The EMBO Journal vol.7 no.4 pp.1159-1165, 1988

The topogenic signal of the glycosomal (microbody) phosphoglycerate kinase of *Crithidia fasciculata* resides in a carboxy-terminal extension

Bart W.Swinkels, Raymond Evers[1] and Piet Borst

Division of Molecular Biology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands
[1]Present address: Molecular Parasitology Unit, Max Planck-Institut für Biologie, Spemannstrasse 34, D-7400 Tübingen, FRG

Communicated by P.Borst

To determine how microbody proteins enter microbodies, we have previously compared the genes for the cytosolic and glycosomal (microbody) phosphoglycerate kinases (PGKs) of *Trypanosoma brucei* and found the microbody enzyme to differ from other PGKs and the cytosolic form in two respects: a high net positive charge and a C-terminal extension of 20 amino acids (Osinga *et al.*, 1985). Here we present the comparison of the genes for the cytosolic and glycosomal PGKs of *Crithidia fasciculata*, another kinetoplastid organism. The amino acid sequences of the two *Crithidia* isoenzymes are virtually identical, except for a C-terminal extension of 38 amino acids. We conclude that this extension must direct the glycosomal PGK to the glycosome. The extensions of the *Crithidia* and *Trypanosoma* enzymes are both rich in small hydrophobic and hydroxyl amino acids.
Key words: glycosome/topogenesis/peroxisome biogenesis/phosphoglycerate kinase/*Crithidia*

Introduction

Most eukaryote proteins are made in the cytosol, during or after which they are routed to their correct location in the cell or secreted. Topogenic signals involved in the initial steps of this routing have been identified for several subcellular destinations, such as the endoplasmic reticulum, mitochondria, chloroplasts and the nucleus. This information is still lacking for one group of organelles, collectively known as microbodies, which include peroxisomes, glyoxysomes and glycosomes. Microbody proteins are encoded in the nucleus, synthesized on free polysomes and imported into the organelle post-translationally (reviewed by Lazarow and Fujiki, 1985; Borst, 1986). Although most microbody proteins are synthesized at their final size, some are made as larger precursors. In no case, however, is proteolytic processing directly coupled to the import process and whether the pre-part of these precursors is involved in topogenesis at all is uncertain. Hence, signals for uptake into the organelle have not been identified with certainty, even though a substantial number of genes for microbody proteins have been sequenced.

Trypanosomes and other Kinetoplastida are characterized by an unusual microbody that harbours enzymes of the glycolytic pathway (Opperdoes and Borst, 1977; Opperdoes *et al.*, 1984), in addition to (parts of) several other metabolic pathways (see Opperdoes, 1987). We have used this transfer of glycolysis from its usual location in the cytosol into a microbody to study what changes are required in cytosolic enzymes to allow their delivery into microbodies. Several genes for glycolytic enzymes in *Trypanosoma brucei* have been analysed. These are the genes for aldolase (Clayton, 1985), phosphoglycerate kinase (PGK; Osinga *et al.*, 1985), triosephosphate isomerase (TIM; Swinkels *et al.*, 1986) and glyceraldehyde-phosphate dehydrogenase (GAPDH, Michels *et al.*, 1986). PGK has been most informative on the signals involved in microbody uptake, since trypanosomes contain a cytosolic and a glycosomal PGK that differ only in 7% of their amino acid sequences. The differences mainly affect two points: at neutral pH the glycosomal PGK has a high net positive charge, whereas the cytosolic PGK is non-charged; and the glycosomal PGK has a C-terminal extension of 20 amino acids, not present in the cytosolic enzyme (Osinga *et al.*, 1985). Either of these differences or a combination of both must determine the difference in routing of the two PGKs.

The other three glycolytic enzymes analysed in *T.brucei* either have no cytosolic counterpart (aldolase and TIM) or its sequence is not yet available (GAPDH). However, in comparison to the (cytosolic) homologues in other organisms, the glycosomal aldolase, TIM and GAPDH lack terminal extensions but do have a distinctly higher isoelectric point (Misset *et al.*, 1986). The analysis of the spatial organization of the extra positive charges in TIM, PGK and GAPDH led Wierenga *et al.* (1987) to propose a novel type of signal for import into the glycosome: two regions of excess positive charge (hot spots), separated by a distance of about 40 Å on the surface of the proteins and each including a pair of positive residues about 7 Å apart. However, this proposal has left some observations unexplained (discussed by Borst, 1986; Wierenga *et al.*, 1987).

To critically test the proposed topogenic signal, a system is required in which import of proteins derived from mutagenized genes can be assayed. Although we have obtained conditions in which the glycosomal PGK, synthesized in a coupled *in vitro* transcription-translation system, is specifically bound to a *T.brucei* glycosome fraction whereas the cytosolic PGK is not, all our attempts to get organelle uptake, as defined by protection against proteolysis, have failed thus far (our unpublished experiments, 1986). Short of an import assay, we have turned to PGK genes from other kinetoplastid organisms, hoping that genetic change during evolution would bring into focus the differences between cytosolic and glycosomal PGK that are essential for topogenesis. The insect parasite *Crithidia* is suitable for this purpose: it contains glycosomes (Taylor *et al.*, 1980) and both a cytosolic and glycosomal PGK (Klein and Miller, 1981). Within the Trypanosomatidae studied, *Crithidia* is the least related to *T.brucei* as judged from a comparison of the mitochondrial rRNA genes (Sloof *et al.*, 1985; unpublished results of de la Cruz *et al.*, quoted by Muhich *et*

©IRL Press Limited, Oxford, England

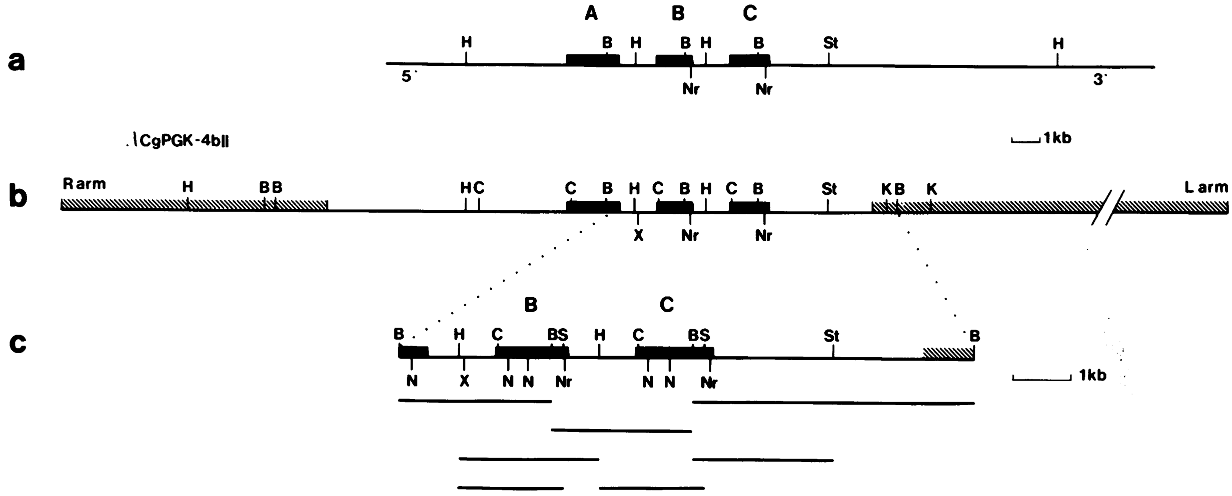

Fig. 1. (**a**) Genomic environment of the three tandemly linked PGK genes in *C.fasiculata*. Solid blocks indicate protein coding sequences. 5′ and 3′ indicate direction of transcription. (**b**) Physical map of λ CgPGK-4bII, containing a genomic fragment cloned into EMBL 3 (indicated by hatched boxes). (**c**) Enlargement of the area analysed. Bars below the map represent subclones used to characterize and sequence the genes B and C. Restriction sites: B = *Bgl*II, C = *Cla*I, H = *Hin*dIII, K = *Kpn*I, N = *Nae*I, Nr = *Nar*I, S = *Sal*I, St = *Stu*I and X = *Xba*I.

```
Gene B   GAGCAAACTGCTGTCTTTTCTTTCACGATAACACACTCATATTAAACGCGAGTATTACTCATCAGTCAACGTCACATTCCGCTCTGTCCACTTCGACCTTACACCTCTAC
         |  |    || |||     ||  |        |  | ||  |||||  |  |                  |  |  |||  |    |  | ||   | | | |   -29
Gene C   AATTATTACTCTCTCATCAGAGGACCGCAACTTGTTATACCCTCAAGTCGAGTCTCTCCACCTGTCACCGTCAATTATTAGCTTGCCCGCAACCTGCATTCTTTCTTTTC

TTGTCCAACTATCTTCCACTTGTCAAGC met ser leu ala pro lys lys thr ile asp asp ala val val lys gly lys lys val leu ile arg val
|   ||    |   |   |||        ATG TCT CTC GCC CCG AAG AAG ACG ATC GAT GAT GCC GTG GTG AAG GGC AAG AAG GTC CTC ATC CGT GTG   69
TCTCGCACAACACCCGAAGACTTCACAA             TT              GC
                                         val             ser

asp phe asn val pro val lys asn gly glu ile thr asn asp phe arg ile arg ser ala leu pro thr ile gln lys val leu lys glu
GAC TTC AAC GTG CCG GTG AAG AAC GGC GAG ATC ACG AAC GAC TTC CGC ATC CGC TCC GCC CTG CCG ACG ATC CAG AAG GTG CTG AAG GAG   159

gly gly ser cys ile leu met ser his leu gly arg pro lys gly ala lys met ser asp pro lys pro ala lys ser val arg gly tyr
GGC GGC TCC TGC ATC CTG ATG AGC CAC CTC GGC CGC CCG AAG GGT GCG AAG ATG AGC GAC CCG AAG CCG GCC AAG AGC GTG CGC GGG TAC   249

glu glu ala ala thr leu arg pro val ala ala arg leu ser glu leu leu gly gln lys val glu phe ala pro asp cys leu asp ala
GAG GAG GCT GCG ACG CTG CGG CCG GTG GCT GCG CGG CTG TCG GAG CTG CTG GGG CAG AAG GTG GAG TTC GCG CCG GAC TGC CTG GAC GCT   339

ala ser tyr ala ala lys leu lys gly gly asp val leu leu leu glu asn val arg phe tyr ala glu glu gly ser lys lys glu glu
GCG TCG TAC GCT GCT AAG CTG AAG GGC GGC GAC GTG CTG CTG CTG GAG AAC GTG CGC TTC TAC GCG GAG GAG GGC AGC AAG AAG GAG GAG   429

glu arg asp ala met ala lys val leu ala ala tyr gly asp val tyr val ser asp ala phe gly thr ala his arg asp ser ala thr
GAG CGC GAC GCG ATG GCG AAG GTG CTT GCG GCG TAC GGC GAC GTG TAC GTG AGC GAC GCC TTC GGC ACT GCG CAC CGC GAC AGC GCG GAC   519

met thr gly ile pro lys val leu gly ala gly tyr ala gly tyr leu met glu lys glu ile asn tyr phe ala gln val leu asn asn
ATG ACG GGC ATC CCG AAG GTG CTG GGT GCG GGC TAC GCC GGC TAC CTG ATG GAG AAG GAG ATC AAC TAC TTC GCG CAG GTG CTG AAC AAC   609

pro pro arg pro leu val ala ile val gly gly ala lys val ser asp lys ile gln leu leu asp asn met leu gly arg ile asn tyr
CCG CCG CGT CCG CTG GTG GCC ATC GTG GGC GGT GCG AAG GTT AGC GAC AAG ATC CAG CTG CTG GAC AAC ATG CTG GGC CGC ATC AAC TAC   699

leu val ile gly gly ala met ala tyr thr phe gln lys ala gln gly his ala ile gly ile ser met cys glu glu asp lys leu asp
CTG GTG ATT GGC GGC GCG ATG GCG TAC ACG TTC CAG AAG GCG CAG GGC CAC GCG ATC GGC ATC TCG ATG TGC GAG GAG GAC AAG CTG GAC   789

leu ala lys ser leu leu lys lys ala gln glu arg asn val glu val leu leu pro val asp his val cys asn lys glu phe lys ala
CTT GCC AAG TCG CTG CTG AAG AAG GCG CAG GAG CGC AAC GTG GAG GTG CTT CTG CCG GTG GAC CAC GTG TGC AAC AAG GAG TTC CAA GGC   879

val asp ala pro leu val thr lys asp val glu ile pro glu gly tyr met ala leu asp ile gly pro lys thr ile lys ile tyr glu
GTG GAC GCG CCG CTG GTG ACG AAG GAC GTG GAG ATC CCG GAG GGG TAC ATG GCG CTG GAC ATT GGC CCG AAG ACG ATC AAG ATC TAC GAG   969

asp val ile ala lys cys lys ser thr ile trp asn gly pro met gly val phe glu met pro cys tyr ser lys gly thr phe ala val
GAC GTG ATT GCC AAG TGC AAG AGC ACG ATC TGG AAC GGC CCG ATG GGT GTG TTC GAG ATG CCG TGC TAC TCG AAG GGC ACG TTC GCT GTG   1059

ala lys ala met gly asn gly thr gln lys asn gly leu met ser ile ile gly gly gly asp thr ala ser ala ala glu leu ser gly
GCC AAG GCG ATG GGC AAC GGG ACG CAG AAG AAC GGG CTG ATG AGC ATC ATC GGC GGC GGC GAC ACG GCG AGC GCT GCG GAG CTG AGC GGG   1149

glu ala lys asn met ser his val ser thr gly gly gly ala ser leu glu leu leu glu gly lys ser leu pro gly val thr val leu
GAG GCG AAG AAC ATG TCG CAC GTG TCG ACT GGC GGC GGC GCC TCG CTG GAG CTG CTG GAG GGC AAG TCG CTG CCG GGT GTC ACC GTT CTC   1239
                                                  T                                               C

thr asn lys glu *** 1254
ACC AAC AAG GAG TAG AATGTTGAGTCGAGTGCGAGACGCGGAATTGCACGGTTTTCGCCGTCATCGTCGTGATATGCGGCCG   1321
                  C GCG AAG GCG CCT GCT GCT GCT GCC GCT GCC GGT GGT GAC TGC CCG TGC GGT TCC GGC TGC GCT GCG GTT CCC GCC GCA   1329
                asp ala lys ala pro ala ala ala ala ala ala gly gly asp cys pro cys gly ser gly cys ala ala val pro ala ala

GCC ACG GCA ACG GTT TCG ATG GTG CTC GCC TCG CCA TGA AGATTCTGAAGCTGCTCGCCGCGGTGC   1395
ala thr ala thr val ser met val leu ala ser pro *** 1368
```

Fig. 2. Comparison of the nucleotide and amino acid sequences of genes B and C of *C.fasciculata*. The coding regions of genes B and C start at position 1. Upstream of the coding regions positional identities are indicated with vertical bars. Possible splice acceptor AG sequences and the pyrimidine-rich stretches possibly also involved in splicing (see Laird *et al.*, 1987) are underlined. Nucleotide and amino acid substitutions in gene C relative to gene B are shown below the gene B sequence.

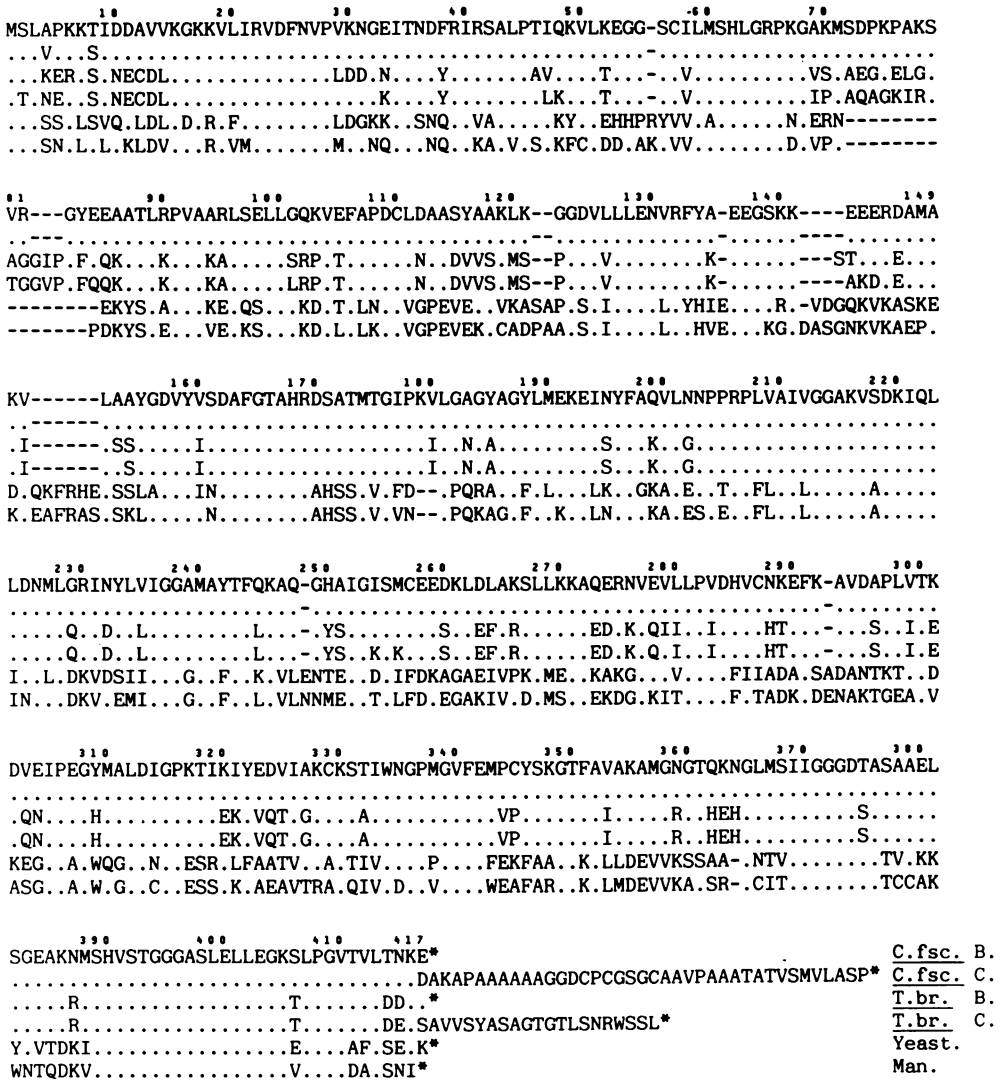

Fig. 3. A comparison of the amino acid sequences of *Crithidia* and *Trypanosoma* cytosolic (B) and glycosomal (C) PGKs with yeast and human PGKs. The numbering refers to the *Crithidia* cytosolic PGK and only this sequence is shown in full. Identical residues in the other sequences are indicated by dots. Sequences have been aligned to obtain maximal positional identity. Insertions and deletions (indicated by dashes) have been positioned such that they do not fall within secondary structure areas, as determined for the yeast PGK. Sequences are from yeast (*Saccharomyces cerevisiae*) (Watson *et al.*, 1982), man (Michelson *et al.*, 1983) and trypanosome (Osinga *et al.*, 1985).

al., 1987), facilitating the search for conserved elements.

In this paper we present our analysis of the PGK genes and proteins in *C.fasciculata*. The cytosolic and glycosomal isoenzymes are in essence identical but for a C-terminal extension on the glycosomal PGK. It follows that this extension must direct the glycosomal PGK to the glycosome.

Results

Isolation of the Crithidia PGK genes

We used a gene-internal DNA probe from one of the PGK genes of *T.brucei* to identify the corresponding gene(s) in *Crithidia* nuclear DNA and in a genomic DNA recombinant clone bank. The physical maps of the genomic area and of the genomic clone and subclones used to characterize the genes are presented in Figure 1. As in *T.brucei*, we found three genes with homology to the trypanosome probe and to each other in a tandem array. The homology between the *Crithidia* gene A and the other two genes (B and C) is not as high as in *T.brucei* (Osinga *et al.*, 1985; Le Blancq *et al.*, 1988); at reduced stringency (3 × SSC, 65°C) only weak cross-hybridization is seen (not shown) and at higher stringencies hybridization is lost (see Figure 4b). These are the only sequences in the genome of *C.fasciculata* that hybridize with homologous or *T.brucei* PGK probes under these conditions (results not shown).

Sequence of Crithidia PGK genes

The nucleotide sequences and derived amino acid sequences of the genes B and C as well as the nucleotide differences between the genes are presented in Figure 2. Figure 3 shows a comparison of the amino acid sequences of the proteins encoded by *Crithidia* genes B and C and the PGKs from *T.brucei*, yeast and man. Partial sequence analysis of the *Crithidia* gene A shows the gene to be the homologue of the corresponding gene in *T.brucei*, which does not encode one of the major PGKs (Osinga *et al.*, 1985). Instead, the *Crithidia* and *T.brucei* genes A encode PGK-related proteins of unknown function, characterized by an insertion of about 100 amino acids (see also Le Blancq *et al.*, 1988). As the genes A are not directly relevant to glycosome topogenesis, they will not be further considered here.

The sequence comparison of Figure 3 shows a high degree of similarity between the deduced amino acid sequences of the *Crithidia* genes B and C and the other eukaryote PGKs. The sizes of the predicted proteins are similar, the amino acid positional identity is about 45%, including the conservation of all residues essential for enzyme function with respect to horse and yeast PGKs (Banks *et al.*, 1979; Watson *et al.*, 1982). The *Crithidia* and yeast genes can be aligned without major insertions or deletions, indicating that the *Crithidia* genes have no introns. We conclude from these results that both genes B and C code for a functional PGK. The *Crithidia* and trypanosome genes resemble each other more than the other eukaryote PGKs: ~70% identity at the nucleotide level and ~75% at the amino acid level. They are both characterized by several minor trypanosomatid-specific insertions and deletions.

The striking and unexpected feature of the sequence comparison is, however, that the *Crithidia* genes B and C are virtually identical. The deduced amino acid sequences differ only in two conservative amino acid subsitutions near the N-terminus and a 38 amino acid extension of the gene C protein at its C-terminus. The nucleotide sequences of both genes differ in only six positions, four of which account for the two amino acid substitutions near the N-terminus; two other silent single base substitutions are found towards the 3′ end of the genes. Immediately upstream of the initiator Met codon and downstream of the gene B stopcodon, the sequence identity of both genes is abruptly lost.

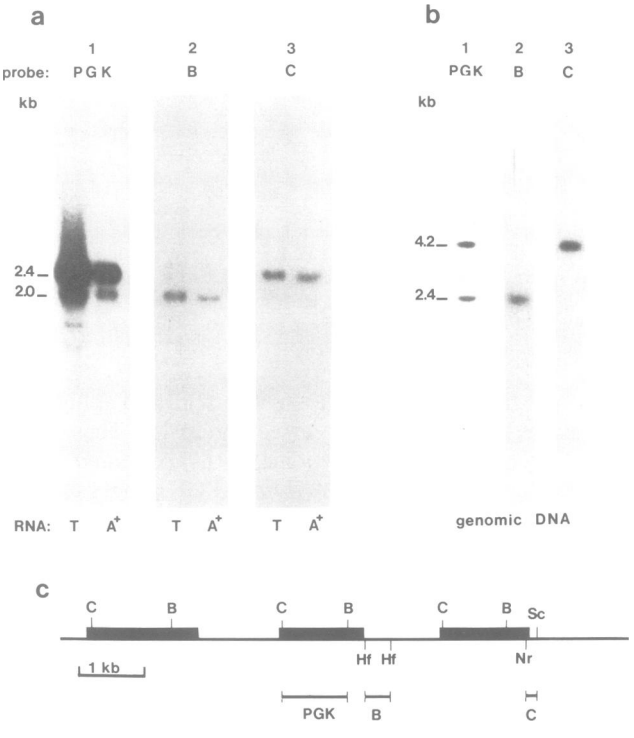

Fig. 4. (a) Total (T) and poly(A)$^+$ RNA was size-fractionated by agarose gel electrophoresis and transferred to nitrocellulose filters. Identical strips of the filter were hybridized with ^{32}P-labelled PGK and gene specific probes as indicated. (b) Nuclear DNA from *C.fasciculata* was digested with *Hin*dIII and *Stu*I, size-fractionated by agarose gel electrophoresis, transferred to nitrocellulose and hybridized with probes used in panel (a), to demonstrate their specificity. (c) Simplified map from Figure 1, indicating the positions of the probes. For *Hin*dIII and *Stu*I sites, see Figure 1.

Crithidia PGK transcripts and proteins

The striking similarity of the PGK gene arrays of *T.brucei* and *Crithidia* suggests that *Crithidia* gene B codes for the cytosolic PGK and gene C for the glycosomal isoenzyme, as in *T.brucei*. To verify this, we have studied the transcripts of the *Crithidia* genes and the *Crithidia* PGKs.

On a RNA blot we detect two RNA species with a gene-internal PGK probe from gene B, in both total and poly(A)-enriched RNA (see Figure 4a, lane 1): one abundant transcript of 2.4 kb and one less abundant transcript of 2.0 kb, present at about half the level of the 2.4 kb transcript (see Table I). To match these RNAs with their corresponding genes, we hybridized identical strips of the RNA blot with probes specific for either gene B or C, the positions of which are indicated in Figure 4c. The specificity of the probes is documented in Figure 4b. It is clear from Figure 4 that the 2.0 kb transcript comes from gene B and the abundant 2.4 kb transcript from gene C.

Although glycosomes have been found in *C.fasciculata* (Taylor *et al.*, 1980), and cytosolic as well as glycosomal PGK activities in *C.luciliae* (Klein and Miller, 1981), the subcellular distribution and size of PGKs in *C.fasciculata* have not been studied. We have therefore analysed subcellular fractions of *C.fasciculata* using an antiserum raised against the glycosomal PGK of *T.brucei*. Figure 5

Table I. The PGK genes, transcripts and proteins of *C.fasciculata*

Genes	B	C
Size mRNA (kb)	2.0	2.4
Relative amount mRNA[a]	28%	72%
Predicted mol. wt of protein (kd)	44.465	47.706
Proteins	**Cytosol**	**Glycosome**
Relative amount protein[b]	33%	67%
Mol. wt (kd)[c]	43.5 ± 0.7	48.0 ± 1.0

[a,b]Determined from densitometric scans of the autoradiographs in Figures 4 and 5, respectively.
[c]Determined by SDS−PAGE and antibody staining as in Figure 5 (average values of four determinations).

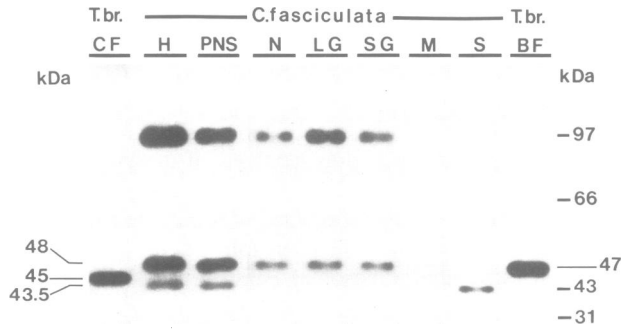

Fig. 5. Western blot analysis of proteins from *C.fasciculata* subcellular fractions. For comparison the extreme left and right lanes show the cytosolic and glycosomal PGKs in *T.brucei* cultured insect form and bloodstream form homogenates respectively. The *Crithidia* fractions from left to right are: H: homogenate, PNS: post nuclear supernatant, N: nuclear (1000 *g*) pellet, LG: large granular (5000 g) pellet, SG: small granular (14 500 *g*) pellet, M: microsomal (140 000 *g*) pellet and S: soluble fraction. Proteins were size-fractionated by SDS−PAGE, electroblotted onto nitrocellulose and detected using antibodies against the glycosomal PGK of *T.brucei* and ^{125}I-labelled staphylococcal protein A. On the left the positions of the mol. wt markers are indicated (bars). Thin lines indicate the mol. wt of each of the four PGKs.

shows a Western blot of the *Crithidia* fractions compared to the cytosolic and glycosomal PGKs from *T.brucei*. Proteins of 43.5, 48 and 96 kd in the *C.fasciculata* homogenate react with the heterologous antiserum. Since all PGKs analysed thus far have a mass of around 45 kd (Scopes, 1973), the most simple interpretation is to assign the 43.5 and 48 kd proteins to the *Crithidia* PGKs. The product of gene A is not detectable under these conditions. The 96 kd protein remains unidentified and is probably unrelated to the PGKs, as our PGK gene specific probes only detect genes A, B and C. Note that no reaction with a 96 kd protein occurs in the *T.brucei* homogenates. The subcellular fractionation shows that the 48 kd protein is particle-bound whereas the 43.5 kd protein is cytosolic. We conclude that *C.fasciculata* contains cytosolic and glycosomal PGK isoenzymes with a mass of 43.5 and 48 kd, respectively.

Table I presents a comparison of the relative amounts of the transcripts and proteins as well as a comparison of the sizes of the proteins deduced from the gene sequences and determined by gel electrophoresis. We infer from these data that the cytosolic PGK is encoded by gene B and the glycosomal PGK by gene C, analogous to the situation in *T.brucei*.

Isoelectric focusing of the Crithidia PGKs

We have previously found that the main difference between glycosomal enzymes from *T.brucei* and their counterparts in the cytosol, or in other organisms, is the high net positive charge on the glycosomal enzymes at neutral pH (Osinga *et al.*, 1985. Wierenga *et al.*, 1987). In contrast, the gene C encoded glycosomal PGK of *C.fasciculata* has a net charge which is equal to that of its cytosolic counterpart and which is only slightly positive (3+). To exclude the remote possibility that the particle-bound PGK of *Crithidia* would acquire positive charge by secondary modification of the protein, we have compared the *Crithidia* PGK proteins with the glycosomal PGK of *T.brucei* on a two-dimensional gel, with non-equilibrium pH gradient electrophoresis in the first dimension and SDS−PAGE in the second dimension. Figure 6, panel a shows the result obtained with a *Crithidia* protein lysate. The two *Crithidia* PGKs migrate to the centre of the gel in the first dimension and only a single spot is seen for each isoenzyme, indicating that both PGKs have a neutral p*I* and that no allelic charge polymorphisms are present. Figure 6, panel b shows an identical experiment in which the *Crithidia* lysate was mixed with a *T.brucei* bloodstream form lysate, containing the glycosomal PGK for comparison. Several new spots appear in the basic area of the gel, some of which with a mol. wt of the *T.brucei* glycosomal PGK (47 kd) and some with a mol. wt of 43 kd. Because similar results were obtained when the purified *T.brucei* glycosomal PGK was mixed with the *Crithidia* lysate, the cytosolic PGK of *T.brucei* does not cause the 43 kd spots (data not shown). We attribute these smaller proteins to proteolysis of the *T.brucei* glycosomal PGK in the homogenate, probably occuring during denaturation in urea.

We conclude from these data that the 48 kd particle-bound PGK of *Crithidia*, detected by our PGK antiserum has a neutral p*I*, as predicted by the sequence of gene C.

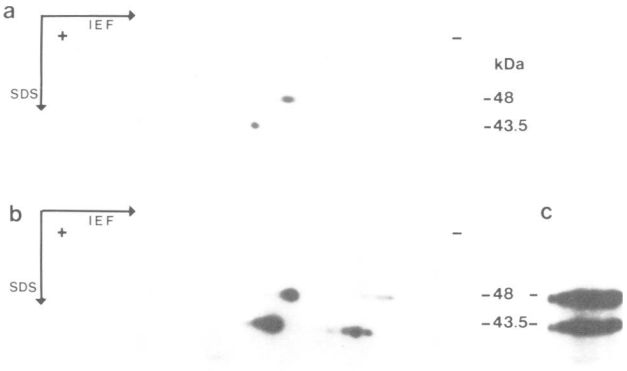

Fig. 6. (**a**) Two-dimensional gel electrophoresis of a *Crithidia* homogenate. Non-equilibrium pH gradient gel electrophoresis in the first dimension (sample was applied at the anodic side), SDS−PAGE in the second dimension. Proteins were electro-blotted onto nitrocellulose and detected as in Figure 5. (**b**) As in (a), except that *T.brucei* bloodstream form homogenate was mixed with the *Crithidia* homogenate. (**c**) The *Crithidia* homogenate was run only in the second dimension to serve as mol. wt markers. In the second dimension samples of panels (a), (b) and (c) were run on a single SDS−PA gel. Only the 45 kd region of this gel is shown. The positions of the anode (+) and cathode (−) are indicated.

Discussion

We have cloned the PGK genes of *C.fasciculata* and found three genes (A, B and C) in a tandem array. The evidence presented strongly suggests that the two major PGKs of *C.fasciculata*, the cytosolic and glycosomal PGKs, are encoded by genes B and C respectively, although we have no direct amino acid sequence data to link the genes to the proteins. The deduced amino acid sequences of the two isoenzymes are almost identical. The only differences are two amino acid substitutions near the N-terminus and a C-terminal extension of 38 amino acids on the glycosomal PGK. These differences must determine the difference in routing of the two isoenzymes. Both N-terminal substitutions are conservative (Val−Ala; Ser−Thr) and the *T.brucei* isoenzymes do not have similar differences at these positions. It is therefore highly unlikely that these substitutions as such (or via post-translational modification) are involved in topogenesis. We conclude that the 38 amino acid extension of the glycosomal PGK is essential for glycosomal uptake, probably by acting as a topogenic signal. However, the unlikely alternative that the extension changes the folding of the protein and thus leads to exposure of an internal signal

```
- D A K A P A A A A A A G G D C P C G S G C A A V P A   A   A   V   M V L A S P *
                                                    A   T   T   S
                                                    |   |   |   |       |
                                                    A   T   T   S
- - - - - - - - - - - - - - - - - - - - S A V V S Y A S   G   G   L   N R W S S L *
```

Fig. 7. A comparison of the C-terminal extensions of the glycosomal PGKs of *C.fasciculata* (top) and *T.brucei* (bottom). Identical residues are indicated.

remains to be excluded.

Topogenic signals are highly conserved throughout nature (see Borst, 1986), albeit not so much in their amino acid positional identity as in their functional physicochemical characteristics (Von Heijne, 1985, 1986). Our results therefore imply that the high net positive charge on the *T.brucei* glycosomal enzymes, either as such or in the 'hot spot' configuration (Wierenga *et al.*, 1987), cannot be essential for uptake into the glycosome. The only feature conserved between both glycosomal PGKs is a C-terminal extension. The length of this extension is not conserved, suggesting that part of the extension on the *Crithidia* PGK may not be essential. This could be the N-terminal half of the *Crithidia* extension with its funny string of alanines. Both extensions are rich in small hydrophobic and hydroxyl amino acids (Ala, Gly, Ser, Thr, Val, Pro and Cys) and have only 5% charged residues with no preponderance of positive or negative charge. Amino acid positional identity is low, but if the C-termini of both extensions are aligned, six identical residues occur, four of which in one block are alternating with non-identical resiudes, suggesting that this part of the C-termini might be in a β-sheet conformation of which one side is functionally more important that the other (see Figure 7). We do not see other similarities between both extensions that might provide clues to the functional requirements of glycosomal topogenesis.

The topogenic signals in the glycosomal TIM, GAPDH and aldolase from *T.brucei* are not present in extensions and must be internal. There are several candidate areas in each of them with credible resemblance to the common features of the PGK extensions in Figure 7. As these common features are so meagre, sequence comparisons do not provide a stringent test, however.

We have previously argued that all microbodies, including glycosomes, share a common ancestor (Osinga *et al.*, 1985; Borst, 1986). As addressing signals for other organelles have been found to be highly conserved in evolution, we expect that this will also hold for microbody signals. Indeed, it has recently been shown that the peroxisomal targeting signal of the firefly luciferase is recognized by mammalian peroxisomes (Keller *et al.*, 1987). A search for these signals in published sequences of peroxisomal and glyoxysomal, proteins is hampered by the lack of a stringent sequence requirement test. Moreover, sequences of isoenzymes in other cellular compartments are not available to limit the number of candidate signals. The only exception is the rat peroxisomal and mitochondrial 3-ketoacyl-CoA thiolase isoenzyme couple for which the sequences recently became available (Hijikata *et al.*, 1987; Arakawa *et al.*, 1987). In this exceptional case the peroxisomal isoenzyme is endowed with a cleavable N-terminal extension and the mitochondrial is not. Cleavage of the extension is not directly coupled to import; moreover the amino acid sequences of the isoenzymes are only 37% identical, allowing ample opportunity for internal signal sequences. Nevertheless, the N-terminal extension of 26 amino acids is a plausible candidate for the peroxisomal signal sequence. It is therefore of interest that this extension is rich in small residues and includes a cluster of serine residues.

Materials and methods

Isolation and blotting analysis of DNA and RNA

C.fasciculata was grown in 3.7% (w/v) Brain Heart Infusion (Difco) supplemented with 20 mg/l hemin. Nuclear DNA was isolated as described by Van der Ploeg *et al.* (1982). Total RNA was isolated by LiCl precipitation (Auffray and Rougeon, 1980) and the poly(A)$^+$ fractions were purified by oligo(dT)−cellulose chromatography (Hoeijmakers *et al.*, 1980). DNA and RNA were size-fractionated and blotted onto nitrocellulose by standard procedures (Maniatis *et al.*, 1982). DNA and RNA blots were hybridized with nick-translated ^{32}P-labelled (Rigby *et al.*, 1977) probes as described by Jeffreys and Flavell (1977), with the addition of 10% (w/v) Dextran sulphate (Wahl *et al.* 1979).

Construction and screening of the Crithidia genomic library

A *Crithidia* genomic library in phage λ EMBL 3 was constructed and screened essentially as described (Frischauf *et al.*, 1983; Kaiser and Murray, 1985). Briefly, *C.fasciculata* genomic DNA was partially digested with *Sau* 3A, treated with Calf Intestine Alkaline Phosphatase and ligated onto *Bam*HI digested EMBL 3 arms. Packaging, plating and screening of the bank were standard. Post-hybridizational washes were carried out for 3 h at 65°C with several changes of 3 × SSC (SSC = 0.15 M NaCl, 0.015 M sodium citrate, pH 7.0), 0.1% SDS. Phage DNA was purified as described (Maniatis *et al.*, 1982).

DNA sequence analysis

DNA sequence analysis was performed using the dideoxy method (Sanger *et al.*, 1977) with the modifications described by Biggin *et al.* (1983). Areas with high GC content were sequenced using 7-deaza dGTP (Boehringer, Mannheim, FRG) as substitute for dGTP. The entire sequences of both genes as presented in Figure 2 have been determined on both strands. Genomic fragments from λ CgPGK-4bII (see Figure 1) were subcloned into pEMBL vectors (Dente *et al.*, 1983). Subclones were linearized at unique restriction sites in the insert and treated with Bal 31 nuclease to obtain series of progressively deleted clones. Cloning and preparation of template DNA were standard (Maniatis *et al.*, 1982; Dente *et al.*, 1983).

For several areas oligodeoxynucleotides were synthesized using the phosphotriester approach (Marugg *et al.*, 1985) on a fully automatic synthesizer (Biosearch SAM 1).

Subcellular fractionation and Western blotting

C.fasciculata cells were disrupted and fractionated as described by Steiger *et al.* (1980). Fractions were analysed by SDS−PAGE [10% (w/v)] (Laemmli, 1970) and Western blotting as described (Vaessen *et al.*, 1981). Protein concentrations were determined by the method of Bradford (1976) using lysozyme as standard.

Antiserum preparation

Rabbit antiserum against purified glycosomal PGK of *T.brucei* was prepared using standard procedures. Briefly, an intramuscular injection of 100 μg glycosomal PGK purified from *T.brucei* (Misset and Opperdoes, 1984) together with Freund's complete adjuvant, was followed by a second injection of antigen and Freund's incomplete adjuvant after 2 weeks. In weeks 5 and 6 after the initial injection, serum was collected by plasmapheresis. In Western blot experiments (Figures 5 and 6) a 1:500 dilution of the serum was used.

Two-dimensional gel electrophoresis

Two-dimensional gel electrophoresis was performed as described by O'Farrell *et al.* (1977), with non-equilibrium pH gradient gel electrophoresis in the first dimension and SDS−PAGE in the second dimension. The first dimension gel contained 1.6% Ampholines (pH 5−7) and 0.4% Ampholines (pH 3.5−10) from LKB (Uppsala, Sweden). Samples were adjusted to 9 M urea and 2% Nonidet P40, applied at the anodic side and run for 4.5 h at 800 V. Approximately 100 μg protein was layered and in the case of the mixed *Crithidia* and *T.brucei* homogenates, 75 μg of *Crithidia* proteins were mixed with 25 μg of *T.brucei* proteins. The dimensions of our SDS−PA gels are such that two first dimension tube-gels can be accommodated on a single SDS−PA gel to ensure reproducibility. SDS−PAGE, electro-blotting and immunodetection were performed as in Figure 5 (described above).

Acknowledgements

We thank Drs Onno Misset and Fred Opperdoes (International Institute for Cellular and Molecular Pathology, Brussels) for providing us with the glycosomal PGK purified from *T.brucei* and P.C.van Mourik and Dr C.J.M.Melief of the Central Laboratory Bloodtransfusion Service, Amsterdam for raising antibodies against this enzyme. We are grateful to Nico Stam (Division of Cellular Biochemistry, The Netherlands Cancer Institute, Amsterdam) for performing the first dimension in our two-

dimensional gel electrophoresis experiments. We thank Jan Kooter, Peter Laird, Joost Zomerdijk, Carsten Lincke, Anneloor ten Asbroek and Drs Hidde Ploegh, Ted White, Maarten de Bruin, Michèle Crozatier and Marc Ouellette for critical comments on the manuscript. This work was supported by The Netherlands Foundation for Chemical Research (SON) with financial aid from The Netherlands Organization for the Advancement of Pure Research (ZWO).

References

Arakawa,H., Takiguchi,M., Amaya,Y., Nagata,S., Hayashi,H. and Mori,Y. (1987) *EMBO J.*, **6**, 1361–1366.

Auffray,C. and Rougeon,F. (1980) *Eur. J. Biochem.*, **107**, 303–314.

Banks,R.D., Blake,C.C.F., Evans,P.R., Haser,R., Rice,D.W., Hardly,G.W., Merret,M. and Phillips,A.W. (1979) *Nature*, **279**, 773–777.

Biggin,M.D., Gibson,T.J. and Hong,G.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963–3965.

Borst,P. (1986) *Biochim. Biophys. Acta*, **866**, 179–203.

Bradford,M.M. (1976) *Anal. Biochem.*, **72**, 248–254.

Clayton,C.E. (1985) *EMBO J.*, **4**, 2997–3003.

Dente,L., Cesarini,G. and Cortese,R. (1983) *Nucleic Acids Res.*, **11**, 1645–1655.

Fairlamb,A.H. and Bowman,I.B.R. (1980) *Mol. Biochem. Parasitol.*, **1**, 315–334.

Frischauf,A.M., Lehrach,H., Poustka,A. and Murray,N. (1983) *J. Mol. Biol.*, **170**, 827–842.

Hijikata,M., Ishii,N., Kagamiyama,H., Osumi,T. and Hashimoto,T. (1987) *J. Biol. Chem.*, **262**, 8151–8158.

Hoeijmakers,J.H.J., Borst,P., Van den Burg,J., Weissmann,C., Cross,G.A.M. (1980) *Gene*, **8**, 391–417.

Jeffreys,A.J. and Flavell,R.A. (1977) *Cell*, **12**, 429–439.

Kaiser,K. and Murray,N.E. (1985) In Glover,D.M. (ed.), *DNA Cloning: A Practical Approach*. IRL Press, Oxford, Vol. 1, pp. 1–47.

Keller,G.-A., Gould,S., Deluca,M. and Subramani,S. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 3264–3268.

Klein,R.A. and Miller,P.G.G. (1981) *Parasitology*, **82**, 1–30.

Laemmli,U.K. (1970) *Nature*, **227**, 680–685.

Laird,P.W., Zomerdijk,J.C.B.M., De Korte,D. and Borst,P. (1987) *EMBO J.*, **6**, 1055–1062.

Lazarow,P.B. and Fujiki,Y. (1985) *Annu. Rev. Cell. Biol.*, **1**, 489–530.

Le Blancq,S.M., Swinkels,B.W., Gibson,W.C. and Borst,P. (1988) *J. Mol. Biol.*, in press.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Marrug,J.E., Piel,N., McLaughlin,L.W., Tromp,M., Veeneman,G.H., Van der Marel,G.A. and Van Boom,J.H. (1984) *Nucleic Acids Res.*, **12**, 8639–8651.

Michels,P.A.M., Polsizczak,A., Osinga,K.A., Misset,O., Van Beeumen,J., Wierenga,R.K., Borst,P. and Opperdoes,F.R. (1986) *EMBO J.*, **5**, 1049–1056.

Michelson,A.M., Markham,A.F. and Orkin,S.H. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 472–476.

Misset,O. and Opperdoes,F.R. (1984) *Eur. J. Biochem.*, **144**, 475–483.

Misset,O., Bos,O.J.M. and Opperdoes,F.R. (1986) *Eur. J. Biochem.*, **157**, 441–453.

Misset,O., van Beeumen,J., Lambeir,A.-M., Van der Meer,R. and Opperdoes,F.R. (1987) *Eur. J. Biochem.*, **162**, 501–507.

Muhich,M.L., Hughes,D.E., Simpson,A.M. and Simpson,L. (1987) *Nucleic Acids Res.*, **15**, 3141–3153.

O'Farrell,P.Z., Goodman,H.M. and O'Farrell,P.H. (1977) *Cell*, **12**, 1133–1142.

Opperdoes,F.R. (1987) *Annu. Rev. Microbiol.*, **41**, 127–151.

Opperdoes,F.R. and Borst,P. (1977) *FEBS Lett.*, **80**, 360–364.

Opperdoes,F.R., Markos,A. and Steiger,R.F. (1981) *Mol. Biochem. Parasitol.*, **4**, 291–309.

Opperdoes,F.R., Baudhuin,P., Coppens,I., De Roe,C., Edwards,S.W., Weijers,P.J. and Misset,O. (1984) *J. Cell Biol.*, **98**, 1178–1184.

Osinga,K.A., Swinkels,B.W., Gibson,W.C., Borst,P., Veeneman,G.H., Van Boom,J.H., Michels,P.A.M. and Opperdoes,F.R. (1985) *EMBO J.*, **4**, 3811–3817.

Rigby,P.W.J., Dieckmann,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.*, **113**, 237–251.

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

Scopes,R.K. (1973) In Boyer,P. (ed.), *The Enzymes*. 3rd edn. Academic Press, NY, Vol. 8, pp. 335–351.

Sloof,P., Van Den Burg,J., Voogd,A., Benne,R., Agostinelli,M., Borst,P., Gutell,R. and Noller,H. (1985) *Nucleic Acids Res.*, **13**, 4171–4190.

Steiger,R.F., Opperdoes,F.R. and Bontemps,J. (1980) *Eur. J. Biochem.*, **105**, 163–175.

Swinkels,B.W., Gibson,W.C., Osinga,K.A., Kramer,R., Veeneman,G.H., Van Boom,J.H. and Borst,P. (1986) *EMBO J.*, **5**, 1291–1298.

Taylor,M.B., Berghausen,H., Heyworth,P., Messenger,N., Rees,L.J. and Gutterridge,W.E. (1980) *Int. J. Biochem.*, **11**, 117–120.

Vaessen,R.T.M.J., Kreike,J. and Groot,G.S.P. (1981) *FEBS Lett.*, **124**, 193–196.

Van der Ploeg,L.H.T., Bernards,A., Rijsewijk,F.A.M. and Borst,P. (1982) *Nucleic Acids Res.*, **10**, 593–609.

Von Heijne,G. (1985) *J. Mol. Biol.*, **184**, 99–105.

Von Heijne,G. (1986) *EMBO J.*, **5**, 1335–1342.

Wahl,G.M., Stern,M. and Stark,P.G.R. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 3683–3687.

Watson,H.L., Walker,N.P.C., Shaw,P.J., Bryant,T.N., Wendell,P.L., Fothergill,L.A., Perkins,R.E., Conroy,S.C., Dobson,M.J., Tuite,M.F., Kingsman,A.J. and Kingsman,S.M. (1982) *EMBO J.*, **1**, 1635–1640.

Wierenga,R.K., Swinkels,B.W., Michels,P.A.M., Osinga,K.A., Misset,O., Van Beeumen,J., Gibson,W.C., Postma,J.P.M., Borst,P., Opperdoes,F.R. and Hol,W.G.J. (1987) *EMBO J.*, **6**, 215–221.

Received on December 11, 1987; revised on January 18, 1988