chou and Fasman reveal several common features amongst all the HGPRTs. In particular, a B-x-B structure characteristic of a mononucleotide binding fold

is found before the purine binding motif in all HGPRTs examined to date.

Molecular characterization revealed a single transcript of 1.8 kb. for the *T. cruzi* and 1.9kb. for the *L. donovani hgprt*. In contrast, two transcripts were seen for the *T. brucei hgprt*, one at 1.4kb and the other at 1.9kb. Both transcripts were expressed in the mammalian and insect stages of the parasite. Analysis of the 5' and 3' portion of the mature transcript suggests that the heterogeneity of the transcript is found in the 3' end of the message. Southern blot analysis also showed a dissimilarity between *T. brucei* and the other two trypanosomatids. The *T. brucei hgprt* appears to be a multicopy gene whereas the *T. cruzi* and *L. donovani* appear to be a single copy. It is interesting to note that the *T. brucei* locus is not arranged in a tandem repeat as are many other genes in the trypanosomatids.

The overexpression of the HGPRT in an *E. coli* background that is deficient in HPRT and XGPRT activity has greatly facilitated the analysis and purification of the recombinant enzyme. Substrate specificity, K_m, and K_{cat} analysis have confirmed that the purine analog, allopurinol, is indeed a substrate for the trypanosomatids HGPRT enzyme. In addition, the failure of the recombinant enzyme to phosphoribosylate adenine or xanthine supports earlier reports that there are separate PRTs for these purine bases. Inhibitions studies involving the *T. brucei* and *T. cruzi* HGPRT recombinant enzyme showed that GTP, IMP, and GMP were all capable of affecting the ability of the recombinant protein to phosphoribosylate hypoxanthine. The effect of AMP and ATP however were markedly different between the two trypanosomes. Whereas AMP and ATP diminished phosphoribosylation activity for the *T. brucei* recombinant protein these two effectors appeared to have no effect on the *T. cruzi* enzyme.

Further analysis of the recombinant *T. cruzi* and *T. brucei* HGPRT enzyme involved an analysis of the quaternary structure of the active protein. Both enzymes were found to be in a dimer conformation in their active state. From these experiments it was also obvious that the cofactor, PRPP, was necessary for the enzyme to form the dimer. In addition, the pH at which the recombinant protein exhibit the greatest amount of phosphoribosylation activity was studied. The T. *brucei* HGPRT showed an optimum pH of 8.5 while for the T. *cruzi* the pH was 8.0.

Finally, a phylogenetic analysis was performed using the HGPRT enzyme from nine different organisms including the three trypanosomatids. Analysis revealed that the trypanosomatids diverge very early from the main branch of the eukaryotic lineage supporting previous studies. Although the branching amongst the trypanosomatids themselves is not strongly supported by this analysis, this work does suggest a closer relationship between *T. cruzi* and *L. donovani* than *T. brucei* and *L. donovani* show to each other.

Future Directions

The cloning and overexpression of the HGPRT enzyme from the trypanosomatids is just the first step in a thorough investigation into the molecular and biochemical characteristics of this important enzyme. It will be necessary to probe the mechanistic along with the structural features of this protein while also exploiting the enzyme as a probe into the biochemistry and cellular biology of the organisms. The following is a brief discussion of just a few of the directions this research can take and some of the investigations that are currently in progress.

The ability to purify large quantities of recombinant protein should enable a detailed examination of the steady-state kinetics of the HGPRT enzyme. At the present time only the *Schistosoma mansoni* HGPRT has been examined in any detail and their determination of an ordered bi-bi mechanism differs with earlier work which proposed that the human HGPRT enzyme uses a ping-pong kinetic mechanism. Although the study of the human HGPRT needs to be followed up with a more detailed analysis, the current evidence suggests that the kinetic mechanism may be dissimilar facilitating the hope that the difference could be exploited in the design of antischistosomal drugs. For the trypanosomatids, the HGPRT enzyme studies should include examination of the dynamics of the reaction in both the forward and reverse direction with each of the substrates. In addition, the binding order of the substrates can be determined by product inhibition studies of both the forward and reverse directions.

In addition to the kinetic studies, the solving of the crystal structure for the HGPRT enzyme is a prerequisite to the successful design of new drugs based on structure-function criteria. Currently a collaboration has been established with Dr. Dick Brennan at the Oregon Health Sciences University to crystalize and solve the protein structure. Once again, the ability to produce large quantities of recombinant protein should make the solving of the structure much easier. At the present time, initial "first screens" for potential crystallization conditions have been carried out with encouraging results.

To further our understanding of the parameters involved in the unique substrate specificity of the HGPRT enzyme from the trypanosomatids, an analysis of the potential purine binding site has been initiated. An alignment of each of the HGPRT's that have been isolated to date revealed several discrepancies in the purine binding site that may explain the differences in substrate specificity. Using site directed mutagenesis, potential residues in the purine binding site are being systematically altered to reflect the human HGPRT enzyme. In addition, an attempt is being made to mimic the substrate specificity of the malaria HGXPRT which is capable of phosphoribosylation of xanthine, a reaction that the trypanosomatid HGPRT's are incapable of catalyzing. It is also our intention to create mutations that mimic the defects found in patients who suffer from lesch-nyhan syndrome.

Advances in computer technology and our understanding of the dynamics of protein structure permits a rapid comparison of secondary structure predictions of proteins. Partial analysis of the secondary structure of the trypanosomatids HGPRT protein has already revealed several structures that are conserved amongst all of the HGPRT proteins even though the % identity at the amino acid level is very low. In particular a potential

dinucleotide-binding fold toward the amino terminus of the protein is seen in all HGPRT's examined to date. The conservation of structure even though the primary sequence is poorly conserved should not be surprising since changes that do not effect the tertiary structure of a protein would be allowed at the primary level. It is for this reason that it is difficult to predict qualities of a protein without a crystal structure to be a guide. However, if a crystal structure is available, the use of molecular modeling may prove useful in helping to make predictions as to the structure of a protein based on it's primary sequence and secondary structure predictions. At the present time several groups are in the process of collecting crystal structure data and it is our hope that this information will be available in the near future. Using a solved structure for an HGPRT protein, we can then begin to model the trypanosomatid HGPRT proteins using the solved structure as a backbone to build our model on. Although this method has not advanced to the point were it becomes unnecessary to solve the structure for our own proteins, the technique can allow for an early start into the exploration of the substrate binding characteristics and for development of novel inhibitors of the HGPRT enzyme in the absences of a solved protein structure.

To further explore the structure of the HGPRT enzyme, circular dichroism spectroscopy (CD) analysis is planned. CD analysis can predict the general secondary structure of proteins by measuring amide-amide interactions. This work will help in our analysis of secondary and tertiary structure based on our computer generated models and will give insight into how accurate these models are.

With the development of transfections systems for the trypanosomatids, a powerful tool became available to look at the dynamics of purine salvage in the parasite. Transfection

experiments will be carried out involving the deletion of the *hgprt* gene in the parasite. In addition, the recent cloning of the Adenine phosphoribosyltransferase (APRT) gene will allow for single and double knockouts thus allowing for analysis of the importance of these enzymes in purine metabolism. Even more interesting however, will be the ability to transfect *hgprt* sequences into the genome of the parasite that have been altered using site directed mutagenesis.

Another interesting question involving the parasites HGPRT's is their localization in the cell. Recent studies on the location of the malaria HGXPRT suggests that the enzyme may be associated with membrane bound compartments in merozoites. In the trypanosomatids, several studies have suggested that HGPRT activity is found localized in the glycosomes which are membrane bound organelles involved in glycolysis. It is also intriguing to note that the reported targeting sequence at the carboxyl terminus of proteins found localized to the glycosomes is found in all three HGPRT enzymes reported in this thesis. The localization of the HGPRT enzyme could have implications involving drug delivery and metabolism. At this time, recombinant HGPRT for both *T. cruzi* and *T. brucei* has been inoculated into rabbits and the first round of blood collection completed. Western blot analysis confirms that specific antibody is being produce to both recombinant enzymes and the purification of monospecific antibody will be completed in the near future.

Finally, with the onset of the AIDS epidemic, a desperate need is present for novel treatments to combat opportunistic pathogens of AIDS patients such as *Pneumocystis* and *Toxoplasma*. With the molecular tools now available to us we intend to isolate the *hgprt* gene from these organisms so that an attempt may be made at controlling these infections

through an understanding of their purine pathway.

Appendix A

Cloning and Overexpression of the Hypoxanthine-Guanine Phosphoribosyltransferase Gene from *Leishmania donovani*

by

Thomas Allen, Ho-Yon Hwang, Sheri Hanson, Armando Jordon,

and Buddy Ullman

To further characterize the enzyme HGPRT, an effort was made to clone the *Leishmania donovani hgprt* gene. Initial experiments using full length *hgprt* cDNA clones from the human, *S. mansoni*, *P. falciparum*, mouse, *T. brucei* and *T. cruzi* in the screening of *L. donovani* libraries proved unsuccessful. In addition, extensive use of degenerate oligonucleotides designed to conserved regions of the *hgprt* gene were used in PCR experiments on both DNA and RNA and in direct screenings of several different *L. donovani* libraries. This approach also proved unsuccessful. An attempt was then made to isolate the native HGPRT enzyme from the promastigote form of the parasite for amino acid terminus sequencing using the Edman degradation procedure. The purification scheme was essentially the same as that described for the *Trypanosoma* recombinant proteins except that 3 liters of confluent promastigotes were used and an additional wash step involving 10 column volumes of TMD buffer containing 50mM KCL buffer was added. Figure 1 shows the partially purified *L. donovani* HGPRT.

RNA was purified and cDNA was generated as previously described for the

trypanosomes. PCR was performed with the cDNA produced using a reverse compliment degenerate oligonucleotide designed to the protein sequence and a primer designed to the splice leader sequence. A single major product of 153 b.p. was observed and isolated for subcloning (figure 2). The cloned 153 b.p. fragment was then used to screen a *L. donovani* cosmid library at high stringency. Several clones were isolated and a sequence revealing an open reading frame of 632 b.p. encoding for a protein of approximately M_r = 23.6 Kd was observed (figure 3). The use of cDNA and PCR was a necessary approach since attempts to use the degenerate oligonucleotide designed from the protein sequence in the direct screening of libraries was unsuccessful. Although several clones were isolated using this more direct approach, upon analysis of the clones it was obvious that they did not encode the *L. donovani hgprt*.

Optimal sequence alignment of the *L. donovani* gene exhibited a 19% to 21% identity with its eukaryotic counterparts (figure 4). More interestingly, amongst the trypanosomatids it showed a 50% identity to *T. cruzi* while showing only a 34% identity to *T. brucei*. As with the trypanosomes, *L. donovani hgprt* showed a higher degree of identity to the prokaryotic HPT than to its eukaryotic counterparts. The phylogenetic relationships between the various HGPRT's will be discussed in further detail in appendix C.

Southern blot analysis revealed that the *L. donovani hgprt* is a single copy gene (Figure 5). Northern blot analysis revealed the presences of a single transcript at 1.5 Kb (figure 6). Since the cloning of the hgprt gene required the amplification of the 5' end of the mature message, no further experiments were required to determine the splice junction site for the splice leader sequence. The junction is marked by an # in figure 3.

The *L. donovani hgprt* was engineered using PCR to allow for cloning into the prokaryotic expression vector pBAce with the subsequent transformation of the vector into the *E.coli* stain S0606. HGPRT activity was observed in crude extracts of cells transformed with the *hgprt* construct whereas crude extracts of cells transformed with pBAce alone showed no activity (figure 7). In addition, substrate specificity for the recombinant enzyme was determined. As expected, hypoxanthine, guanine and allopurinol were all substrates for the *L. donovani* HGPRT while adenine and xanthine were not.

We feel that the work involving *L. donovani* was an important and integral part of our exploration of the enzyme HGPRT. The ability to examine kinetic parameters and to solve the HGPRT protein structures from this diverse group of organisms will allow for the identification of important amino acid residues and secondary structures motifs which should give us insights into the parameters that will need to be considered in successful drug design.

Figure 1-A. HGPRT purification from *L. donovani* promastigotes

Lane A - No additional wash step performed

Lane B - Additional 50mM KCL TMD wash step performed

Figure 2. cDNA PCR amplification

Lane A - 123 b.p. molecular weight markers

Lane B - 153 b.p. fragment

A B

Figure 3. Nucleotide and predicted amino acid sequence for *L. donovani hgprt*

TGCGTGCACCCATAAATATACATTTAAGTCCACATCACAGCGACTCAACGATCTTGCTCGATCCCTCCTCCCTCCCTTCTTC
CCTCTCTGTGAATCCTCGCTTACTCTCATGTGTATGGAAGCAACCATTATCCATCAGACCCCACCTCTCTTTTCCTT TACC
TTCACTATCCC

```
1   MET SER ASN SER ALA LYS SER PRO SER GLY PRO VAL GLY ASP GLU GLY ARG ARG ASN TYR 20
1   ATG AGC AAC TCG GCC AAG TCG CCC TCC GGC CCC GTC GGT GAT GAG GGG CGG CGC AAC TAT 60

21  PRO MET SER ALA HIS THR LEU VAL THR GLN GLU GLN VAL TRP ALA ALA THR ALA LYS CYS 40
61  CCG ATG TCA GCC CAC ACC CTC GTC ACA CAG GAG CAG GTG TGG GCC GCC ACG GCG AAG TGC 120

41  ALA LYS LYS ILE ALA GLU ASP TYR ARG SER PHE LYS LEU THR THR ASP ASN PRO LEU TYR 60
121 GCA AAG AAG ATT GCA GAG GAC TAC AGA AGT TTT AAG TTG ACG ACC GAC AAC CCG CTC TAC 180

61  LEU LEU CYS VAL LEU LYS GLY SER PHE ILE PHE THR ALA ASP LEU ALA ARG PHE LEU ALA 80
181 CTG CTG TGC GTG CTC AAG GGC AGC TTC ATC TTC ACG GCC GAC CTT GCC CGC TTT CTC GCC 240

81  ASP GLU GLY VAL PRO VAL LYS VAL GLU PHE ILE CYS ALA SER SER TYR GLY THR GLY VAL 100
241 GAC GAG GGT GTC CCG GTG AAG GTG GAG TTT ATT TGC GCG AGC TCG TAC GGC ACG GGC GTG 300

101 GLU THR SER GLY GLN VAL ARG MET LEU LEU ASP VAL ARG ASP SER VAL GLU ASN ARG HIS 120
301 GAG ACG TCG GGC CAG GTG CGC ATG CTC CTC GAC GTG CGT GAC TCC GTC GAG AAT CGC CAT 360

121 ILE LEU ILE VAL GLU ASP ILE VAL ASP SER ALA ILE THR LEU GLN TYR LEU MET ARG PHE 140
361 ATT CTG ATT GTC GAG GAC ATC GTC GAC AGC GCC ATC ACG CTG CAG TAT CTG ATG CGG TTC 420

141 MET LEU ALA LYS LYS PRO ALA SER LEU LYS THR VAL VAL LEU LEU ASP LYS PRO SER GLY 160
421 ATG CTC GCC AAG AAG CCG GCC TCG CTC AAG ACG GTG GTG CTG CTG GAC AAG CCG TCG GGG 480

161 ARG LYS VAL GLU VAL LEU VAL ASP TYR PRO VAL ILE THR ILE PRO HIS ALA PHE VAL ILE 180
481 CGA AAG GTG GAG GTG CTA GTC GAC TAC CCC GTC ATC ACG ATC CCG CAC GCG TTT GTG ATT 540

181 GLY TYR GLY MET ASP TYR PRO GLU SER TYR ARG GLU LEU ARG ASP ILE CYS VAL LEU LYS 200
541 GGC TAC GGC ATG GAC TAC CCC GAG TCG TAT CGA GAG CTG CGC GAT ATC TGC GTG CTC AAG 600

201 LYS GLU TYR TYR GLU LYS PRO GLU SER LYS VAL                                    211
601 AAG GAG TAC TAC GAG AAG CCG GAG AGC AAG GTG                                    633
```

TAGCGGTGACGAGCTATACCGCTCGTGTCGGTGGGAACACCTGCCCGCCTCTCTCCTTCTCTATCNCNCNCTCTCACAGAAAC
GCACACCGACATGCCAACAAGCGTGCTCGTGGGCGATGGAAGGGGTGAGACCGCCGTAGSACTGCGCTGCGTACTCAAGTGGA
GCCGACCATACCACNCNCNGCCTTTTTCTTTTCTTCGTTCTCTAACTTCTTCTTACCCATTT

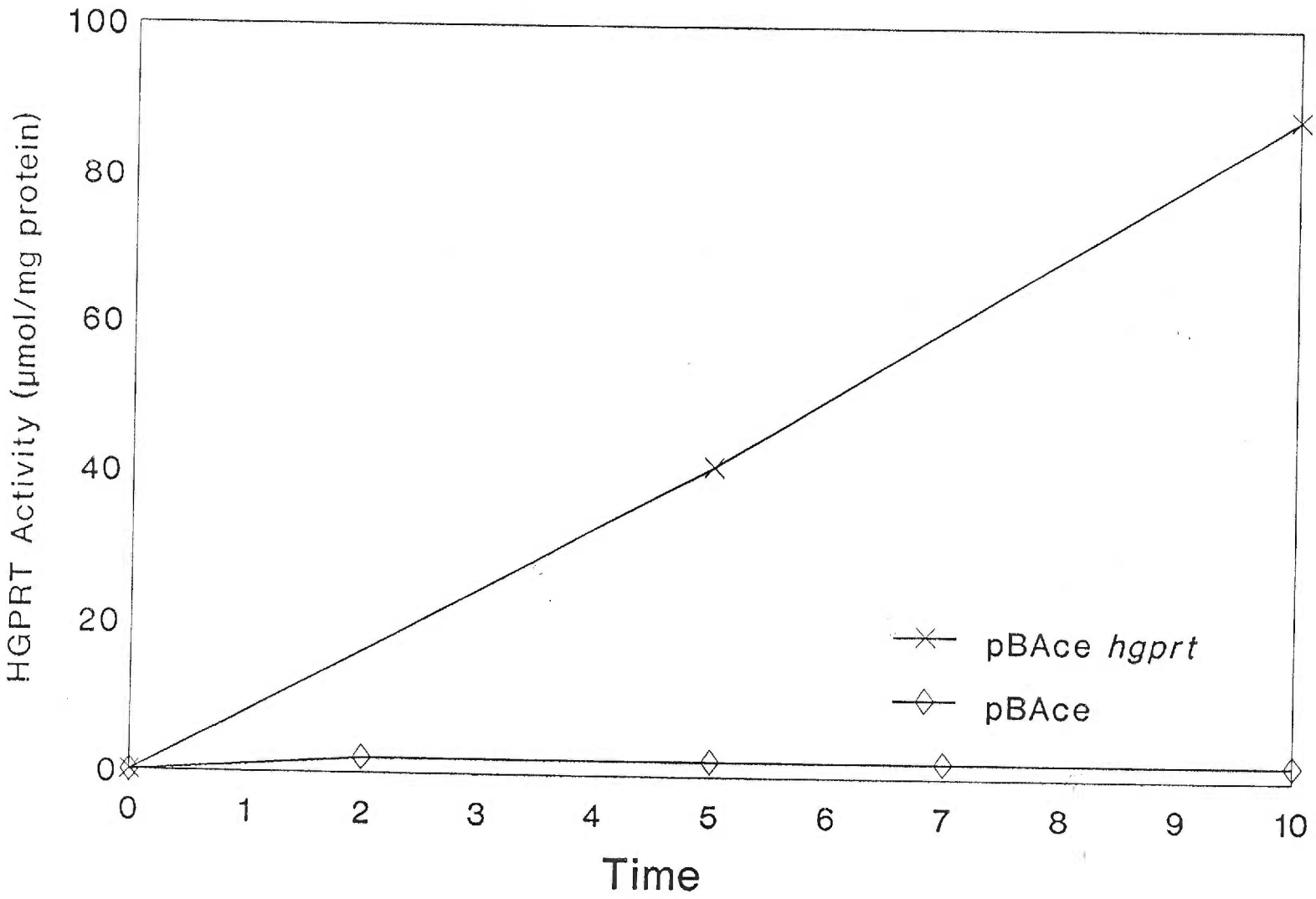
Figure 4. Alignment of the amino acid sequences of *L. donovani* with other HGPRTs

```
HUMAN           MAT---RSPG-----VVISDDEPGYDLDLFCIPNHYAEDLERVFIPHGLI
S.mansoni       MSSNMIKADC-----VVIEDSFRGFPTEYFCTSPRYDECLDYVLIPNGMI
P.falciparum    MPIPNNPGAGENAFDPVFVKDDDGYDLDSFMIPAHYKKYLTKVLVPNGVI
T.brucei        MEPACK-----------------YDFA-----------TSVLFTEAEL
T.cruzi         MP--------------------REYEFA-----------EKILFTEEEI
L.donovani      MSNSAKSPSG-----PVGDEGRRNYPMS-----------AHTLVTQEQV
V.harveyi       MKHT-----------------------------------VEVMISEQEV
                *                                          . .

                                                 Purine Site
HUMAN           MDRTERLARDVMKEMGGHH-------IVALCVLKGGYKFFADLLDYIKAL
S.mansoni       KDRLEKMSMDIVDYYEACN----ATSITLMCVLKGGFKFLADLVDGLERT
P.falciparum    KNRIEKLAYDIKKVY---N----NEEFHILCLLKGSRGFFTALLKHLSRI
T.brucei        HTRMRGVAQRIADDYSNCNLKPLENPLVIVSVLKGSFVFTADMVRILG--
T.cruzi         RTRIMEVAKRIADDYKGKGLRPYVNPLVLISVLKGSFMFTADLCRALS--
L.donovani      WAATAKCAKKIAEDYRSFKLTT-DNPLYLLCVLKGSFIFTADLARFLA--
V.harveyi       QERIRELGKQITERYQG------SEDLVMVGLLRGSFVFMADLARAI---
                      . .                . . .*.*.  * . .   .

HUMAN           NRNSDRSIPMTV---DFIRLKSYCNDQSTGDIKVIGGDDLSTLTGKNVLI
S.mansoni       VRARGIVLPMSV---EFVRVKSYVNDVSIHEPILTGLGDPSEYKDKNVLV
P.falciparum    HNYSAVEMSKPLFGEHYVRVKSYCNDQSTGTLEIVS-EDLSCLKGKHVLI
T.brucei        ----DFGVPTRV---EFLRASSYGHDTKSCGRVDVKADGLCDIRGKHVLV
T.cruzi         ----DFNVPVRM---EFICVSSYGEGVTSSGQVRMLLDTRHSIEGHHVLI
L.donovani      ----DEGVPVKV---EFICASSYGTGVETSGQVRMLLDVRDSVENRHILI
V.harveyi       ------ELTHQV---DFMTASSTGNTMESSRDVRILKDLDDDIKGKDVLI
                      .   .   ...   *                    ..*.
                Prim 1
HUMAN           VEDIIDTGKTMQTLLSLVRQYNPKMVKVASLLVKRTPRSVGYKPDFVGFE
S.mansoni       VEDIIDTGKTITKLISHLDSLSTKSVKVASLLVKRTSPRNDYRPDFVGFE
P.falciparum    VEDIIDTGKTLVKFCEYLKKFEIKTVAIACLFIKRTPLWNGFKADFVGFS
T.brucei        LEDILDTALTLREVVDSLKKSEPASIKTLVAIDKPGGRKIPFTAEYVVAD
T.cruzi         VEDIVDTALTLNYLYHMYFTRRPASLKTVVLLDKREGRRVPFSADYVVAN
L.donovani      VEDIVDSAITLQYLMRFMLAKKPASLKTVVLLDKPSGRKVEVLVDYPVIT
V.harveyi       VEDIIDTGNTLNKIREILSLREPKSIAICTLLDKPSRREVEVPVDYVGFA
                .***.*.. *.             .     . *         ..
                   Prim II
HUMAN           IPDKFVVGYALDYNEYFRDLNHVCVI---------SETGKAKYKA-----
S.mansoni       VPNRFVVGYALDYNDNFRDLHHICVI---------NEVGQKKFSVPCTSK
P.falciparum    IPDHFVVGYSLDYNEIFRDLDHCCLV---------NDEGKKKYK----AT
T.brucei        VPNVFVVGYGLDYDQSYREVRDVVILKPSVYETWGKELERRK-------A
T.cruzi         IPNAFVIGYGLDYDDTYRELRDIVVLRPEVYAEREAARQKKQRAIGSADT
L.donovani      IPHAFVIGYGMDYPESYRELRDICVLKKEYYEKPE--------------
V.harveyi       IPDEFVVGVGIDYAQKYRDLPFIGKVVPQE-------------------
                .*. **.* ..** . .*..     .

HUMAN           ------------
S.mansoni       P----------V
P.falciparum    S----------L
T.brucei        AGEAKR------
T.cruzi         DRDAKREFHSKY
L.donovani      ---------SKV
V.harveyi       ------------
```

Figure 5. Southern blot analysis of *L. donovani hgprt*

Figure 6. Northern blot analysis of *L. donovani* promastigote RNA

Figure 7. HGPRT expression in *E. coli*


100
80
60
40
20
0
HGPRT Activity (µmol/mg protein)
0
1
2
3
4
5
6
7
8
9
10
Time
pBAce hgprt
pBAce

Appendix B

Quaternary Structure, Stability, and pH Optimization Analysis of the *Trypanosoma* HGPRT

by

Thomas Allen and Buddy Ullman

Since the goal of producing recombinant HGPRT protein is to be able to perform detailed mechanistic and structural studies, several physical parameters of the HGPRT enzyme were examined. These parameters included a study of the pH at which the HGPRT enzyme exhibits the greatest amount of activity and an exploration of the stability of the stored protein. In addition, an attempt was made to ascertain the quaternary structure of the active protein. It is our hope that an understanding of these parameters will aid in our ongoing investigation into the nature of the phosphoribosylation reaction that the HGPRT protein catalyzes.

Experimental evidence involving the quaternary structure of the HGPRT active protein has been both confusing and contradictory. The quaternary structure of a protein refers to the number and spatial arrangements of individual subunits. To determine the apparent molecular weight and the number of subunits making up the quaternary structure of the HGPRT enzyme, high pressure liquid chromatography (HPLC) was employed using

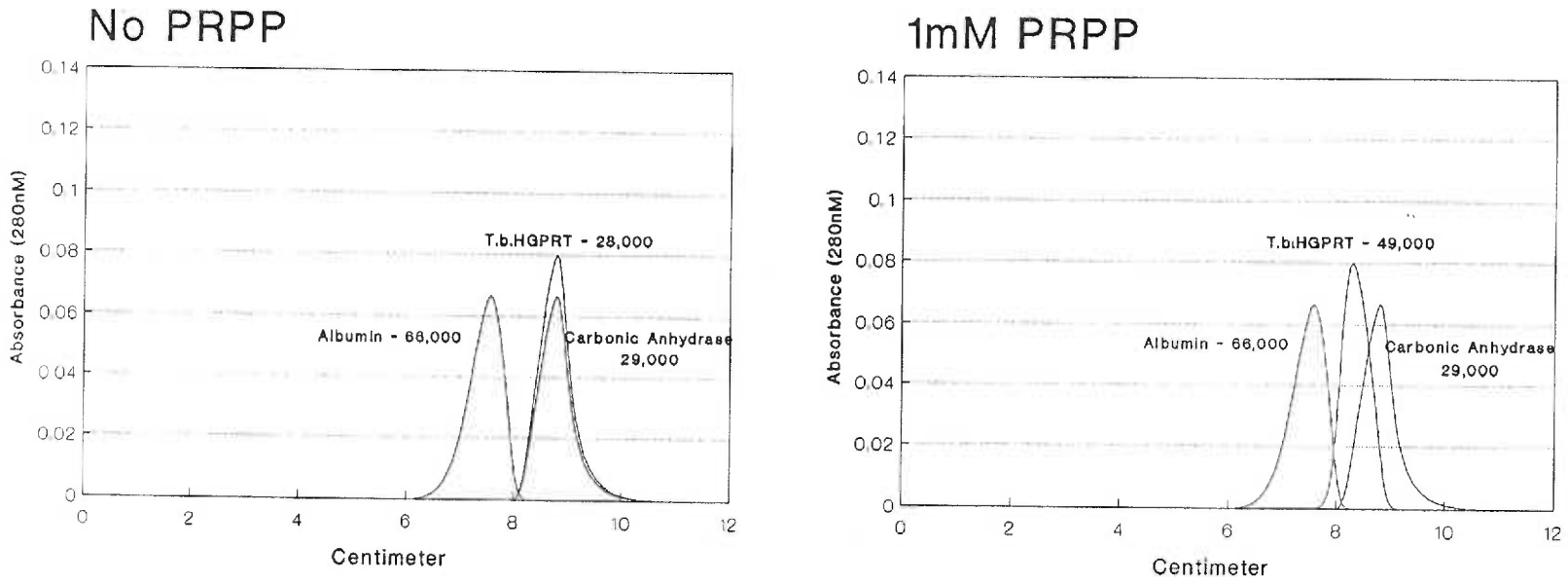
two different gel filtration column. The first column was capable of separating proteins between 10,000 to 300,000 in molecular weight while the second was capable of separating proteins in the 2,000-80,000 range. Freshly purified recombinant HGPRT from both *T. brucei* and *T.cruzi* were run through the filtration column which had been equilibrated in the buffer TMD (20mM Tris,pH7.5, 5mM $MgCl_2$, 1mM DTT) in the presence or absence of 5-phospho-1-pyrophosphate (PRPP) and fractions were collected. When HGPRT was run through the column without PRPP being added to the buffer, a single peak of approximately M_r = 28 kd for *T. brucei* and M_r = 29 kd for *T. cruzi* was observed. In the presence of PRPP however, the molecular weight for *T. brucei* and for *T. cruzi* was approximately M_r = 50kd. Since the calculated molecular weights for the *T. brucei* and *T. cruzi* are M_r = 23.4 kd and M_r = 26 kd respectively, this suggests that in the presences of PRPP the HGPRT enzyme is in a dimeric conformation. The results suggest that HGPRT is found in a dimer conformation only when in the presence of PRPP and since the enzyme is only active in the presences of PRPP, the dimer conformation must be the active form of the enzyme (figure 1).

Studies were also conducted to determine the pH at which the enzymes exhibit the greatest capacity to phosphoribosylate [^{14}C] hypoxanthine. The reaction buffer TMD described above was modified to produce a range of pH levels from pH 4 to pH 10. Samples of freshly purified recombinant HGPRT were diluted in the individual buffers and then added to prewarmed reaction cocktails at the same pH as the enzyme. Levels of phosphoribosylation were recorded and are shown in figure 2. These experiments suggest that the recombinant enzyme for *T. brucei* demonstrated a pH optimum at 8.5 whereas the

T. cruzi enzyme showed an optimization of pH 8.0. It is interesting that these values are more basic then for other HGPRT enzymes.

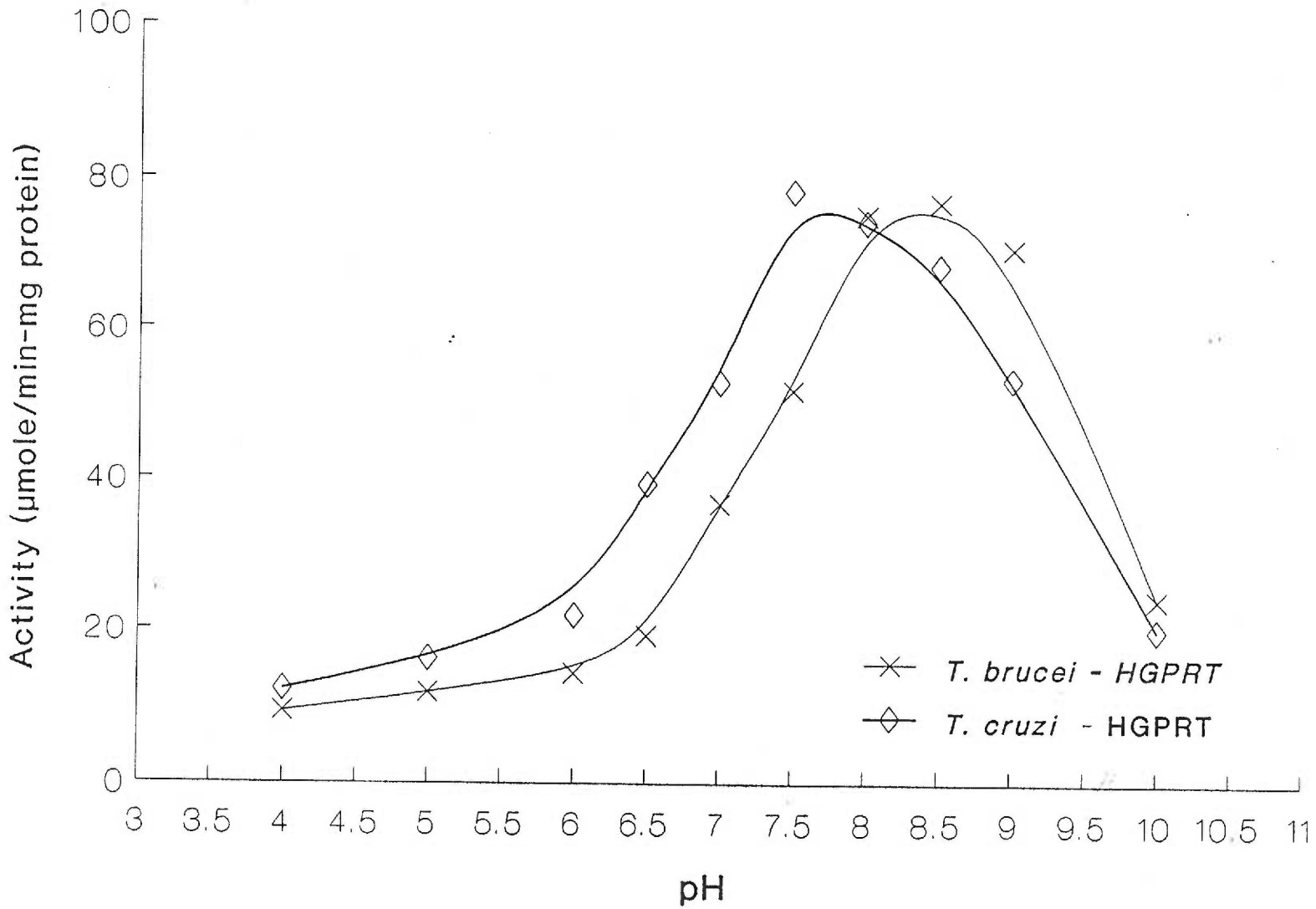
Finally, since it would be beneficial to use previously purified protein for continued experimental analysis, the stability of the recombinant protein was explored. A detailed experiment involving the recombinant *T. cruzi* HGPRT was conducted while a smaller study involving the *T. brucei* protein was also perused. For the *T. cruzi* protein, samples of purified enzyme in the presences of 1mM PRPP were stored at temperatures ranging from 4° to -70° C. while in the presence or absence of 50% glycerol. The activity of the enzyme was then measured over a six month period. The results of this study showed that the recombinant *T. cruzi* HGPRT enzyme is very stable at all temperatures tested and in the presence or absence of glycerol. It has been observed however that the purified protein when stored at very high concentrations (greater then 5mg/ml) can precipitate out of solution over time suggesting that a more dilute concentration of the enzyme may increase the ability of the protein to be stored. The *T. brucei* enzyme also showed an excellent ability to remain active for long periods of time. As of the writing of this thesis, recombinant enzyme has been stored at -70°C for twelve months without any loss of activity. The ability to store large quantities of recombinant enzyme will enhance our ability to carry out detailed mechanistic studies and protein crystallization experiments.

Figure 1. Quaternary and apparent molecular weight determination for *T. brucei* HGPRT (*T. cruzi* HGPRT is virtually identical)



Molecular Weight Determination

Figure 2. pH Optimization


Activity (µmole/min-mg protein)
100
80
60
40
20
0
3
3.5
4
4.5
5
5.5
6
6.5
7
7.5
8
8.5
9
9.5
10
10.5
11
pH
T. brucei - HGPRT
T. cruzi - HGPRT

Appendix C

The use of the enzyme HGPRT in a Phylogenetic Study

by

Thomas E. Allen and Buddy Ullman

The ever increasing speed whereby genes are identified, cloned, sequenced, and the proteins they encode analyzed is beginning to permit a detailed comparison of the relationships between different organisms. This is the study of phylogenies which concerns itself with the inference of the evolutionary history and relationships based on the divergence of genes and the products they encode (1). At the present time several excellent studies have been carried out to examine the relationship between the Trypanosomatidae family and a variety of eukaryotic and prokaryotic organisms (2,3,4). In particular, the use of small-subunit rRNA genes have given us insight into the relationship between organisms even at the deepest and earliest levels. However, few studies have been performed using genes that are widespread and indispensable to the survival of both eukaryotic and prokaryotic organisms. It is true that enzymes involved in the glycolytic pathway have been used in several studies, however the validity of using mitochondrial genes in studies involving widely divergent organisms is in doubt (2). For this reason and to further the understanding of the relationship between members of the trypanosomatidae family to each other,

phylogenetic analysis was performed on the enzyme hypoxanthine-guanine phosphoribosyltransferase (HGPRT). Two separate methods were used to estimate the relationship of nine species to each other using HGPRT as a indicator protein, the parsimony method for protein sequence and the distance matrix method for both protein and DNA sequences (5).

The parsimony or compatibility method involves producing a unrooted tree using an algorithm that weighs changes in protein sequence based on the number of substitutions that are necessary at the nucleotide level. In this way a single residue change that requires a single step at the DNA level is not treated equally as a change that requires two or more substitutions. This has the advantage over early versions of parsimony calculations in that it allows for a more sensitive determination of the distances between species. This method also assumes that a change in one residue that is found in the same place in another organisms requires that the two organisms be on the same branch of the tree. The difficulty in analysis occurs when all the changes are examined and inconsistences arise. The algorithm then builds a tree with the least amount of discrepancies.

The first step in producing a phylogeny using the parsimony method is to create a multiple sequence alignment (6). Figure 1 shows the result of the alignment between the nine different organisms at the protein level. Parsimony calculations were then performed using a random number generator to determine the order of input into the program. These calculations were performed ten separate times to assure consistency in the results. In addition, calculations were performed assuming the prokaryotic HPT sequence as an outgroup and also with no outgroup being designated. Figure 2 shows the trees that were

supported by the protein data. In all calculations, the same tree was produced along with the same number of cycle needed to produce it whether an outgroup was designated or not. These results suggests that this is indeed the most relevant tree for the data set. The designation of an outgroup makes the assumption that this is the root of the tree. In these calculations, the designation of an outgroup did not change the structure of the tree thus lending confidence to the results.

There are many different distance matrix algorithms available. The one used in this analysis was introduced by Saitou and Nei and is known as the neighbor joining method (7). This algorithm uses a distance matrix table to build its trees. In addition to using this algorithm to look at the relationship between the nine different organisms that were used in the protein parsimony calculations, the relationship between only the trypanosomatids was also examined at the DNA level.

As with the protein parsimony programs it is necessary to first align the sequence using a multiple sequence alignment program. The neighbor joining method then calculates the percent divergence between each species to build a matrix distance table. For calculations at the DNA level, a different "weight" is given for transition (A - G, C - T) than for transversions (A - C/T). Phylogenetic trees are then constructed based on the matrix table with the tree that minimizes the discrepancy between the observed distances and expected distances chosen as the best tree. Figures 3 shows the results of these calculations for the protein level.

A resampling method, known as 'bootstrapping", was used to analyze the calculations made using the neighbor joining method (8). Bootstrapping involves a random sampling of

positions in the initial alignment process with the calculation of a distance matrix with the subsequent production of a tree based on the sampling. If the data strongly supports the tree calculated using the entire sequence, the tree from the sampling will be the same. This method is especially useful for branches that are weakly supported by the data. In this analysis, a minimum of 1000 independent sampling were performed to look at the confidence levels for the trees calculated. Only the branches used in the final tree (figure 4) were supported by the resampling calculations. However, the branch that divides the trypanosomatids themselves was relatively weakly supported by the protein data. To further examine the relationship between the members of the trypanosomatidae family, neighbor-joining calculations were performed using the DNA sequence as an indicator of relativeness. Figure 4 shows the alignment of the DNA sequences while figure 5 shows the tree that was supported by the data. Although still rather weak, the DNA calculations also support a closer relationship between *T. cruzi* and *L. donovani* then between *T. brucei* and *L. donovani*.

The results of this study supports previous phylogenetic analysis that suggest that trypanosomatids are found on some of the earliest branchings among eukaryotes (9). In addition, evidence suggesting a closer relationship between *L. donovani* and *T. cruzi* than between *T. brucei* and *L. donovani* are also supported by this data (10). The lack of robustness in the calculations however regarding the branching order of the trypanosomatids is still not clearly supported by this data set. Further analysis will need to be performed to give insight into the relationship between this diverse group of organisms.

Figure 1. **Alignment of protein sequences**

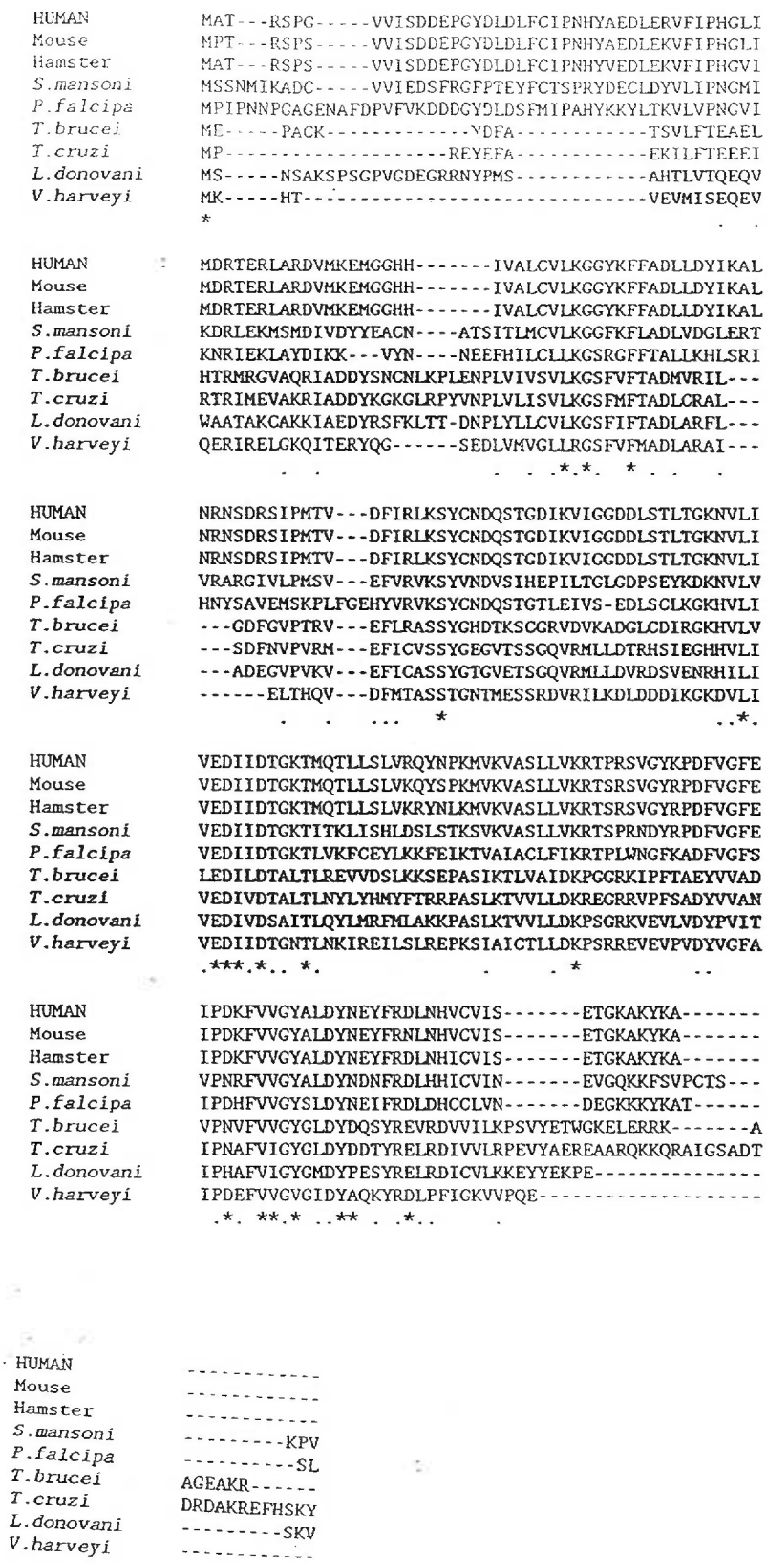
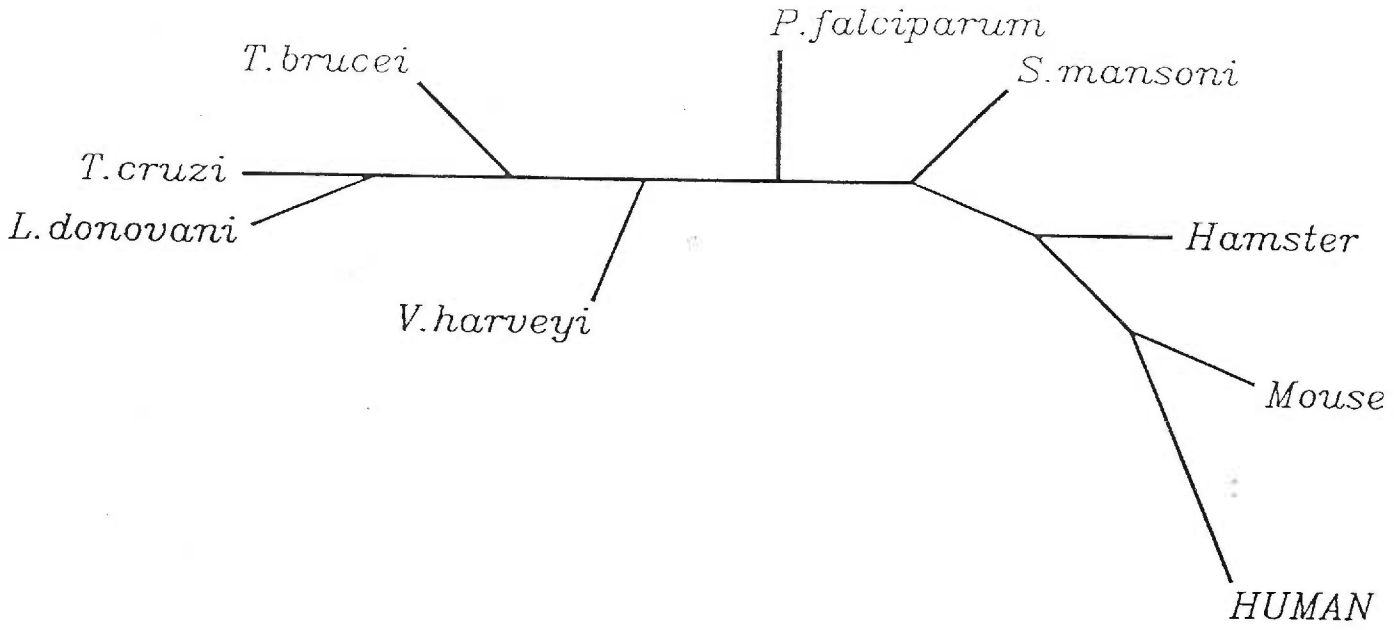

HUMAN MAT---RSPG-----VVISDDEPGYDLDLFCIPNHYAEDLERVFIPHGLI
Mouse MPT---RSPS-----VVISDDEPGYDLDLFCIPNHYAEDLEKVFIPHGLI
Hamster MAT---RSPS-----VVISDDEPGYDLDLFCIPNHYVEDLEKVFIPHGVI
S.mansoni MSSNMIKADC-----VVIEDSFRGFPTEYFCTSPRYDECLDYVLIPNGMI
P.falcipa MPIPNNPGAGENAFDPVFVKDDDGYDLDSFMIPAHYKKYLTKVLVPNGVI
T.brucei ME-----PACK----------YDFA----------TSVLFTEAEL
T.cruzi MP----------REYEFA----------EKILFTEEEI
L.donovani MS-----NSAKSPSGPVGDEGRRNYPMS----------AHTLVTQEQV
V.harveyi MK-----HT----------VEVMISEQEV
HUMAN MDRTERLARDVMKEMGGHH-------IVALCVLKGGYKFFADLLDYIKAL
Mouse MDRTERLARDVMKEMGGHH-------IVALCVLKGGYKFFADLLDYIKAL
Hamster MDRTERLARDVMKEMGGHH-------IVALCVLKGGYKFFADLLDYIKAL
S.mansoni KDRLEKMSMDIVDYYEACN----ATSITLMCVLKGGFKFLADLVDGLERT
P.falcipa KNRIEKLAYDIKK---VYN----NEEFHILCLLKGSRGFFTALLKHLSRI
T.brucei HTRMRGVAQRIADDYSNCNLKPLENPLVIVSVLKGSFVFTADMVRIL---
T.cruzi RTRIMEVAKRIADDYKGKGLRPYVNPLVLISVLKGSFMFTADLCRAL---
L.donovani WAATAKCAKKIAEDYRSFKLTT-DNPLYLLCVLKGSFIFTADLARFL---
V.harveyi QERIRELGKQITERYQG------SEDLVMVGLLRGSFVFMADLARAI---
HUMAN NRNSDRSIPMTV---DFIRLKSYCNDQSTGDIKVIGGDDLSTLTGKNVLI
Mouse NRNSDRSIPMTV---DFIRLKSYCNDQSTGDIKVIGGDDLSTLTGKNVLI
Hamster NRNSDRSIPMTV---DFIRLKSYCNDQSTGDIKVIGGDDLSTLTGKNVLI
S.mansoni VRARGIVLPMSV---EFVRVKSYVNDVSIHEPILTGLGDPSEYKDKNVLV
P.falcipa HNYSAVEMSKPLFGEHYVRVKSYCNDQSTGTLEIVS-EDLSCLKGKHVLI
T.brucei ---GDFGVPTRV---EFLRASSYGHDTKSCGRVDVKADGLCDIRGKHVLV
T.cruzi ---SDFNVPVRM---EFICVSSYGEGVTSSGQVRMLLDTRHSIEGHHVLI
L.donovani ---ADEGVPVKV---EFICASSYGTGVETSGQVRMLLDVRDSVENRHILI
V.harveyi ------ELTHQV---DFMTASSTGNTMESSRDVRILKDLDDDIKGKDVLI
HUMAN VEDIIDTGKTMQTLLSLVRQYNPKMVKVASLLVKRTPRSVGYKPDFVGFE
Mouse VEDIIDTGKTMQTLLSLVKQYSPKMVKVASLLVKRTSRSVGYRPDFVGFE
Hamster VEDIIDTGKTMQTLLSLVKRYNLKMVKVASLLVKRTSRSVGYRPDFVGFE
S.mansoni VEDIIDTGKTITKLISHLDSLSTKSVKVASLLVKRTSPRNDYRPDFVGFE
P.falcipa VEDIIDTGKTLVKFCEYLKKFEIKTVAIACLFIKRTPLWNGFKADFVGFS
T.brucei LEDILDTALTLREVVDSLKKSEPASIKTLVAIDKPGGRKIPFTAEYVVAD
T.cruzi VEDIVDTALTLNYLYHMYFTRRPASLKTVVLLDKREGRRVPFSADYVVAN
L.donovani VEDIVDSAITLQYLMRFMLAKKPASLKTVVLLDKPSGRKVEVLVDYPVIT
V.harveyi VEDIIDTGNTLNKIREILSLREPKSIAICTLLDKPSRREVEVPVDYVGFA
HUMAN IPDKFVVGYALDYNEYFRDLNHVCVIS-------ETGKAKYKA-------
Mouse IPDKFVVGYALDYNEYFRNLNHVCVIS-------ETGKAKYKA-------
Hamster IPDKFVVGYALDYNEYFRDLNHICVIS-------ETGKAKYKA-------
S.mansoni VPNRFVVGYALDYNDNFRDLHHICVIN-------EVGQKKFSVPCTS---
P.falcipa IPDHFVVGYSLDYNEIFRDLDHCCLVN-------DEGKKKYKAT------
T.brucei VPNVFVVGYGLDYDQSYREVRDVVILKPSVYETWGKELERRK-------A
T.cruzi IPNAFVIGYGLDYDDTYRELRDIVVLRPEVYAEREAARQKKQRAIGSADT
L.donovani IPHAFVIGYGMDYPESYRELRDICVLKKEYYEKPE----------
V.harveyi IPDEFVVGVGIDYAQKYRDLPFIGKVVPQE----------
HUMAN ----------
Mouse ----------
Hamster ----------
S.mansoni ---------KPV
P.falcipa ----------SL
T.brucei AGEAKR------
T.cruzi DRDAKREFHSKY
L.donovani ---------SKV
V.harveyi ----------

Figure 2. **Parsimony trees**

a. No outgroup

b. Outgroup (*V. harveyi*)

A


P.falciparum
T.brucei
S.mansoni
T.cruzi
L.donovani
Hamster
V.harveyi
Mouse
HUMAN

B

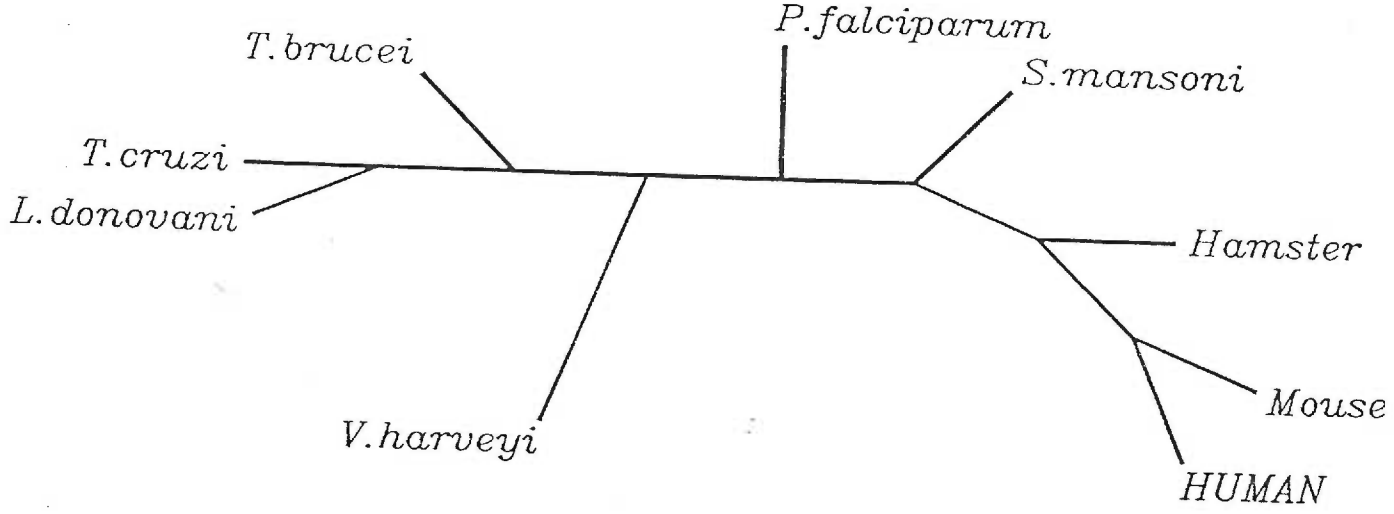

P.falciparum
T.brucei
S.mansoni
T.cruzi
L.donovani
Hamster
Mouse
V.harveyi
HUMAN

Figure 3. **Neighbor Joining Tree** (protein)

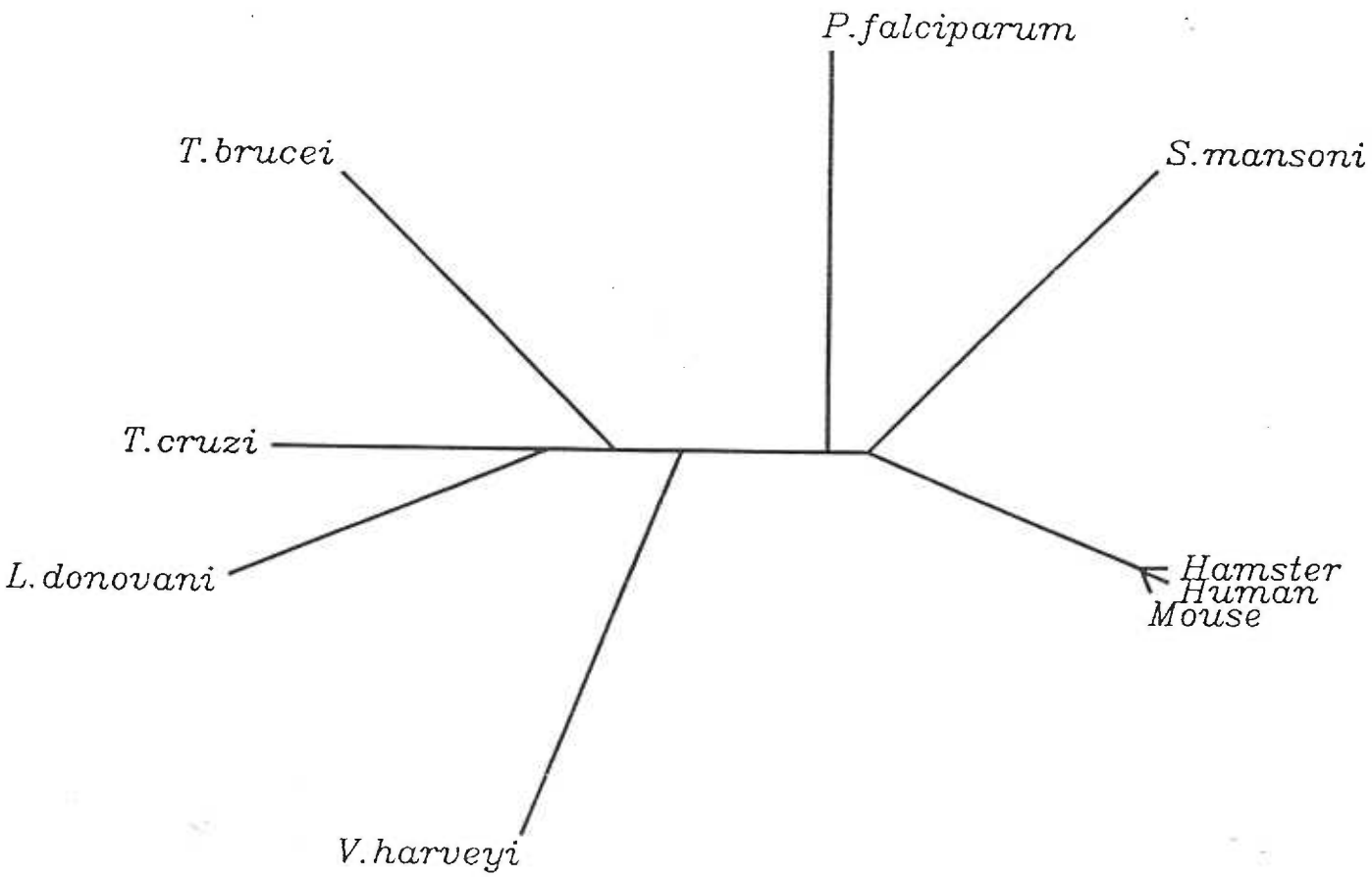

P.falciparum
T.brucei
S.mansoni
T.cruzi
L.donovani
Hamster
Human
Mouse
V.harveyi

Figure 4. **Alignment of DNA sequences**

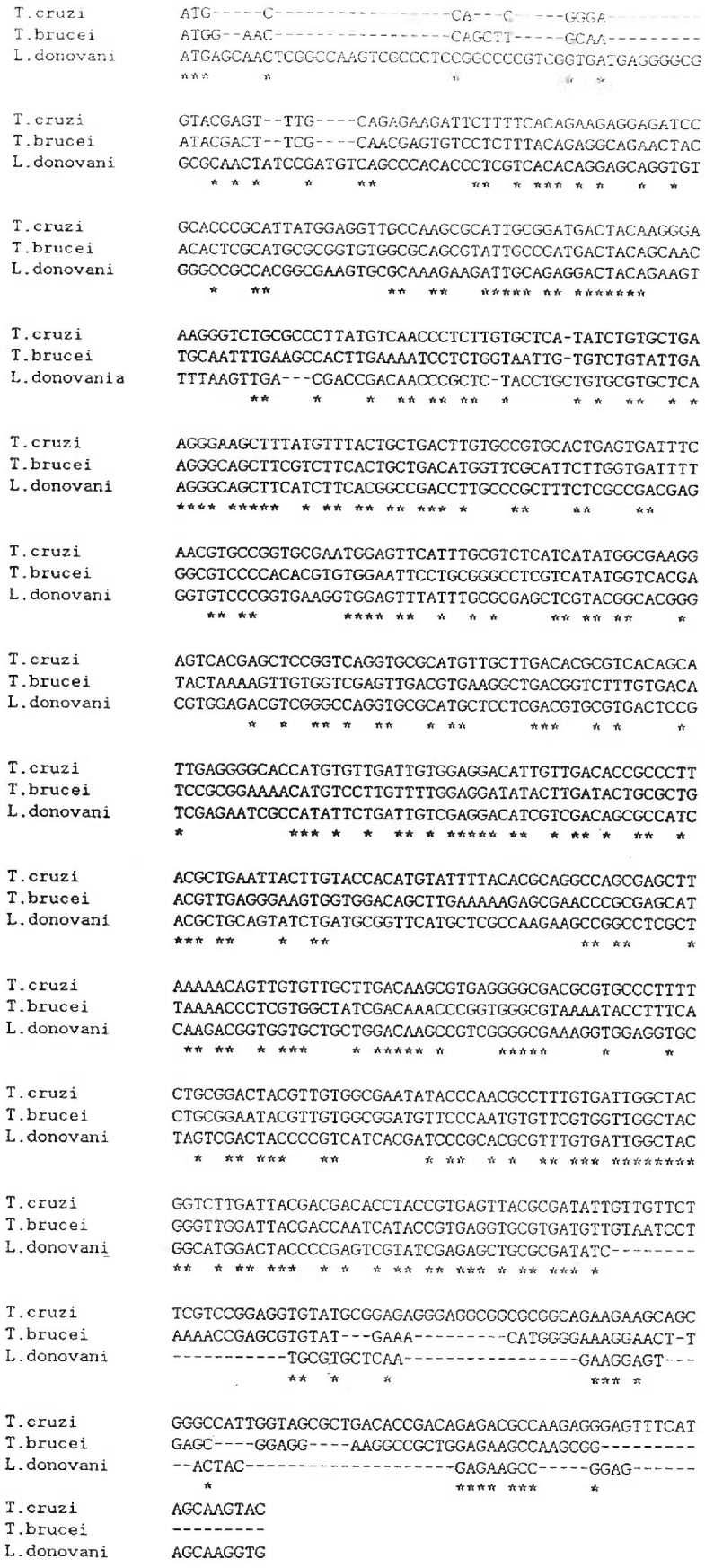
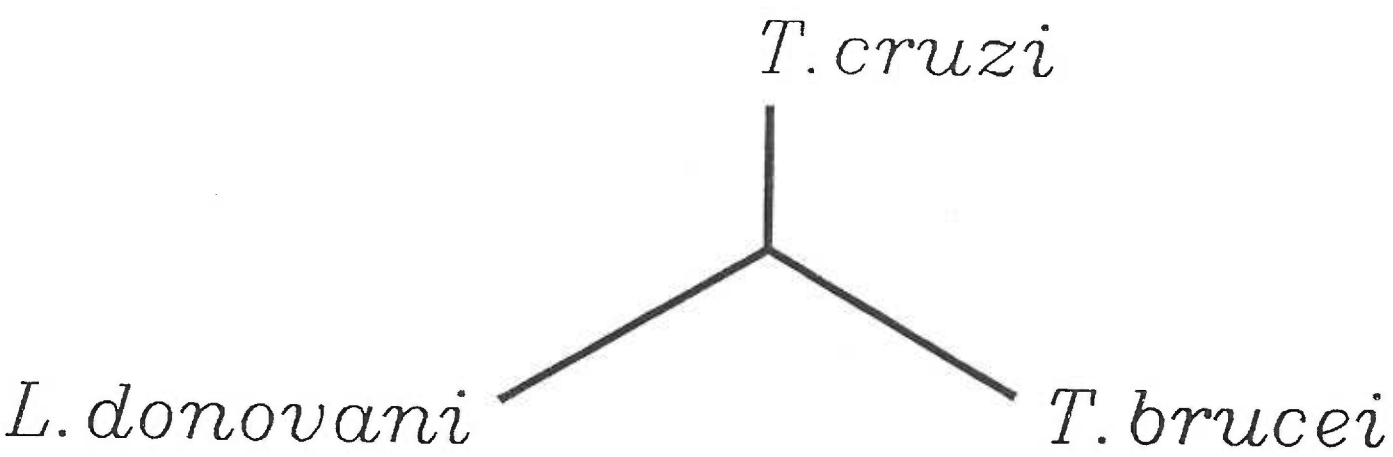

T.cruzi
T.brucei
L.donovani
T.cruzi
T.brucei
L.donovani
T.cruzi
T.brucei
L.donovani
T.cruzi
T.brucei
L.donovania
T.cruzi
T.brucei
L.donovani
T.cruzi
T.brucei
L.donovani
T.cruzi
T.brucei
L.donovani
T.cruzi
T.brucei
L.donovani
T.cruzi
T.brucei
L.donovani
T.cruzi
T.brucei
L.donovani
T.cruzi
T.brucei
L.donovani
T.cruzi
T.brucei
L.donovani
T.cruzi
T.brucei
L.donovani

Figure 5. **Neighbor Joining Tree (DNA)**


T. cruzi
L. donovani
T. brucei

References

1. Felsenstein, J. (1988) Annu.Rev.Genet. 22, 521-565

2. Sogin, M.L., Elwood, H.J. & Gunderson, J.H. (1986) Proc.Natl.Acad.Sci.USA 83, 1383-1387

3. Lake, J.A., de la Cruz, V.F., Ferreira, P.C.G., Morel, C. & Simpson, L. (1988) Proc.Natl.Acad.Sci.USA 85, 4779-4783

4. Sogin, M.L. & Gunderson, J.H. (1993) Annal.New York Acad.Sci. 125-139

5. Analysis of phylogenetic relationships were all performed using the computer programs contained within **PHYLIP** (Phylogeny Inference Package) version 3.4

6. Multiple sequence alignment was accoplished with the computer program, CLUSTAL version V

7. Saitou,N. & Nei,M. (1987) Mol.Bio.Evol. 4, 406-425

8. Felstein,J. (1985) Evolution 39,783-791

9. Michels, P.A.M. (1986) J.Mol.Evol. 24, 45-52

10. Briones, M.R.S., Nelson, K., Beverly, S.M., Affonso, H.T., Camargo, E.P. & Floeter-Winter, L.M. (1992) Mol.Biochem.Parasitol. 53, 121-128

11. Hernandez, R., Rios, P., Valdes, M. & Pinero, D. (1990) Mol.Biochem.Parasitol. 41, 207-212

12. Gomez, E., Valdez, A.M., Pinero, D. & Hernandez, R. (1991) J.Mol.Evol. 8, 254-259