The EMBO Journal vol.9 no.9 pp.2751 – 2758, 1990

# *Trypanosoma cruzi* glycosomal glyceraldehyde-3-phosphate dehydrogenase does not conform to the 'hotspot' topogenic signal model

**Giles Kendall[1,3], Andrew F.Wilderspin[2], Frank Ashall[1,4], Michael A.Miles[1] and John M.Kelly[1,5]**

[1]Department of Medical Parasitology, London School of Hygiene and Tropical Medicine, Keppel Street (Gower Street), London WC1E 7HT, UK and [2]Laboratory of Molecular Biology, Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, UK

[3]Present address: International Institute of Cellular and Molecular Pathology, Research Unit for Tropical Diseases, Ave Hippocrate 74, Brussels B-1200, Belgium

[4]Present address: Department of Pure and Applied Biology, Imperial College of Science and Technology, London SW7 2BB, UK

[5]To whom correspondence should be addressed

Communicated by P.Borst

**The genes which encode glycosomal glyceraldehyde-3-phosphate dehydrogenase (gGAPDH) of *Trypanosoma cruzi* are arranged as a tandemly repeated pair on a single chromosome and are identical at the level of nucleotide sequence. They are separated by an intergenic region which contains a 317 base pair sequence with the properties of a retroposon. The genes express a 1.5 kb mRNA and a 38 kd protein. The amino acid sequence contains features characteristic of glycosomal enzymes such as peptide insertions and a C-terminal extension. However, *T.cruzi* gGAPDH lacks one of the positively charged 'hotspot' motifs which have been proposed as topogenic signals for import into the glycosome, a unique microbody-like organelle. Molecular modelling of the *T.cruzi* and *T.brucei* enzymes suggests that neither structure would fulfil the requirements of the 'hotspot' glycosomal import model.**

*Key words:* glyceraldehyde-3-phosphate dehydrogenase/glycosome/retroposon/*Trypanosoma cruzi*

## Introduction

The kinetoplastid protozoan parasite *Trypanosoma cruzi* is the causative agent of Chagas' disease, an acute and/or chronic infection which afflicts up to 20 million people in Central and South America (Miles, 1983). There are no effective vaccines (Brener, 1986) and treatment with drugs currently available is unsatisfactory (Miles, 1983). The development of more effective chemotherapeutic agents is urgently required.

One potential target for chemotherapy against trypanosomatid parasites is the glycosome, a microbody-like organelle unique to the kinetoplastids (Opperdoes, 1987). Several housekeeping enzymes are sequestered within the glycosomes including the first seven enzymes of glycolysis (reviewed by Michels, 1989). The glycosomal glycolytic enzymes which have been most intensely studied are those of the African trypanosome, *Trypanosoma brucei*, where the bloodstream form of the organism is totally dependent upon glycolysis for ATP production (reviewed by Fairlamb and Opperdoes, 1986). *T.brucei* glycosomal enzymes have peptide insertions in their sequence and/or terminal extensions which distinguish them from their cytosolic counterparts and which are absent from homologous enzymes in other organisms studied. These differences may provide the basis for a rational approach to drug design (Michels, 1989).

Glycosomal enzymes are encoded by nuclear DNA (Opperdoes, 1987) and are synthesized on free ribosomes in the cytosol. Import into the glycosome appears to be a rapid post-translational event and does not involve protein modification or peptide cleavage (Hart *et al.*, 1987; Clayton, 1987). By inference the topogenic signal resides in the sequence or structure of the mature enzyme. The elucidation of the translocation signal has proven difficult due to problems associated with developing a convincing *in vitro* import assay or an *in vivo* expression system (Borst, 1989). Analysis has instead relied upon comparison of glycosomal enzyme sequences with those of cytosolic isoenzymes that are in some cases also present in the parasite or with cytosolic homologues of other organisms (reviewed by Borst, 1986b, 1989; Wierenga *et al.*, 1987).

A study of four glycolytic enzymes from *T.brucei* led to the proposal that the presence of two clusters of positively charged amino acids located on their surface and separated by ~40 Å may be important for glycosome translocation (Wierenga *et al.*, 1987). These have been called 'hotspots' and share a common motif with respect to charged amino acids of positive-$(neutral)_{1-2}$-positive. One hotspot was always present in a unique peptide insertion in the sequence of each glycosomal enzyme. Hotspots were absent at the corresponding positions in cytosolic isoenzymes and cytosolic homologues of other organisms (Wierenga *et al.*, 1987). Evidence from *Crithidia fasiculata* phosphoglycerate kinase (PGK) genes has argued against this model (Swinkels *et al.*, 1988) and it has alternatively been suggested that hotspots may be involved in intraglycosomal substrate interactions (Opperdoes, 1988; Michels, 1989). To investigate common features amongst trypanosome glycosomal enzymes which may be involved in protein translocation, we have carried out the first structural study of a glycosomal enzyme (gGAPDH) from the South American trypanosome, *T.cruzi*. The results for *T.cruzi* gGAPDH are not consistent with the hotspot model of glycosomal localization. We also report the discovery of a retroposon-like element (RLE) in the intergenic region of the tandemly repeated gGAPDH genes.

## Results

### *Organization of T.cruzi gGAPDH genes*

A *T.cruzi* genomic DNA library (strain X10/6) in λEMBL 3 was screened with a cDNA probe derived from *T.brucei*

gGAPDH (Materials and methods) and a clone containing a 16.5 kb insert was isolated. Restriction mapping and DNA sequence analysis (below) indicated the presence of two tandemly repeated gGAPDH genes within this insert (Figure 1). Southern analysis of genomic DNA revealed the absence of other gGAPDH related sequences or of allelic variants in the genome, although the presence of identical alleles could not be excluded. A single chromosomal location (0.85 megabase pairs) of the gGAPDH genes was determined for the X10/6 strain after contour clamped homogeneous electric field (CHEF) separation and hybridization (data not shown).

### *Sequence of T.cruzi gGAPDH*

The sequences of the two gGAPDH genes (1080 nucleotides) are identical (Figure 2). Similarity between the first 115 nucleotides upstream of the inferred start codon of each of the *T.cruzi* genes is 100%. Further upstream of this the sequences are unrelated. Downstream of each gene the region of similarity stretches 350 nucleotides from the stop codon with only occasional point differences (Figure 2). The *T.cruzi* genes are 81% similar to those of *T.brucei*.

The inferred amino acid sequence of *T.cruzi* gGAPDH reveals a protein with features characteristic of a glycosomal enzyme, i.e. the presence of distinctive peptide insertions and of a five amino acid C-terminal extension (Figure 3). These features are also found in the *T.brucei* gGAPDH enzyme sequence (Michels *et al.*, 1986). The *T.cruzi* gGAPDH amino acid sequence is 90% similar to that of *T.brucei* gGAPDH but only 54% similar to the cytosolically located human GAPDH (Figure 3).

One of the regions designated as a positively charged hotspot (Wierenga *et al.*, 1987) in *T.brucei* gGAPDH (HSII, $_{222}$KRASK) is conserved in *T.cruzi*, but the other (HSI) is not. The positive-neutral-positive amino acid charge motif in *T.brucei* ($_{66}$KSK) is found to be KSS in *T.cruzi* (Figure 3). Differences, sequence or structural, between the *T.cruzi* and *T.brucei* enzymes presumably identify amino acids which are not important for glycosomal localization assuming a common import mechanism. On this basis we suggest that HSI as designated in *T.brucei* does not play an essential role in mediating glycosomal uptake.

A conjugated peptide of identical sequence to the final 10 C-terminal amino acids of the *T.cruzi* enzyme gave rise to antibodies in mice (Materials and methods) which recognized a major protein of ~38 kd on Western blots (Figure 4A). This equates to the predicted subunit molecular weight of the *T.cruzi* enzyme. Since the synthetic peptide corresponds to a region of the enzyme that is not highly conserved between *T.cruzi* and *T.brucei* (Figure 3), the recognition of gGAPDH is specific for the *T.cruzi* enzyme (Figure 4A).

In common with other published *T.cruzi* gene sequences (Dragon *et al.*, 1987; Maignon *et al.*, 1988; Requena *et al.*, 1988; Swindle *et al.*, 1988) there is a strong bias towards C and G residues in the third positions of codons in the gGAPDH genes. This contrasts with the situation in *T.brucei* where the bias is only towards C residues in various housekeeping genes, including gGAPDH (Michels, 1986).

### *Identification of a retroposon-like element*

A DNA sequence with the properties of a retroposon was identified in the intergenic region of the *T.cruzi* gGAPDH genes (Figure 2). It is 317 bp long, lies in the opposite orientation to the gGAPDH genes, has a polyadenylated region at its putative 3'-end (24 out of 25 A residues) and is flanked by short direct repeat sequences, including a match at nine consecutive nucleotides (underlined, Figure 2). These features are thought to arise by reverse transcription of mRNA into DNA and its subsequent integration into the genome (reviewed by Wiener *et al.*, 1986). Such elements may resemble fully processed retropseudogenes and have a 3'-poly(A) tract. The sequence 5'-ATAATTTC which occurs adjacent to the polyadenylated tract of the retroposon-like element (RLE) described here is similar to one found in a corresponding position of the RIME element of *T.brucei* (5'-AAAATTTC; Hasan *et al.*, 1984).

The appearance of a RLE in the *T.cruzi* intergenic region has occurred since species divergence since no comparable sequences are present around the gGAPDH genes of *T.brucei* (P.A.M.Michels, personal communication). An open reading frame was observed which extended from close to the 5'-end of the RLE to beyond the polyadenylated region. The RLE did not appear to be transcribed independently of the gGAPDH genes since signals were not detected on Northern blots or in S1 nuclease protection assays (below) using appropriate single stranded DNA probes (probe b$^1$, Figure 5A). The element was repeated in truncated forms in the flanking regions of the gene repeat (Figure 1).

When *T.cruzi* genomic DNA (strain X10/6) was analysed by Southern blotting using low stringency conditions and a probe derived largely from the intergenic region, additional bands were observed which were not detected when a probe with sequences confined to the gGAPDH gene was used (data not shown). This suggests a wider distribution of sequences related to the RLE in the genome of X10/6 strain, but at a low copy number.

### *Transcript mapping*

A band of about 1.5 kb was detected on Northern blots of *T.cruzi* epimastigote RNA hybridized with gGAPDH probes (Figure 4B). Accumulation of multicistronic transcripts due to heat shock induced disruption of the machinery of RNA processing has been observed in trypanosomes (e.g. tubulin

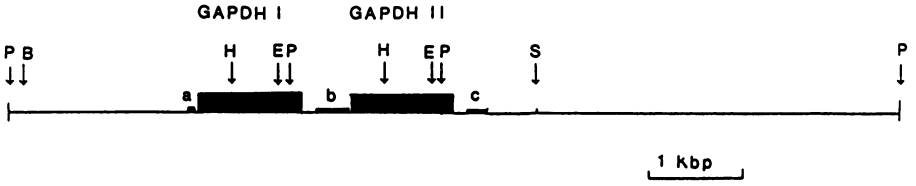


**Fig. 1.** Genomic organization of *T.cruzi* gGAPDH genes. Restriction map of the region of *T.cruzi* genomic DNA containing the gGAPDH genes (I and II). The small bars correspond to the RLE (**b**) or truncated forms thereof (**a,c**). B, *Bam*HI; D, *Dra*I; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; S, *Sal*I.

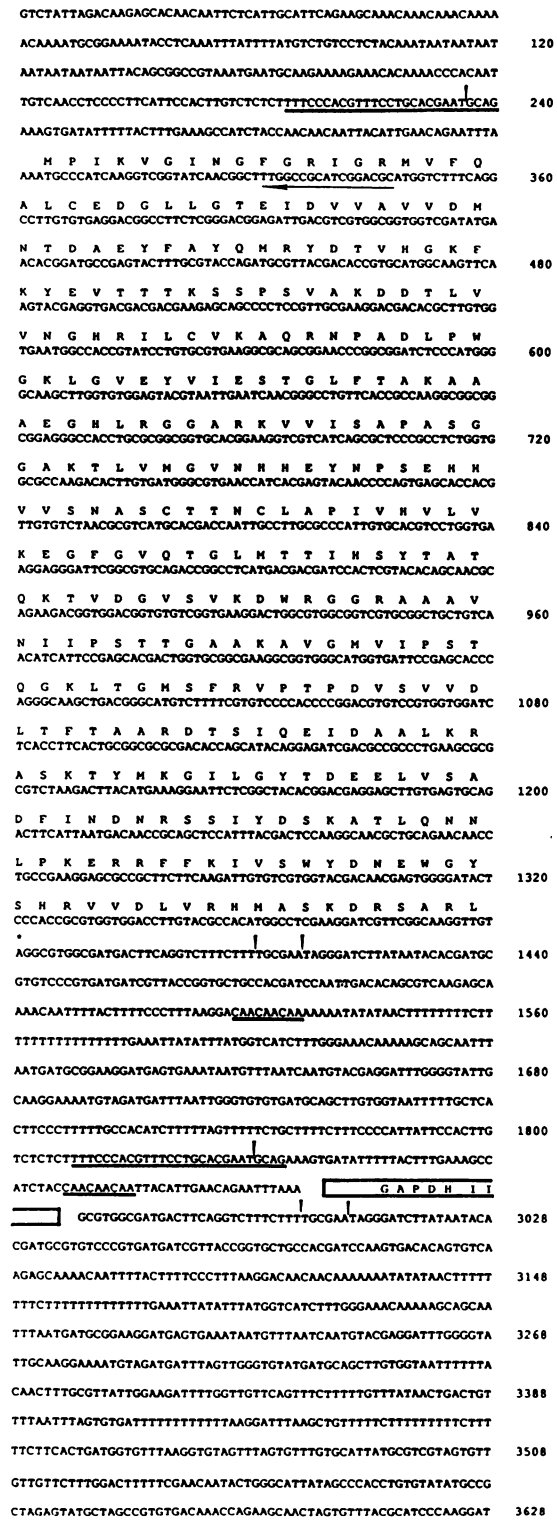

**Fig. 2.** Sequence of *T.cruzi* gGAPDH genes. The nucleotide and amino acid sequences are shown for gGAPDH I. Those of gene II are identical. The mini-exon acceptor site consensus sequence (Laird *et al.*, 1987) upstream of each coding region is underlined. Arrows denote points of mini-exon addition upstream of each gene, or points of poly(A) addition downstream of each gene, as determined by S1 nuclease protection assays (Figure 5). In the intergenic region, the direct nucleotide repeats flanking the RLE (5'-CAACAACAA) are also underlined. The synthetic oligonucleotide used in the production of probe a (Figure 5A) is indicated by a horizontal arrow (nucleotides 347–330). This sequence has been given the EMBL database accession number X52898.

transcripts of *T.brucei*; Muhich and Boothroyd, 1988). No such additional higher molecular weight bands were detected on autoradiographs of Northern blots containing RNA isolated from epimastigotes cultured at 37°C and 42°C (data not shown).

S1 nuclease protection assays were used to delineate the regions of genomic DNA which encode gGAPDH mRNA (Materials and methods, Figure 5). The 5'-end of the mRNA coding regions [in the Kinetoplastida this corresponds to the site of addition of the mini-exon sequence (for review see Ralph *et al.*, 1988)] were determined from the sizes of radiolabelled protected fragments derived from probes a (GAPDH I) and b (GAPDH II) (Figure 5A). Using the conditions described (Materials and methods, Figure 5B) a single protected fragment was detected on polyacrylamide gels (Figure 5B) indicative of a site of splice leader addition 66 nucleotides upstream (denoted by arrows, Figure 2) of the inferred start codon. It is not possible at this stage to determine whether this fragment is derived from the transcript of one or both genes.

The 3'-ends of the genes were determined using probe c (Figure 5A). On polyacrylamide gels (Figure 5C), protected fragments corresponding to a 3'-untranslated region of 34 and 28 nucleotides could be detected (denoted by arrows in Figure 2). As a result of the procedure by which probe c was prepared (Materials and methods), the unlabelled complementary strand was co-purified from a polyacrylamide gel. The presence of this fragment does not affect interpretation of the data but rehybridization with probe c can occur and results in the appearance of the fully protected probe fragment (817 nucleotides, Figure 5C). Sequences similar to the eukaryotic polyadenylation signal (AATAAA) were not observed adjacent to the 3'-ends of the genes.

### *Molecular modelling of T.cruzi gGAPDH*

The three-dimensional structures of *T.cruzi* and *T.brucei* gGAPDH were modelled using the computer program COMPOSER (Birkbeck College, London) and the X-ray crystallographic co-ordinates from GAPDH of human (3.5 Å) (Tso *et al.*, 1985), lobster (2.9 Å) (Davidson *et al.*, 1967) and *Bacillus stearothermophilus* (1.8 Å) (Skarzynski *et al.*, 1987). The technique requires at least 50% amino acid homology to give reliable models. Model building constructs a core enzyme using the known X-ray coordinates as a reference and incorporating as much of the unknown enzyme sequence as possible (73.5% in the case of the trypanosome gGAPDHs). Regions that cannot be directly modelled (identified, Figure 3) are then joined by loops drawn from the Brookhaven database of all known protein structures (Sutcliffe *et al.*, 1987a,b).

In the resulting structure for *T.cruzi* gGAPDH (Figure 6), the region corresponding to HSI was separated from HSII by ~51 Å. Moreover, in the *T.brucei* model, the separation between HSI and HSII was also ~51 Å. These distances are too great to fit the proposed model for glycosomal uptake and also to permit binding of the anti-trypanosomal drug suramin as speculated by Wierenga *et al.* (1987). Consistent with this, 'anti-hotspot' molecules having two negatively charged groups separated by ~40 Å (similar to the suramin structure) were unable to suppress significantly *T.brucei* gGAPDH activity in *in vitro* assays (Opperdoes *et al.*, 1990).

The distances between hotspots were determined using only the C-alpha atom co-ordinates since precise placement

```
     **                  *************    **                  **
Tc   PIKVGINGFGRIGRMVFQALCEDGLLGTEIDVVAVVDMNTDAEYFAYQMRYDTVHGKFKY   60
Tb   T.....................D.....N.............R......K..S......H   60
Bs   AV............N..R.ALKNP     D.E....N. L.N.DGL.HLLK..S...RLDA   55
Hu    V...V........L.TR.AFNS.     KV.I..IN.PFI.LN.MV.MFQ..ST....HG   57
Lo    S.I..D.......L.LR.ALSC.      AQ....N.PFIAL..MV.MFK..ST..V..G   54

     *   *************                          ***            **
Tc   EVTTTKSSPSVAKDDTLVVNGHRILCVKAQRNPADLPWGKLGVEYVIESTGLFTAKAAAE  120
Tb   S.S....K.....................................................V.S...  120
Bs   ..VVN         .GDVS...KE.IVKA E...EN.A..EI..DI.V....R..KRED.A  106
Hu   T.KAE         NGK..I..NP.TIFQ E.D.SKIK..DA.A...V....V..TMEK.G  106
Lo   ..KME         .GA...D.KK.TVFN EMK.ENI..S.A.A..IV....V..TIEK.S  105

          **           ***       ***     **
Tc   GHLRGGARKVVISAPASGGAKTLVMGVNHHEYNPSEHHVVSNASCTTNCLAPIVHVLVKE  180
Tb   ......................F......NN...R.Q..............L.......  180
Bs   K..EA..K..I.....KVENI.V.....QDK.D.KA...I............FAK..HQE  166
Hu   A..Q...KR.I....S AD.PMF......EK.D NSLKII............LAK.IHDN  166
Lo   A.FK...K.......S AD.PMF.C...LEK.S KDMT..............VAK..HEN  163

                     ***   ****************       ******
Tc   GFGVQTGLMTTIHSYTATQKTVDGVSVKDWRGGRAAAVNIIPSTTGAAKAVGMVIPSTQG  240
Tb   ...IS......V.........................L.....................  240
Bs    ..IVR.M...V....NN.RIL.LP.H..L..A....ES...T........AL.L.ELK.  225
Hu    ..IVE.....V.AI.........P.G.L..D..G.LQ....AS........K...ELN.  225
Lo    .EIVE.....V.AV.........P.A.......G..Q.....S........K...ELD.  222

                                                            **   *
Tc   KLTGMSFRVPTPDVSVVDLTFTAARDTSIQEIDAALKRASKTYMKGILGYTDEELVSADF  300
Tb   .....A.....A.........I.T.....K..............N..............  300
Bs   ..N..AM.....N.......VAELEKEVTVE.VN....A.AEGEL....A.SE.P...RNY  285
Hu   .....A.....AN.......CRLEKPAKYDD.KKVV.Q..EGPL.......EHQV..S..  285
Lo   .....A..............VRLGKECSYDD.K..M.T..EGPLQ.F....EDDV..S..  282

        **       ***********                                ******
Tc   INDNRSSIYDSKATLQNNLPKERRFFKIVSWYDNEWGYSHRVVDLVRHMASKDRSARL    358
Tb   .S.S.................N.............................AR..A.K.    358
Bs   NGSTV..TI.ALS.MVIDGK     MV.V.......T.........AAYINA.GL        335
Hu   NS.TH..TF.AG.GIAL.DH     .V.LI......F...N.....MA...ARE         334
Lo   .G......F.A..GIQLSKT     .V.V.......F...Q..I..LK..QKV.SE        333
```

**Fig. 3.** Optical alignments of GAPDH enzymes. Dots denote identity with the *T.cruzi* gGAPDH sequence. Residues are numbered starting after the amino terminal L-Met. Loop regions determined from molecular modelling are superscribed by asterisks. Tc, *T.cruzi*; Tb, *T.brucei*; Bs, *B.stearothermophilus*; Hu, human; Lo, Lobster.

of side chains in such models is invariably sub-optimal. Attempts were made to reduce hotspot separation in the *T.brucei* model by remodelling the region containing HSI (Thr66–Leu78) which corresponds to a loop structure selected automatically by the COMPOSER program (Materials and methods). No satisfactory alternative loop structure could be found.

## Discussion

### *Gene structure and expression*

The gGAPDH genes of *T.cruzi* are organized as a pair of direct tandem repeats of identical nucleotide sequence similar to the arrangement in *T.brucei* (Michels *et al.*, 1986). It can therefore be inferred that duplication of the gGAPDH genes predates the divergence of these trypanosomatid species and that mechanisms operate to maintain sequence identity within each species.

Gene coversion has been invoked as one means by which sequence identity of tandemly repeated genes could be maintained in trypanosomes (e.g. gGAPDH, Michels *et al.*, 1986; aldolase, Marchand *et al.*, 1988; PGK, Le Blancq *et al.*, 1988) and as a mechanism by which variant surface glycoprotein (VSG) genes can be transferred to their active sites (see Borst, 1986a). In the PGK genes of *T.brucei*, the

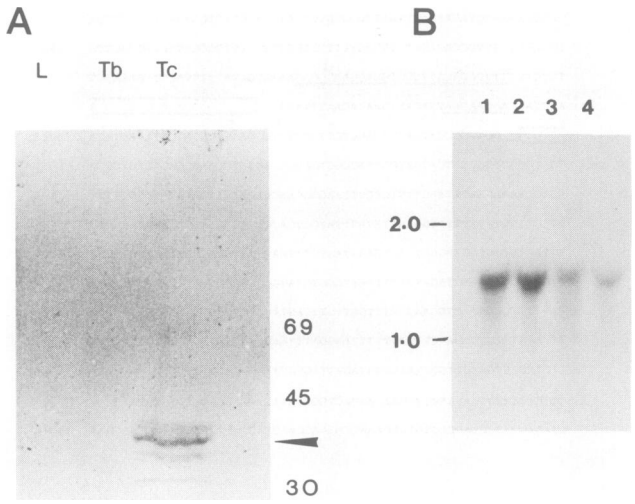


**Fig. 4.** Expression of *T.cruzi* gGAPDH. (**A**) Western blot of trypanosomatid soluble antigen tested with antibodies against the C-terminal 10 amino acids of *T.cruzi* gGAPDH (Materials and methods). Antibody (1:100) was applied for 48 h at 4°C and visualized using anti-mouse antibody conjugated to peroxidase at 1:1000 dilution. Molecular sizes are shown in kilodaltons (kd). Tb, *T.brucei*; Tc, *T.cruzi*; L, *L.donovani*. (**B**) Northern blot of total RNA from *T.cruzi* strains hybridized with a radiolabelled gGAPDH M13 probe. Tracks (1) X10/6 strain; (2) CAN III strain; (3) 92:80 strain; (4) Y strain. Molecular sizes are in kilobases (kb).

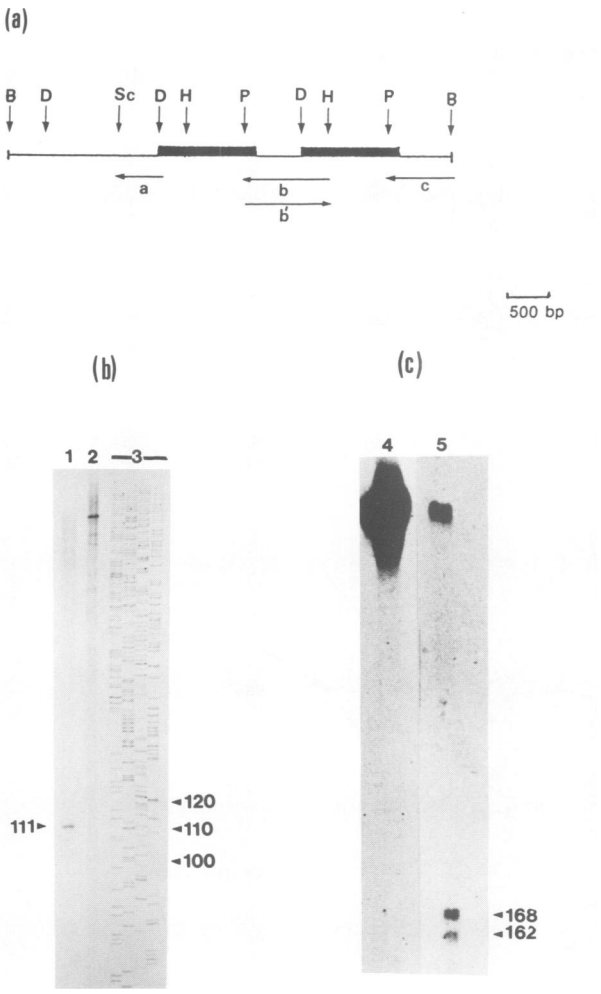


**Fig. 5.** S1 nuclease mapping. (**a**) Restriction map of the 5.0 kb *Bam*HI fragment containing both gGAPDH genes showing location of probes used in S1 nuclease protection assays. Sc, *Sac*II. The probes were single stranded, radiolabelled chain-extension products of M13 subclones. The arrow denotes the direction of DNA synthesis during labelling. (**b**) Localization of the 5′-end of *T.cruzi* (X10/6 strain) gGAPDH mRNA. 25 μg of total RNA was hybridized (75% formamide/42°C) with single stranded radiolabelled probe a (Materials and methods) then treated with S1 nuclease (4°C/4 h). Tracks: (1) probe a, treated; (2) probe a, untreated; (3) Size markers. (**c**) Localization of the 3′-end. Hybridization conditions were as above. Tracks: (4) probe c, untreated; (5) probe c, treated.

presence of sequences similar to the chi sequences of phage lambda (5′-GCTGGTGG), which is known to function as a recombinational hotspot, is suggested as a possible mediator of the gene conversion process (Le Blancq *et al.*, 1988). Chi-like sequences have been identified as being associated with regions which undergo gene conversion in several eukaryotic systems, but their precise role remains to be determined (reviewed by Thaler and Stahl, 1988). In *T.cruzi*, at least three chi-like sites are present in or near the gGAPDH genes. Of these, one occurs in the RLE (Figure 3, 5′-GCTTGTGG; 177 nucleotides upstream of the GAPDH II start codon). The sequence located in the RLE occurs close to the point at which divergence between the 3′-flanking regions of the genes begins. Chi-like sites have been observed near to the 3′-end of the tandemly repeated *T.brucei* gGAPDH (Le Blancq *et al.*, 1988) and the aldolase genes (Marchand *et al.*, 1988). In *T.cruzi* gGAPDH the RLE could also play a role in recombination by mediating heteroduplex formation and subsequent mismatch repair leading to gene homogenization. Whatever the precise mechanism, the result is that the sequences of the mRNA coding regions of both genes are identical. In the immediate flanking regions some point mutations have occurred. Beyond this the sequences diverge completely.

Transcript mapping indicated that the site of mini-exon addition did not occur immediately after an AG dinucleotide in the genomic DNA sequence (Figures 2 and 5B). Variations in sequence adjacent to the site of mini-exon addition have been observed in the dihydrofolate reductase-thymidylate synthetase genes of *Leishmania major* (Kapler *et al.*, 1987), the calmodulin genes of *T.brucei* (Tschudi *et al.*, 1985) and in VSG genes of *Trypanosoma equiperdum* (Layden and Eisen, 1988) and *T.brucei* (Scholler *et al.*, 1988). It has been proposed that base-pairing between the mini-exon donor (med) RNA and the primary transcript influence exactly which splice site is preferred by stabilizing interactions with the med RNA (Layden and Eisen, 1988). The sequence immediately upstream of the 5′-mini-exon addition site of the *T.cruzi* gGAPDH mRNA transcript (−66 nucleotides) is consistent with the consensus trypanosome mini-exon acceptor sites proposed by Laird *et al.* (1987), i.e. TTCPy occurring in one or more copies followed by a T-rich region. The precise definition of factors which govern selection of the sites of mini-exon addition cannot be addressed directly at present.

### *Glycosomal import*

The data presented here suggest that import of *T.cruzi* gGAPDH is not mediated by two positively charged hotspots at its surface. In addition, it can be implied that if two hotspots are involved in glycosomal import of *T.brucei* gGAPDH, then their separation by 40 Å is not essential. If a common mechanism has been conserved during evolution, then glycosomal import must be mediated by structural features other than hotspots. It has been inferred from comparisons of the cytosolic and glycosomal forms of phosphoglycerate kinase (PGK) in *C.fasciculata* that the topogenic signal for glycosome uptake of this enzyme probably resides in the C-terminal extension (Swinkels *et al.*, 1988). In the *T.cruzi* gGAPDH model (Figure 6), the five amino acid C-terminal extension forms a helix which has a high hydrophobic moment ($\mu H = 4.32$, Eisenberg *et al.*, 1984; Von Heijne, 1986) and which extends from the main body of the enzyme. Such structures are known to be important in the uptake of some mitochondrial proteins (Von Heijne, 1986). In the *T.brucei* model (Figure 6), the C-terminal region forms a similar structure despite a high clustering of amino acid substitutions (Figure 3) (four out of eight residues differ; all conservative substitutions).

Peroxisomes are microbodies that may be closely related to glycosomes (Borst, 1989). The C-terminal tripeptide of the peroxisomal enzyme firefly luciferase (SKL*) probably functions as a minimal signal for translocation of proteins across the peroxisomal membranes of species as diverse as plants, yeast, insects and mammals (Gould *et al.*, 1989). The terminal tripeptide of *T.brucei* gGAPDH (AKL*) and *T.brucei* glucosephosphate isomerase (SHL*) (Marchand *et al.*, 1989) can replace the corresponding sequence of firefly luciferase and still permit peroxisomal import (Gould *et al.*, 1989). Other published glycosomal enzyme sequences did not have C-terminal tripeptides which will permit import in

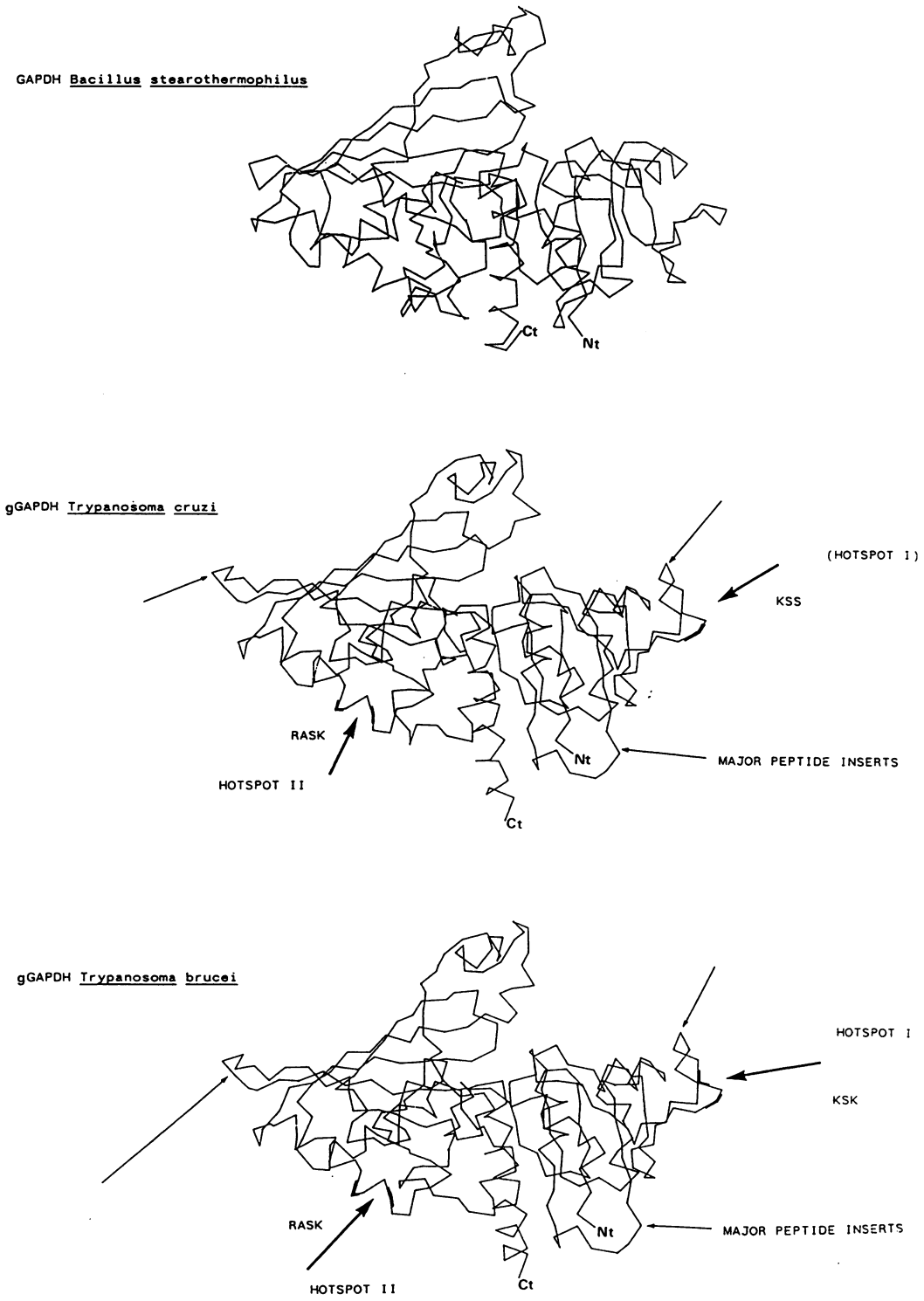


**Fig. 6.** Molecular modelling of trypanosome gGAPDH and comparison with the *B.stearothermophilus* GAPDH X-ray co-ordinates (Skarzynski *et al.*, 1987). Only C-alpha atoms are shown. The single-letter amino acid codes of residues in the regions designated as hotspots are given. Positively charged residues in various hotspot motifs are shown in bold. The major peptide insertions/extensions of trypanosomal gGAPDHs are indicated (thin arrows). Distances between hotspots were measured as average separation between the C-alpha atoms of the charged residues in the hotspot motifs.

this system. That of *T.cruzi* gGAPDH is ARL* (Figure 3) which occurs at the C-terminal of the peroxisomal enzyme spinach glycolate oxidase (Gould *et al.*, 1989). However, experiments *in vitro* to induce the uptake by peroxisomes of various glycosomal enzymes have not been successful (Borst, 1989).

It is not yet possible to assay directly for glycosomal import *in vivo* because of difficulties in the development of suitable trypanosomatid genetic transformation systems. The mechanisms of gene expression in these organisms are unusual in several features (for review, see Clayton, 1988) with the result that promoter sequences have not yet been identified unequivocally for any gene. Recent reports, however, suggest some progress in this direction (Bellofatto and Cross, 1989; Kapler *et al.*, 1990; Laban and Wirth, 1990). The use of trypanosomatid expression systems,

site-directed mutagenesis and species specific antibodies (e.g. Figure 4A) would now seem to be the most appropriate approach toward identifying the topogenic signals of glycosomal enzymes.

## Materials and methods

### *Trypanosomes*

Culture forms of *T.cruzi* were grown at 28°C in a liquid medium consisting of RPMI 1640 (Gibco) with a 0.5% (w/v) trypticase (BBL), 0.5% (w/v) HEPES, 0.03 M haemin and 10% (v/v) fetal calf serum (heat-inactivated). Strains used have been described previously (Gibson and Miles, 1986) with the exception of Y strain, an isolate from a Brazilian patient (Silva and Nussenzweig, 1953).

### *Analysis of DNA and RNA*

High molecular weight genomic DNA was isolated from parasite cells by the proteinase K method (Blaxter *et al.*, 1988). RNA was extracted from cells after lysis with 4 M guanidinium thiocyanate and pelleted by centrifugation through a 5.7 M caesium chloride gradient. Northern analysis was performed after fractionation of glyoxalated RNA through 1.2% agarose gels (Maniatis *et al.*, 1982).

### *Isolation of genomic clones*

A *T.cruzi* (strain X10/6) genomic DNA library was constructed using λEMBL3 phage (insert size range 10–20 kb). A $2 \times 10^4$ p.f.u. stock of the library was screened with a radiolabelled cDNA probe corresponding to *T.brucei* gGAPDH (a kind gift of Drs F.R.Opperdoes and P.A.M.Michels) using a stringency of 5 × SSC, 50% formamide and 42°C. Post-hybridization washes were at 2 × SSC at 42°C. After two rounds of screening a positive clone was selected, purified to homogeneity and DNA prepared (Perbal, 1984).

### *DNA sequencing*

The 5 kb *Bam*HI fragment containing both gGAPDH genes (Figure 1A) was cut with *Hin*dIII or *Pst*I and the resulting overlapping fragments were subcloned into the appropriate restriction sites in M13 mp18 or mp19 vectors. These were sequenced by the chain termination method using the Sequenase enzyme (United States Biochemical Corp.). The protein coding regions were sequenced in both strands.

### *S1 nuclease mapping*

S1 nuclease analysis was carried out essentially as described by Davis *et al.* (1986). Radiolabelled single stranded DNA was prepared by primer extension using suitable M13 subclones (Figure 5A) as templates to give probes that were either sense or antisense to the gGAPDH mRNA. The synthetic primer used in the production of probe a (Figure 5A) is identified in Figure 2. The single stranded probes were liberated by digestion with appropriate restriction enzymes and purified from the unlabelled M13 DNA on 7% polyacrylamide gels. In the case of probe c (Figure 5A) this procedure also liberated the unlabelled complementary strand which was co-purified. Probes were hybridized with total *T.cruzi* X10/6 RNA as described (Figure 5) and digested wtih S1 nuclease under a variety of conditions. Products were analysed on polyacrylamide sequencing gels. In control experiments the RNA was pretreated with DNase or RNase (not shown).

### *Production of anti-gGAPDH antibodies*

A peptide corresponding to the final 10 amino acids of *T.cruzi* gGAPDH (MASKDRSARL*) was synthesized (Dr C.Howard, L.S.H.T.M.) using an automated Merrifield process and conjugated to keyhole limpet haemocyanin (KLH) (Cambridge Research Biochemicals). Mice (C57, BL/10, Sn SC ♀) were inoculated with 100 μg of conjugate in Freund's complete adjuvant. After 1 month, they were boosted with a further 100 μg of conjugate in Freund's incomplete adjuvant and finally boosted 1 month later with 100 μg of conjugate alone. They were bled after a further 5 days and the sera used to probe Western blots (Figure 4B; Blaxter *et al.*, 1988).

## Acknowledgements

We thank Dr Paul Kaye for help in preparing anti-peptide antibody and Drs Fred Opperdoes and Paul Michels (International Institute of Molecular Pathology, Brussels) for providing the *T.brucei* gGAPDH cDNA clone, and thank Drs Alan Fairlamb, Paul Michels and Martin Taylor for critical reading of this manuscript. This work was supported by the Wellcome Trust, Wolfson Foundation, UNPD/World Bank/WHO Special Programme for Research into Tropical Diseases. Giles Kendall was supported by a Hazel-Clifford studentship from London University.

## References

Bellofatto,V. and Cross,G.A.M. (1989) *Science*, **244**, 1167–1169.
Blaxter,M.L., Miles,M.A. and Kelly,J.M. (1988) *Mol. Biochem. Parasitol.*, **30**, 259–270.
Borst,P. (1986a) *Annu. Rev. Biochem.*, **55**, 701–732.
Borst,P. (1986b) *Biochim. Biophys. Acta*, **866**, 179–203.
Borst,P. (1989) *Biochim. Biophys. Acta*, **1008**, 1–13.
Brener,Z. (1986) *Parasitol. Today*, **2**, 196–197.
Clayton,C.E. (1987) *J. Cell Biol.*, **105**, 2649–2651.
Clayton,C.E. (1988) In Rigby,P.W. (ed.), *Genetic Engineering*. Academic Press, London, Vol. 7, pp.1–56.
Davidson,B.E., Sajgo,M., Noller,H.F. and Harris,J.I. (1967) *Nature*, **216**, 1181–1185.
Davis,L.G., Dibner,M.D. and Battey,J.F. (1986) *Basic Methods in Molecular Biology*. Elsevier, Amsterdam.
Dragon,E.A., Sias,S.R., Kato,E.A. and Gabe,J.D. (1987) *Mol. Cell. Biol.*, **7**, 1271–1275.
Eisenberg,D., Schwartz,E., Komaromy,M. and Wall,R. (1984) *J. Mol. Biol.*, **179**, 125–142.
Fairlamb,A.H. and Opperdoes,F.R. (1986) In Morgan,M.J. (ed.), *Carbohydrate Metabolism in Cultured Cells*. Plenum Publishing Corp., New York, pp.183–224.
Gibson,W.C. and Miles,M.A. (1986) *EMBO J.*, **5**, 1299–1305.
Gould,S.J., Keller,F.A., Hosker,N., Wilkinson,J. and Subramani,S. (1989) *J. Cell Biol.*, **108**, 1657–1644.
Hart,D.T., Bandhuin,P., Opperdoes,F.R. and de Duve,C. (1987) *EMBO J.*, **6**, 1403–1411.
Hasan,G., Turner,M.J. and Cordingley,J.S. (1984) *Cell*, **37**, 333–341.
Kapler,G.M., Zhang,K. and Beverly,S.M. (1987) *Nucleic Acids Res.*, **15**, 3369–3383.
Kapler,G.M., Coburn,C.M. and Beverly,S.M. (1990) *Mol. Cell. Biol.*, **10**, 1084–1094.
Laban,A., Tobin,J.F., Curotto de Lafaille,M.A. and Wirth,D.F. (1990) *Nature*, **343**, 572–574.
Laird,P.W., Zomerdijk,J.C.B.M., de Korte,D. and Borst,P. (1987) *EMBO J.*, **6**, 1055–1062.
Layden,R.E. and Eisen,H. (1988) *Mol. Cell. Biol.*, **8**, 1352–1360.
Le Blancq,S.M., Swinkels,B.W., Gibson,W.C. and Borst,P. (1988) *J. Mol. Biol.*, **200**, 439–447.
Maignon,R., Gerke,R., Rodriguez,M., Urbina,J., Hoenicka,J., Negri,T., Aguirre,T., Nehlin,J., Knapp,T. and Crampton,J. (1988) *Eur. J. Biochem.*, **171**, 285–291.
Maniatis,T., Fritsch,E.M. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Marchand,M., Poliszczak,A., Gibson,W.C.,Wierenga,R.K., Opperdoes,F.R. and Michels,P.A.M. (1988) *Mol. Biochem. Parasitol.*, **29**, 65–76.
Marchand,M., Kooystra,U., Wierenga,R.K., Lambier,A.-M., Van Beeumen,J., Opperdoes,F.R. and Michels,P.A.M. (1989) *Eur. J. Biochem.*, **184**, 455–464.
Michels,P.A.M. (1986) *J. Mol. Evol.*, **24**, 45–52.
Michels,P.A.M. (1989) *Exp. Parasitol.*, **69**, 310–315.
Michels,P.A.M., Poliszczak,A., Osinga,K.A., Misset,O., Van Beeumen,J., Wierenga,P.K., Borst,P. and Opperdoes,F.R. (1986) *EMBO J.*, **5**, 1049–1056.
Miles,M.A. (1983) *Trans. R. Soc. Trop. Med. Hyg.*, **77**, 5–22.
Muhich,M.L. and Boothroyd,J.C. (1988) *Mol. Cell. Biol.*, **8**, 3837–3846.
Opperdoes,F.R. (1987) *Annu. Rev. Microbiol.*, **41**, 127–151.
Opperdoes,F.R. (1988) *Trends Biochem. Sci.*, **13**, 255–260.
Opperdoes,F.R., Wierenga,R.K., Noble,M.E.M., Hol,W.G.J., Willfon,M., Kuntz,D.A., Callens,M. and Perrie,J. (1990) In Agabian,N. and Cerami,A. (eds) *Parasites; Molecular Biology, Drug and Vaccine Design*. Wiley–Liss, inc., New York, pp. 233–246.
Perbal,B.A. (1984) *A Practical Guide to Molecular Cloning*. John Wiley and Sons Inc., Chichester, UK.
Ralph,D., Huang,J. and Van der Ploeg,L.H.T. (1988) *EMBO J.*, **7**, 2539–2545.
Requena,J.M., Lopez,M.C., Jimenez-Ruiz,.A, de la Torre,J.M. and Alonso,C. (1988) *Nucleic Acids Res.*, **16**, 1393–1406.
Scholler,J.K., Aline,R.F.,Jr and Stuart,K.D. (1988) *Mol. Biochem. Parasitol.*, **29**, 89–104.
Silva,L.H.P. and Nussenzweig,V. (1953) *Folia Clin. Biol.*, **20**, 191–207.

Skarzynski,T., Moody,P.C.E. and Wonacott,A.J. (1987) *J. Mol. Biol.*, **193**, 171–187.
Sutcliffe,M.J., Haneef,I., Carrey,D. and Blundell,T.L. (1987a) *Prot. Eng.*, **1**, 377–384.
Sutcliffe,M.J., Hayes,F.R.F. and Blundell,T.L. (1987b) *Prot. Eng.*, **1**, 385–392.
Swindle,J., Ajioka,J., Eisen,H., Sanwal,B., Jacquemot,C., Browder,Z. and Buck,G. (1988) *EMBO J.*, **7**, 1121–1127.
Swinkels,B.W., Evers,R. and Borst,P. (1988) *EMBO J.*, **7**, 1159–1165.
Thaler,D.S. and Stahl,F.W. (1988) *Annu. Rev. Genet.*, **2**, 169–197.
Tschudi,C., Young,A.S., Ruben,L., Patton,C.L. and Richards,F.F. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 3998–4002.
Tso,J.Y., Sun,X.-H., Kao,T.H., Reece,K.S. and Wu,R. (1985) *Nucleic Acids Res.*, **13**, 2485–2502.
Von Heijne,G. (1986) *EMBO J.*, **5**, 1335–1342.
Wiener,A.M., Dieninger,P.L. and Efstratidis,A. (1986) *Annu. Rev. Biochem.*, **55**, 631–661.
Wierenga,R.K. (1987) *EMBO J.*, **6**, 215–222.

*Received on February 13, 1990; revised on May 28, 1990*