FEBS Letters 350 (1994) 125–129

FEBS 14360

Phosphoenolpyruvate carboxykinase of *Trypanosoma brucei* is targeted to the glycosomes by a C-terminal sequence

Jürg M. Sommer, Tien T. Nguyen, C.C. Wang*

Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, CA 94143-0446, USA

Received 27 April 1994; revised version received 29 June 1994

Abstract

Import of proteins into the glycosomes of *T. brucei* resembles the peroxisomal protein import in that C-terminal SKL-like tripeptide sequences can function as targeting signals. Many of the glycosomal proteins do not, however, possess such C-terminal tripeptide signals. Among these, phosphoenolpyruvate carboxykinase (PEPCK (ATP)) was thought to be targeted to the glycosomes by an N-terminal or an internal targeting signal. A limited similarity to the N-terminal targeting signal of rat peroxisomal thiolase exists at the N-terminus of *T. brucei* PEPCK. However, we found that this peroxisomal targeting signal does not function for glycosomal protein import in *T. brucei*. Further studies of the PEPCK gene revealed that the C-terminus of the predicted protein does not correspond to the previously deduced protein sequence of 472 amino acids due to a −1 frame shift error in the original DNA sequence. Readjusting the reading frame of the sequence results in a predicted protein of 525 amino acids in length ending in a tripeptide serine-arginine-leucine (SRL), which is a potential targeting signal for import into the glycosomes. A fusion protein of firefly luciferase, without its own C-terminal SKL targeting signal, and *T. brucei* PEPCK is efficiently imported into the glycosomes when expressed in procyclic trypanosomes. Deletion of the C-terminal SRL tripeptide or the last 29 amino acids of PEPCK reduced the import only by about 50%, while a deletion of the last 47 amino acids completely abolished the import. These results suggest that *T. brucei* PEPCK may contain a second, internal glycosomal targeting signal upstream of the C-terminal SRL sequence.

Key words: PEPCK; Glycosome; Peroxisome; Protein targeting; *Trypanosoma brucei*; *Trypanosoma cruzi*

1. Introduction

In the bloodstream form of *Trypanosoma brucei*, which lacks a fully functional mitochondrion, energy is supplied primarily by glycolysis. Most of the glycolytic enzymes are confined within a set of membrane-bound organelles, designated the glycosomes [1]. Morphologically, the glycosomes closely resemble the peroxisomes found in higher eukaryotes, although their enzyme contents differ significantly [2]. Recent studies indicated that the import of proteins into the glycosome depends on a conserved C-terminal tripeptide motif, analogous to the C-terminal serine-lysine-leucine (SKL) sequence responsible for peroxisomal targeting of firefly luciferase [3,4]. Subsequent investigations showed that the C-terminal tripeptide sequence requirements for glycosomal import are considerably more relaxed than those for import of proteins into mammalian peroxisomes [3–5]. Among the seven glycosomal proteins of *T. brucei* that have been sequenced, three are likely to have a C-terminal tripeptide targeting signal. Glyceraldehyde-3-phosphate dehydrogenase and glucosephosphate isomerase end in AKL and SHL, respectively [6,7], while 3-phosphoglycerate kinase uses a C-terminal SSL sequence as targeting signal [5]. However, the other four glycosomal proteins, phosphoenolpyruvate carboxykinase (PEPCK) [8,9], triosephosphate isomerase [10], the 56 kDa phosphoglycerate kinase [11] and aldolase [12] do not end in a tripeptide sequence that could qualify as a glycosomal targeting signal. These proteins must depend on other import signals in order to enter the glycosomes.

A small number of peroxisomal proteins are targeted to the microbody by peptide sequences near the N-terminus of the protein [13–16]. Among these, rat 3-ketoacyl-CoA thiolase is targeted to the mammalian peroxisome by a cleavable N-terminal presequence [13,17,18], apparently using an import mechanism that is distinct from the C-terminal tripeptide recognition system [19]. Swinkels et al. [13] have shown that the first 11 amino acids of the rat thiolase B isozyme precursor are sufficient to target the cytosolic protein chloramphenicol acetyltransferase (CAT) to the peroxisome when present at the N-terminus. Since a limited similarity to this thiolase presequence exists near the N-termini of *T. brucei* aldolase and PEPCK, we tested the possibility that the same thiolase targeting signal may also function in glycosomal protein import. The 11 amino acids of the thiolase were thus added to the N-termini of chloramphenicol acetyltransferase (CAT) and β-glucuronidase (GUS), but these fusion proteins failed to enter the glycosomes in vivo, suggesting that the thiolase N-terminal signal does not function in *T. brucei*.

The *T. brucei* PEPCK gene reportedly encodes a protein of 472 amino acids with a molecular mass of 52.5 kDa [8,9]. However, during our studies we found an apparent sequencing error in the PEPCK gene. Readjusting the

*Corresponding author. Fax: (1) (415) 476 3382.

SSDI 0014-5793(94)00747-0

reading frame of the sequence results in a predicted protein of 525 amino acids ending in the tripeptide SRL. Further examinations verified that this is indeed the major glycosomal targeting signal of PEPCK, although the upstream sequence of about 50 amino acids may also play an important role in glycosomal import.

2. Materials and methods

2.1. N-terminal thiolase fusion constructs

The thiolase–CAT fusion construct was made by PCR amplification of the CAT gene in the plasmid pJP25 [3] with an N-terminal PCR primer encoding the first 15 amino acids of the rat 3-ketoacyl-CoA thiolase B gene [20]. The thiolase–GUS fusion construct was made by PCR amplification of the GUS gene in pGUSN358→S (Clontech Laboratories, Inc., Palo Alto, CA) with a 5′ PCR primer in which the start codon of the GUS gene was replaced by the first 11 amino acids of the rat thiolase gene. Both PCR fragments were inserted into the cloning site of the TA vector (Invitrogen, San Diego, CA) and subsequently transferred as *Hin*dIII to *Bam*HI fragments into the *T. brucei* expression vector pTSA-HYG2 [5].

2.2. PEPCK fusion constructs

The *T. brucei* PEPCK gene, a generous gift from Dr. Thomas Seebeck, was purified as an *Xho*I fragment from the plasmid pTbp60B [8] and inserted into pBluescript (pBS+) (Stratagene, San Diego, CA), with the downstream *Hin*dIII site of PEPCK closest to the *Hin*dIII site in pBS. A *Not*I site was introduced at the start of the PEPCK coding region by site-directed mutagenesis (Amersham System Version 2.0, Amersham, Arlington Height, IL.). The open reading frame of the luciferase gene in the expression vector pLUH207 lacks the last nine nucleotides encoding the SKL tripeptide (luciferase⊿SKL), which were replaced by *Not*I, *Stu*I and *Sna*BI sites, followed by stop codons in each reading frame. pLUH207 is essentially identical to pLUH206 [5], except that a four base pair GATC sequence immediately preceding the luciferase coding region was removed in pLUH207 by site-directed mutagenesis for the construction of some of the N-terminal fusions. The 1,440 bp *Not*I–*Hin*dIII fragment containing the first 478 amino acids of PEPCK gene was inserted downstream of the luciferase gene in pLUH207, cut with *Not*I and *Stu*I, resulting in the luciferase⊿SKL–PEPCK478. The truncated PEPCK gene is immediately followed by a stop codon in the vector sequence. To produce the luciferase–PEPCK fusion gene containing the full length PEPCK, the 1,875 bp *Not*I–*Dra*I fragment was inserted into pLUH207 cut with *Not*I and *Stu*I. To delete the C-terminal SRL sequence from the PEPCK protein, the full-length gene was amplified by the polymerase chain reaction (PCR) using a mutagenic C-terminal primer which introduced a stop codon at amino acid position 523. The PCR fragment was then used to replace the PEPCK gene in the above luciferase⊿SKL–PEPCK fusion construct. The luciferase⊿SKL–PEPCK496 construct was made by truncating the PEPCK gene at a unique *Ear*I site, filling in the recessed end with Klenow and ligation to the downstream *Stu*I site. This resulted in addition of a proline and a threonine residue between phenylalanine 496 of PEPCK and the stop codon.

2.3. Transformation of T. brucei and subcellular localization of reporter protein

Linearized plasmid DNA was introduced into procyclic *T. brucei* TREU667 cells by electroporation [3], and stable transformants were isolated as described previously [3]. Subcellular localization of the luciferase or GUS reporter proteins was determined by differential digitonin-solubilization of either transiently transfected cells or stable transformants by a previously described procedure [3, 5].

For the Proteinase K treatment, approximately 5×10^6 *T. brucei* procyclic cells were treated with 4 μg digitonin in a 200 μl volume of HEDS + KOAc buffer [21]. Proteinase K was added to a final concentration of 250 μg/ml and the mixture was incubated at room temperature for 30 min. Phenylmethanesulfonyl fluoride (PMSF) was added to 2 mM and the remaining luciferase activity was assayed immediately.

3. Results and discussion

3.1. The N-terminal presequence of peroxisomal thiolase does not function in glycosomal import

We added the 11 amino acid portion of the rat 3-ketoacyl-CoA thiolase presequence (−26 to −16) to the N-termini of CAT and GUS. The resulting thiolase–CAT fusion protein was thus identical to that shown to be imported into the mammalian peroxisome by Swinkels et al. [13]. The subcellular distribution of the fusion proteins was determined by assaying for CAT or GUS activity in the supernatant and pellet fractions following in vivo expression of the fusion genes in transformed procyclic trypanosomes and solubilization of the plasma membrane by digitonin [3,5]. These experiments showed that in both cases the enzyme activity was located in the cytoplasm (Data not shown), leading us to conclude that the thiolase N-terminal peroxisomal targeting signal does not function or is not sufficient for import of proteins into the glycosomes.

3.2. Fusion constructs containing the PEPCK protein at the N-termini are not expressed in T. brucei

To test for the presence of a potential N-terminal targeting signal in *T. brucei* PEPCK, we made fusion constructs containing the 472 amino acids of the PEPCK protein at the N-termini of GUS and firefly luciferase,

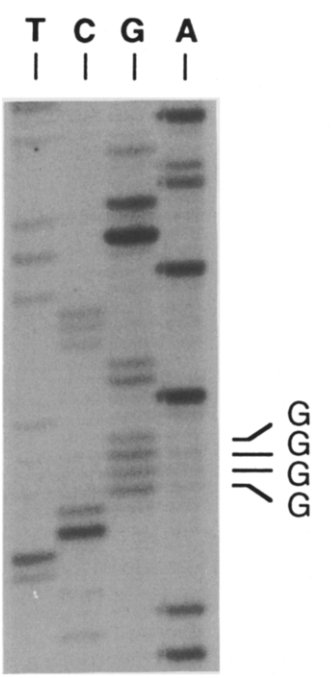

Fig. 1. DNA sequence surrounding amino acid position 458 of the *T. brucei* PEPCK gene. Four sequential guanosine residues were resolved by double-stranded sequencing of plasmid DNA using a thermostable polymerase and high temperature sequencing (Circumvent, New England Biolabs, Beverly, MA). The sequence from top to bottom specifies TAT (Tyr^{451}) GAA (Glu^{452}) GTG (Val^{453}) TAT (Tyr^{454}) CCC (Pro^{455}) GGA (Gly^{456}) TGG (Trp^{457}) GGC (Gly^{458}) CTT (Leu^{459}) and CAC (His^{460}).

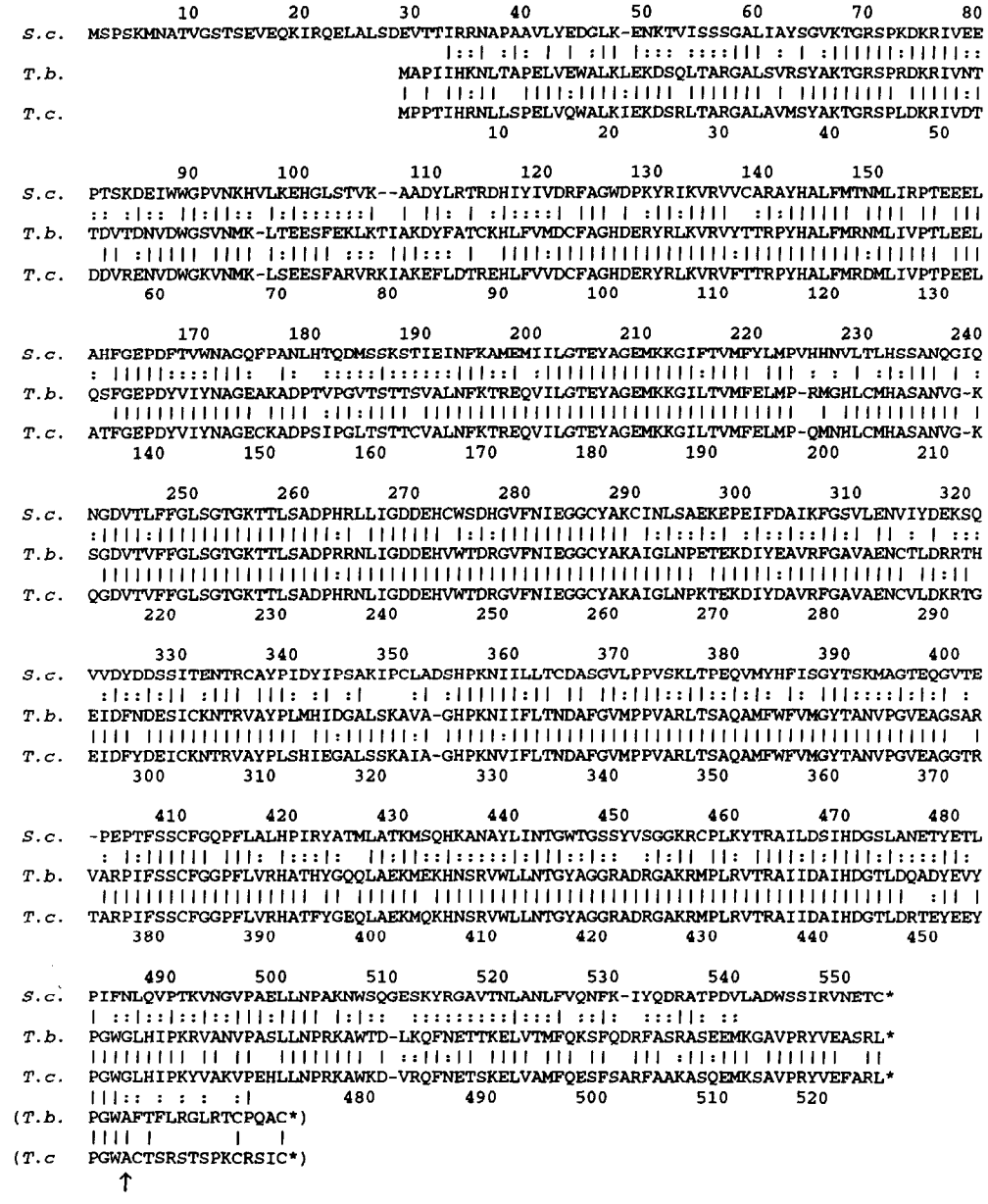

Fig. 2. Amino acid sequence alignment of PEPCK from *Saccharomyces cerevisiae, T. brucei* and *T. cruzi*. Optimal alignment of the *S. cerevisiae* and *T. brucei* PEPCK (ATP) protein sequences (GenBank Accession Nos. X1306 and A33275) was computed at the EMBL (Heidelberg) using the FASTA network service. Alignment of the *T. brucei* and *T. cruzi* (GenBank Accession No. A33275) sequences was performed using MBIR software. The C-termini of the published *T. brucei* and *T. cruzi* sequences are indicated in parentheses. The arrow indicates the position of the frame shift.

which had its own C-terminal SKL targeting signal deleted (luciferaseΔSKL). The in vivo expression of these fusion constructs did not result in any measurable GUS or luciferase activity, and we were unable to detect the fusion proteins by Western blotting (data not shown). Further examination of these constructs showed that the downstream reporter proteins were not in the same reading frame as the PEPCK protein (see below).

3.3. *Determination of the C-terminal sequence of PEPCK*

The originally predicted *T. brucei* PEPCK protein of 472 amino acids ends in the tripeptide glutamine-alanine-cysteine [8,9], which is unlikely to function as a glycosomal import signal by our previous analysis of tripeptide targeting signals [3]. However, upon resequencing of the PEPCK gene, corresponding to amino acid position 458, we found four sequential guanosine residues instead of three, as previously reported (Fig. 1). This shift in the reading frame at the C-terminus of PEPCK thus explains the lack of expression or activities of the reporter enzymes from the PEPCK-luciferase fusion constructs. The length of the newly deduced PEPCK protein increases from 472 to 525 amino acids with a calculated molecular mass of 58.6 kDa, which is close to the 60 kDa molecular

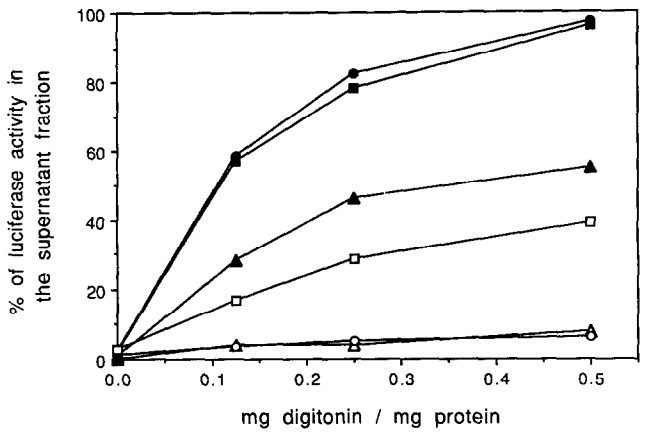

Fig. 3. Digitonin solubilization of luciferase–PEPCK fusion proteins. Cell cultures were transiently transfected with plasmids encoding either wild type luciferase (○), luciferaseΔSKL (●), the fusion proteins luciferaseΔSKL–PEPCK525 (△), luciferaseΔSKL–PEPCK522 (▲), luciferaseΔSKL–PEPCK496 (□) or luciferaseΔSKL–PEPCK478 (■), solubilized at increasing concentrations of digitonin, centrifuged and assayed for luciferase activity in the pellet and supernatant fractions.

weight of *T. brucei* PEPCK estimated by SDS-polyacrylamide gel electrophoresis [8]. The newly derived protein sequence following amino acid 458 also shows significant homology to the PEPCK (ATP) of *Saccharomyces cerevisiae* (Fig. 2). The longer protein ends in the tripeptide serine-arginine-leucine (SRL), which had been previously identified as a functional glycosomal targeting signal when present at the C-terminus of luciferase [3].

3.4. The luciferase–PEPCK fusion protein is imported into the glycosomes

Following in vivo expression of a plasmid construct containing the full length PEPCK gene sequence downstream of luciferase (lucΔSKL–PEPCK525), more than 90% of the luciferase activity remained in the pellet fraction at digitonin concentrations which are known to completely solubilize the cytoplasmic proteins (Fig. 3), indicating possible import of the fusion protein into the glycosomes. The wild type luciferase was also found in the pellet fraction under the same conditions (Fig. 3), and we have previously shown by electron microscopy that this protein is imported into the glycosomes [3,5].

Proteins that are inside an intact, membrane-bound organelle are expected to be protected from externally added proteases. We thus treated the digitonin-permeabilized procyclic transformants with Proteinase K at 250 μg/ml, which inactivates native luciferase and luciferase fusion proteins in the presence of 0.2% Triton X-100 (Fig. 4). In each case, the digitonin-insoluble luciferase activity was highly resistant to Proteinase K treatment, but completely lost when Triton X-100 was included (Fig. 4). The luciferaseΔSKL–PEPCK525 fusion protein was equally resistant to Proteinase K as the wild type luciferase, and also became susceptable to Proteinase K digestion in the presence of Triton X-100. It is thus most likely imported into the *T. brucei* glycosomes.

3.5. The C-terminal tripeptide of PEPCK is required for efficient glycosomal import

In order to show that the C-terminal SRL sequence of PEPCK is specifically required for glycosomal targeting, we expressed a fusion construct lacking the SRL sequence (luciferaseΔSKL–PEPCK522). In this case, about 50% of the luciferase activity was digitonin soluble, indicating that it was present in the cytoplasm, while the remainder may be imported into the glycosome (Fig. 3). Thus, the C-terminal SRL sequence of *T. brucei* PEPCK is indeed required for efficient import of the protein into the glycosomes. The digitonin-insoluble fraction of the luciferase activity was also resistant to Proteinase K digestion (Fig. 4), confirming that it was most likely inside the glycosome. These results suggest that *T. brucei* PEPCK may contain another targeting signal in addition to its C-terminal SRL sequence. In our earlier efforts to produce a luciferaseΔSKL–PEPCK fusion construct, we made use of a *Hin*dIII endonuclease restriction site located 8 nt downstream of the wrongly presumed stop codon, which resulted in a truncation of the actual C-terminus of PEPCK by 47 amino acids. This fusion protein (luciferaseΔSKL–PEPCK478) was found exclusively in the cytoplasm of *T. brucei* and was sensitive to Proteinase K. An additional construct lacking

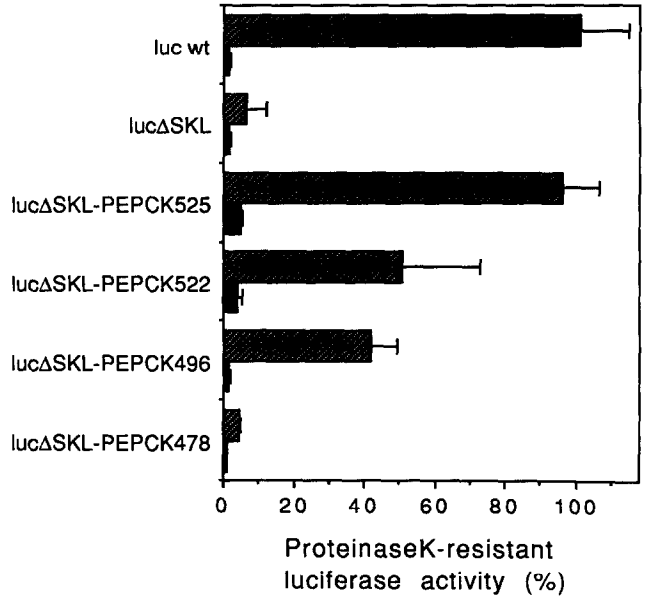

Fig. 4. Proteinase K treatment of digitonin-permeabilized transformants expressing luciferase fusion proteins. Stable procyclic transformants were incubated with 0.25 mg per ml Proteinase K in the presence of digitonin (hatched bars) or digitonin and Triton X-100 (solid bars) for 30 min at room temperature. In each case, the remaining luciferase activity was plotted as percent of enzyme activity in the absence of Proteinase K.

only the last 29 amino acids of PEPCK (luciferase⊿SKL–PEPCK496) is, however, partially imported into the glycosomes (Figs. 3 and 4). This observation raises the possibility that essential elements of a second targeting signal may be contained between amino acids 478 and 496 of PEPCK, approximately 40 amino acids upstream of the C-terminal SRL tripeptide. However, it is also possible that some of these deletions affect the overall structure of the protein and thus alter context- or structure- dependent targeting signals present in the remainder of the protein [22]. The possibility that the fusion proteins containing 496 and 522 amino acids of PEPCK are partially digitonin-insoluble and protease-resistant due to aggregation is unlikely, since they are completely soluble and digestable by Proteinase K at low concentrations of Triton X-100 and preliminary immunoelectron microscopy data indicates that much of the luciferase⊿SKL–PEPCK522 is found inside the glycosomes (Data not shown). Catalase A, a major peroxisomal protein of *Saccharomyces cerevisiae*, was found to contain two independently functioning targeting signals, one of which is an internal signal, which could not be precisely mapped, and the other a C-terminal SKF sequence [22]. Further investigations with localized mutagenesis of the PEPCK protein will determine if the sequence between amino acids 478 and 496 indeed contains a genuine glycosomal targeting signal.

The gene encoding PEPCK from a related trypanosomatid species, *Trypanosoma cruzi*, has also been cloned and sequenced recently [23]. Interestingly, the deduced amino acid sequence of the *T. cruzi* PEPCK protein appears to contain the same frame shift error at amino acid position 458, resulting from a missing guanosine nucleotide in the DNA sequence. By inserting an extra guanosine at the position corresponding to amino acid 458 in the *T. cruzi* gene, the predicted protein sequence also extends to a length of 525 amino acids. Alignment of the corrected *T. brucei* and *T. cruzi* protein sequences restores their C-terminal homology to 74% identity, with 85% identity over the entire length of the two PEPCK proteins (Fig. 2). These changes do not affect the consensus phosphate and adenine binding sites around amino acids 224 and 333, or the putative divalent cation binding site at position 253 of the PEPCK (ATP) [23]. The revised *T. cruzi* PEPCK protein sequence ends in the tripeptide ARL, which is also expected to be responsible for its glycosomal import in *T. cruzi* [3].

In summary, we have shown that the N-terminal presequence of peroxisomal thiolase does not function for import of proteins into *T. brucei* glycosomes, although this N-terminal peroxisomal targeting signal has been well conserved between mammalian cells, plants and fungi [17,18,24,25]. However, we cannot rule out the possibility that a related variant of this signal also functions for import of some proteins into the glycosomes of *T. brucei* . The PEPCK genes of *T. brucei* and *T. cruzi* each encode proteins of 525 amino acids with calculated molecular masses of 58.6 and 58.9 kDa, respectively. The SRL and ARL sequences at the C-termini of these proteins are most likely involved in the glycosomal targeting of these enzymes, but *T. brucei* PEPCK may also contain a second, internal targeting signal, about 40 amino acids upstream of the C-terminus.

Acknowledgements: We thank Dr. T. Seebeck for providing the *T. brucei* PEPCK gene. This work was supported by Grant AI-21782 from the National Institutes of Health.

References

[1] Opperdoes, F.R. and Borst, P. (1977) FEBS Lett. 80, 360–4.
[2] Sommer, J.M. and Wang, C.C. (1994) Annu. Rev. Microbiol. 48, 105–38.
[3] Sommer, J.M., Cheng, Q.L., Keller, G.A. and Wang, C.C. (1992) Mol. Biol. Cell. 3, 749–59.
[4] Blattner, J., Swinkels, B., Dorsam, H., Prospero, T., Subramani, S. and Clayton, C. (1992) J. Cell Biol. 119, 1129–36.
[5] Sommer, J.M., Peterson, G., Keller, G.A., Parsons, M. and Wang, C.C. (1993) FEBS Lett. 316, 53–8.
[6] Michels, P.A., Poliszczak, A., Osinga, K.A., Misset, O., Van Beeumen, J., Wierenga, R.K., Borst, P. and Opperdoes, F.R. (1986) EMBO J. 5, 1049–56.
[7] Marchand, M., Kooystra, U., Wierenga, R.K., Lambeir, A.M., Van Beeumen, J., Opperdoes, F.R. and Michels, P.A. (1989) Eur. J. Biochem. 184, 455–64.
[8] Kueng, V., Schlaeppi, K., Schneider, A. and Seebeck, T. (1989) J. Biol. Chem. 264, 5203–9.
[9] Parsons, M. and Smith, J.M. (1989) Nucleic Acids Res. 17, 6411.
[10] 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, 1291–8.
[11] Alexander, K. and Parsons, M. (1991) Mol. Biochem. Parasitol. 46, 1–10.
[12] Clayton, C.E. (1985) EMBO J. 4, 2997 3003.
[13] Swinkels, B.W., Gould, S.J., Bodnar, A.G., Rachubinski, R.A. and Subramani, S. (1991) EMBO J. 10, 3255–62.
[14] Osumi, T., Tsukamoto, T., Hata, S., Yokota, S., Miura, S., Fujiki, Y., Hijikata, M., Miyazawa, S. and Hashimoto, T. (1991) Biochem. Biophys. Res. Commun. 181, 947–54.
[15] Gietl, C. (1990) Proc. Natl. Acad. Sci. USA 87, 5773–7.
[16] de Hoop, M.J. and Ab, G. (1992) Biochem. J. 657–69.
[17] Glover, J.R., Andrews, D.W., Subramani, S. and Rachubinsky, R.A. (1994) J. Biol. Chem. 269, 7558–63.
[18] Tsukamoto, T., Hata, S., Yokota, S., Miura, S., Fujiki, Y., Hijikata, M., Miyazawa, S., Hashimoto, T. and Osumi, T. (1994) J. Biol. Chem. 269, 6001–10.
[19] McCollum, D., Monosov, E. and Subramani, S. (1993) J. Cell. Biol. 121, 761–74.
[20] Hijikata, M., Wen, J.K., Osumi, T. and Hashimoto, T. (1990) J. Biol. Chem. 265, 4600–6.
[21] Sommer, J.M., Thissen, J.A., Parsons, M. and Wang, C.C. (1990) Mol. Cell. Biol. 10, 4545–54.
[22] Kragler, F., Langeder, A., Raupachova, J., Binder, M. and Hartig, A. (1993) J. Cell. Biol. 120, 665–73.
[23] Linss, J., Goldenberg, S., Urbina, J.A. and Amzel, L.M. (1993) Gene 136, 69–77.
[24] van der Klei, I.J., Faber, K.N., Keizer, G.I., Gietl, C., Harder, W. and Veenhuis, M. (1993) FEBS Lett. 334, 128–32.
[25] Gietl, C., Faber, K.N., van der Klei, I.J. and Veenhuis, M. (1994) Proc. Natl. Acad. Sci. USA 91, 3151–5.