The EMBO Journal vol.6 no.1 pp.215–221, 1987

Common elements on the surface of glycolytic enzymes from *Trypanosoma brucei* may serve as topogenic signals for import into glycosomes

R.K.Wierenga, B.Swinkels[1], P.A.M.Michels[2], K.Osinga[1,4], O.Misset[2,4], J.Van Beeumen[3], W.C.Gibson[1,5], J.P.M.Postma, P.Borst[1], F.R.Opperdoes[2] and W.G.J.Hol

Laboratory of Chemical Physics, Department of Chemistry, University of Groningen, 9747 AG Groningen, [1]The Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands, [2]Research Unit for Tropical Diseases, International Institute of Cellular and Molecular Pathology, 1200 Brussels, and [3]Laboratory of Microbiology and Microbial Genetics, State University of Ghent, 9000 Ghent, Belgium

[4]Present address: Gist Brocades N.V., PO Box 1, 2600 MA Delft, The Netherlands

[5]Present address: Unit of Molecular Genetics, University of Bristol, Bristol BS8 1UG, UK

Communicated by C.de Duve

In *Trypanosoma brucei*, a major pathogenic protozoan parasite of Central Africa, a number of glycolytic enzymes present in the cytosol of other organisms are uniquely segregated in a microbody-like organelle, the glycosome, which they are believed to reach post-translationally after being synthesized by free ribosomes in the cytosol. In a search for possible topogenic signals responsible for import into glycosomes we have compared the amino acid sequences of four glycosomal enzymes: triosephosphate isomerase (TIM), glyceraldehyde-phosphate dehydrogenase (GAPDH), phosphoglycerate kinase (PGK) and aldolase (ALDO), with each other and with their cytosolic counterparts. Each of these enzymes contains a marked excess of positive charges, distributed in two or more clusters along the polypeptide chain. Modelling of the three-dimensional structures of TIM, PGK and GAPDH using the known structural coordinates of homologous enzymes from other organisms indicates that all three may have in common two 'hot spots' about 40 Å apart, which themselves include a pair of basic amino acid residues separated by a distance of about 7 Å. The sequence of glycosomal ALDO, for which no three-dimensional information is available, is compatible with the presence of the same configuration on the surface of this enzyme. We propose that this feature plays an essential role in the import of enzymes into glycosomes.

Key words: glycolysis/Kinetoplastida/microbody/topogenesis/three-dimensional structure

Introduction

The hemoflagellate protozoan parasites of the *Trypanosoma brucei* complex are the causative agents of human trypanosomiasis, or sleeping sickness, and of nagana in cattle in sub-Saharan Africa (Molyneux and Ashford, 1983).

A unique feature of all Trypanosomatidae is the presence of glycosomes, microbody-like organelles, which are vital for the energy supply of the bloodstream-form trypanosome (Opperdoes and Borst, 1977; Fairlamb *et al.*, 1977; Opperdoes, 1985). The main constituents of glycosomes are nine enzymes involved in glucose and glycerol metabolism (Opperdoes, 1985).

Recent investigations on three glycosomal enzymes have shown that these proteins are made in the cytosol on free ribosomes, and are subsequently imported into the organelle without detectable size changes (Opperdoes *et al.*, 1986; Hart *et al.*, unpublished results), suggesting that the topogenic signals for import are part of the mature proteins. Initial studies on the glycosomal (g) and cytosolic (c) phosphoglycerate kinase (PGK) from *T. brucei* have revealed a remarkable difference in charge between these two enzymes with only a limited difference in sequence. We have, therefore, proposed (Osinga *et al.*, 1985) that this charge difference is related to the different localization of the two enzymes within the cell. This suggestion was subsequently supported by further studies showing that several other glycosomal enzymes have a distinctly higher isoelectric point than their cytosolic homologues in other species (Misset *et al.*, 1986).

In the present investigation sequencing results, crystallographic data and modelling of the sequences into known three-dimensional structures of homologous animal or bacterial enzymes have been used in a search for a common configuration of positive residues that might serve as topogenic signal. The results indicate that this configuration could consist of two positively charged patches ('hot spots'), each including a pair of positive residues about 7 Å apart, separated by a distance of about 40 Å on the surface of the molecules.

Results

Amino acid sequences

The amino acid sequences derived from the gene sequences of g-triosephosphate isomerase (TIM) (Swinkels *et al.*, 1986), g-PGK (Osinga *et al.*, 1985), c-PGK (Osinga *et al.*, 1985) and g-glyceraldehyde-phosphate dehydrogenase (GAPDH) (Michels *et al.*, 1986) have already been published by us while the sequence of g-fructose bisphosphate aldolase (ALDO) has been published by Clayton (1985) and found to be essentially identical in our laboratory. A comparison of their subunit size with that of their mammalian counterparts shows that three out of four enzymes (g-PGK, g-GAPDH and g-ALDO) are between 1 and 5 kd larger (Misset *et al.*, 1986). As shown in Figure 1, proper alignment reveals ~50% homology between the glycosomal sequences and their human counterparts, as well as (not shown) other known prokaryotic and eukaryotic sequences. In these alignments, insertions and deletions were placed outside regions of secondary structure whenever possible. It is striking that each of the trypanosomal enzymes shown in Figure 1 has at least one insertion that does not occur in the human nor in any of the other known sequences.

None of the four glycosomal enzymes show evidence for the presence of a cleavable N-terminal leader sequence when compared with the other members within the respective enzyme families (Figure 1). This is in agreement with the following observations. (i) No evidence has been found for any significant change in size upon import of g-ALDO and g-GAPDH into glycosomes (Hart *et al.*, unpublished results; Opperdoes *et al.*, 1986). (ii) For

g-PGK, g-TIM, g-GAPDH and g-ALDO there is an excellent agreement between the mol. wts calculated from the gene sequences and those determined for the purified enzymes (Misset *et al.*, 1986a). (iii) For g-GAPDH any N-terminal processing was ruled out because of the perfect match of the gene and protein sequences at the N-terminus (Michels *et al.*, 1986), while for g-TIM such processing is highly unlikely because of the excellent agreement between electron-density distribution and gene sequence. Only the first three N-terminal residues could not be seen in the electron-density map. (iv) Although Clayton (1985) postulated the presence of a 10-amino acid N-terminal leader sequence for g-ALDO, we come to a different conclusion. In our alignment of g-ALDO there is around the N-terminus a good agreement between the first 10 residues of g-ALDO and human aldolase. Two residues are unchanged and five substitutions are conservative, leaving as only significant differences Ser-1, Glu-5 and Leu-8, replaced by Ala, Pro and Thr, respectively, in the human sequence. We, therefore, conclude that g-ALDO has no N-terminal extension, but instead a 10-amino acid insertion starting at position 11 (Figure 1).

It is also unlikely that the C terminus of the glycosomal enzymes contains a common import signal. Only two of the four enzymes, g-PGK and g-GAPDH, have C-terminal extensions which are respectively 20 and five amino acids in length. There is no significant sequence similarity, either between these two, or with the C termini or internal regions of the other two glycosomal enzymes.

Net charges of glycosomal enzymes

Another characteristic property shared by each of the four sequenced glycosomal enzymes is that each has by far the highest calculated positive charge within its family of homologous proteins (Figure 2). These calculated net charges are in good agreement with the isoelectric points of the four purified enzymes, which are >9 in all cases (Misset *et al.*, 1986).

These observations gave us a first hint as to the possible nature of the topogenic signal. Since it is unlikely that a high overall positive charge alone would ensure effective targetting into a glycosome, we investigated the spatial distribution of these positive charges.

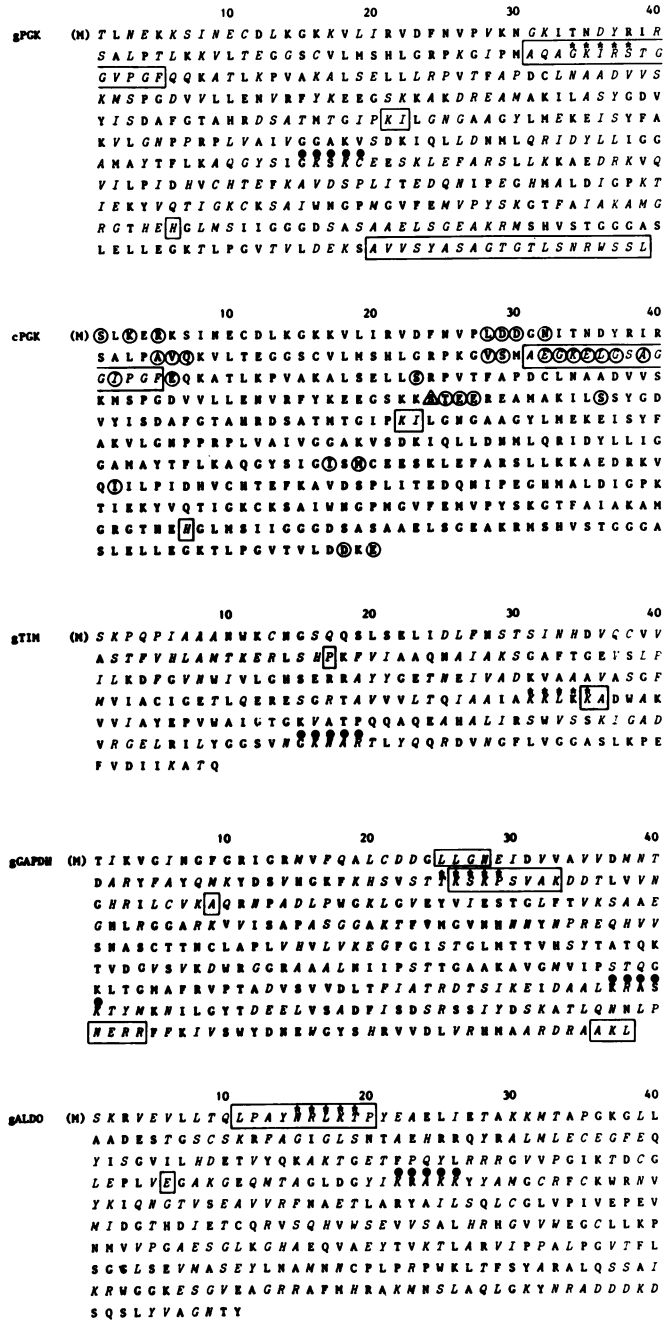

Fig. 1. The sequences of five trypanosomal glycolytic enzymes. The sequences have unique insertions when compared with all other homologous enzymes; these unique insertions are printed within boxes. The amino acids of hot spot sequence I (see text) are emphasized by asterisks. The hot spot II amino acids are indicated by dots. In all four glycosomal sequences (g-PGK, g-TIM, g-GAPDH, g-ALDO) residues that differ from the human sequence are written in italics. The g-PGK sequence is 47% identical with the human liver PGK sequence (Michelson *et al.*, 1983). The c-PGK sequence is 46% identical with the human liver PGK sequence (Michelson *et al.*, 1983), while it is 93% identical with the g-PGK sequence (Osinga *et al.*, 1985). The residue in the triangle is a one-residue insertion in c-PGK which does not occur in g-PGK. Residues that are different in c-PGK from g-PGK are encircled. The g-TIM sequence is 52% identical with the human liver TIM sequence (Maquat *et al.*, 1985). The g-GAPDH sequence is 53% identical with the human GAPDH sequence (Tso *et al.*, 1985). The g-ALDO sequence is 47% identical with the human ALDO sequence (Rottman *et al.*, 1984).

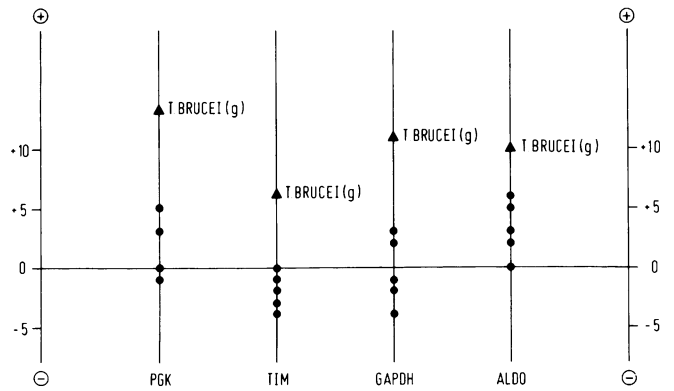

Fig. 2. Overall charges of four glycolytic enzymes. The scale of the overall charge is shown on the vertical axes at left and right of the drawing. The charges of Asp and Glu are taken as −1; Lys and Arg as +1; His and Cys as 0. The overall charges of the glycosomal enzymes from *T. brucei* are indicated by the closed triangles. The overall charges (Q) of the other enzymes have been calculated from the following sequences. PGK: Q = +5 (horse), Q = +3 (human), Q = 0 (yeast), Q = −1 (*T. brucei*, cytosolic). TIM: Q = 0 (rabbit), Q = −1 (chicken), Q = −1 (human liver), Q = −2 (human placenta), Q = −3 (coelacanth), Q = −4 (yeast). GAPDH: Q = +3 (yeast), Q = +3 (chicken), Q = +2 (rat), Q = +2 (human), Q = +2 (*Drosophila melanogaster*), Q = −1 (*Escherichia coli*), Q = −2 (pig), Q = −2 (lobster). ALDO: Q = +6 (chicken), Q = +5 (rat), Q = +3 (rabbit), Q = +2 (human), Q = 0 (*D. melanogaster*). For details on the sequences analyzed see Osinga *et al.* (1985), Swinkels *et al.* (1986), Michels *et al.* (1986) and Clayton (1985) and the references cited therein.

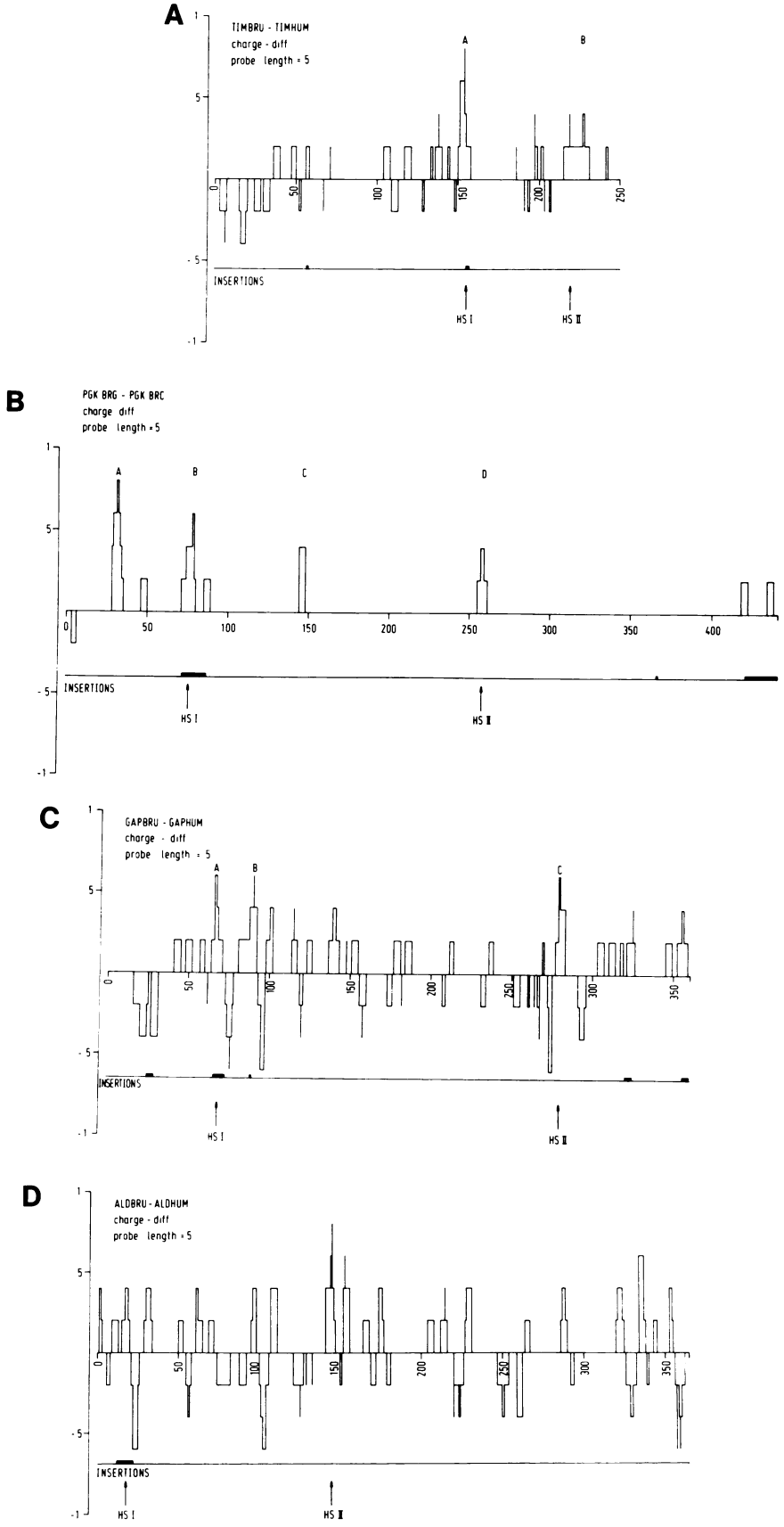

Fig. 3. Charge-difference profiles for four glycosomal enzymes. The charge differences calculated as described under Materials and methods are plotted on the vertical axis for the central position of the probe. Position numbers of the glycosomal sequences are given on the horizontal axis. The unique insertions of trypanosomal enzymes are indicated by black bars on the lower line of the figure. (**A**) g-TIM minus human TIM (Swinkels *et al.*, 1986; Maquat *et al.*, 1985). (**B**) g-PGK minus cPGK (Osinga *et al.*, 1985). (**C**) g-GAPDH minus human GAPDH (Michels *et al.*, 1986; Tso *et al.*, 1985). (**D**) g-ALDO minus human ALDO (Clayton, 1985; Rottman *et al.*, 1984).

Distribution of charges

Figure 3 shows the differential charge distribution profiles of g-TIM, g-GAPDH and g-ALDO versus their human homologues, and of g-PGK versus c-PGK. The extra positive charges of the glycosomal enzymes tend to be congregated in two or more clusters, of which one is always associated with an insertion, whereas the others result from amino acid substitutions. These facts strongly suggest that the common topogenic signal that targets glycosomal enzymes to their intracellular location consists of strategically situated positive clusters, or 'hot spots', as they will be referred to henceforth. Close examination of the g-TIM sequence shows that the excess of positive charges is completely covered by the regions A and B in Figure 3 and that there are no other clusters of positive charges. Therefore, if this hypothesis is correct, the topogenic signal consists of, at most, two hot spots, which we expect to be present in a specific configuration on the surface of the glycosomal enzymes. Therefore one would expect the hot spots to be separated by the same distance on the surface of all the enzymes concerned. To test this point, knowledge of the three-dimensional structure of the enzymes is needed. Such information may also help to identify the putative hot spots where more than two positive clusters exist.

Three-dimensional structures

So far, crystallographic data are available only for g-TIM, at 2.9 Å resolution (Wierenga *et al.*, 1984). With these data, the g-TIM structure could be solved by the molecular replacement method, using the refined coordinates of chicken TIM (Banner *et al.*, 1975; Alber *et al.*, 1981). This shows that g-TIM and chicken TIM have similar structures, as can be expected from the homology between the two sequences (52% identical). In the calculated electron density map the polypeptide chain could be traced, but further refinement, at a higher resolution, is required to obtain the accurate g-TIM structure. Clearly no major structural differences exist between g-TIM and chicken TIM. Therefore the two hot spots have been modelled into the chicken TIM structure, as shown in Figure 4A. They are separated by a distance of ~40 Å. This, therefore, is what we should look for in the structure of the other glycosomal enzymes.

For trypanosomal PGK, GAPDH and aldolase no three-dimensional structures are available yet. However, the percentage identical residues between g-PGK and horse PGK (Merrett, 1981) is 46% and between g-GAPDH and *Bacillus stearothermophilus* GAPDH (Walker *et al.*, 1980) is 53%. This implies that the known three-dimensional structures of horse PGK (Rice, 1981) and bacterial GAPDH (Leslie and Wonacott, 1984) are, for the purpose of this paper, sufficiently accurate models of the corresponding trypanosomal enzymes. Accordingly, these structures were used for the three-dimensional modelling of the amino acid sequences of the trypanosomal enzymes, in order to estimate the distances between positive clusters. The results are shown in Figure 4B and C. No such treatment could be applied to aldolase as no detailed three-dimensional structure from any organism has been published yet, although X-ray studies on several fructose bisphosphate aldolase enzymes have been initiated (Brenner-Holzach and Smit, 1982; Sygush *et al.*, 1985).

Of the four positive clusters that distinguish g-PGK from c-PGK the most likely candidates for hot spots are those labelled B and D in Figure 3B, which are close to 40 Å apart. Regions A, B and C are all within 29 Å from each other, whereas the A−D and C−D distances are both 32−33 Å.

A detailed comparison of the sequences of PGKs (cf. Osinga *et al.*, 1985) provides further support for this tentative assignment. In region C the charge difference between g- and c-PGK arises from a single amino acid substitution (Lys-145 in g-PGK→Glu-146 in c-PGK; Figure 1). Region A in g-PGK carries only one extra positive charge compared with virtually every other non-trypanosomal member of the PGK family (Osinga *et al.*, 1985). Therefore the charge differences at regions A and C appear to be less significant than those at region B which is an insertion unique for trypanosomal PGK. We conclude therefore that the two hot spots in g-PGK are most likely formed by region B of Figure 3B, i.e. residues 74−78, and by region D, i.e. residues 255−259. The three-dimensional arrangement of these two hot spots is shown in Figure 4B.

No two out of three positive clusters found in g-GAPDH (Figure 3C) are exactly 40 Å apart in the assumed three-dimensional structure of this enzyme (Figure 4C). The closest combination is between regions A and C of Figure 3C, i.e. between residues 65−69 and residues 277−281, which are 48 Å apart. Due to the eight-residue insertion containing part of the first cluster, this distance could easily be reduced to 40 Å. This is not the case with the distances in each of the other possible combinations. Consequently, only one combination of hot spots would be consistent with the essential features of a glycosomal import signal as deduced from the g-TIM data and these hot spots are shown in Figure 4C. The localization of one of the hot spots at the eight amino acid insertion starting at position 66 is moreover strongly supported by the fact that the cytosolic GAPDH of *T. brucei* lacks this insertion (Misset *et al.*, 1987).

Without three-dimensional information about aldolase it is impossible to identify with any degree of certainty possible import signals for g-ALDO. However, the two positive clusters as defined in Figure 3D are good candidates. Peptide 142−146 clearly forms a cluster of positive charges, while peptides 15−19, like the other first hot spots is situated in an insert, the only one found in the sequence of the glycosomal enzyme (see Figure 1). Both, as shown below, obey the consensus structure of the hot spots in the other enzymes.

Common structural features of the hot spots

Figure 5 summarizes the information available on the structure of the putative hot spots identified in the four enzymes. Hot spot I, the one nearest to the N terminus, shows remarkable uniformity. In all four enzymes it shows the consensus sequence 'positive-neutral-positive' (p-x-p) and is present in an insert. In contrast, hot spot II has the sequence p-x-p in two enzymes, and p-x-x-p in two others. This difference may be more apparent than real. In the two enzymes (g-TIM and g-GAPDH) that have a hot spot II of p-x-x-p sequence, this stretch is part of an α-helical region, at least in the corresponding cytosolic homologues, whereas the p-x-p hot spot II of PGK appears extended (nothing is known about ALDO).

As illustrated in Figure 6, the α-helical p-x-x-p of GAPDH superimposes over the p-x-p of PGK in such a way that the positive residues of one hot spot coincide almost exactly with those of the other, the distance between the two positive charges of each hot spot being ~7 Å. These facts strongly suggest that an essential and common feature of hot spot II is a pair of positive charges separated by ~7 Å.

Although further refinement of the g-TIM structure and knowledge of the structures of g-PGK and g-GAPDH are required to make definite statements regarding the three-dimensional structure of hot spot I, the p-x-p sequence of hot spot II in g-PGK is the same as the consensus sequence of hot spot I suggesting that the conformations might also be similar.

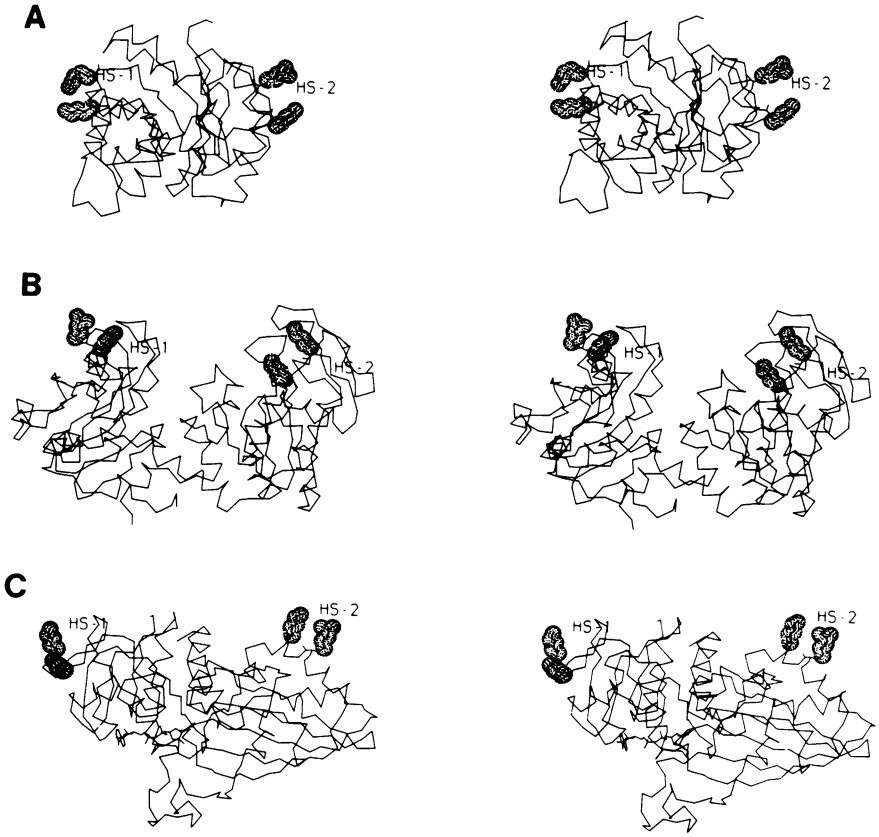

Fig. 4. Putative glycosomal import signals on TIM, PGK and GAPDH. The stereofigures are drawn on the same scale. The position of every hot spot, of which the residues are emphasized by dots, is visualized by two positive side chains. The separation between hot spot I and hot spot II is calculated as the average value of the distances of the Cγ-atoms of the two positively charged side chains in hot spot I to the two corresponding Cγ-atoms in hot spot II. (**A**) TIM. The exact positions of the positively charged side chains are not clearly visible in the current electron density map of g-TIM. Therefore these side chains have been modelled into the structure of the homologous chicken TIM structure. Hot spot I starts at residue 152 of chicken TIM and continues as an insertion between residue 155 and 156. Its position is indicated by two lysines at positions 153 and 155. The second hot spot is homologous to the chicken residues 214–219, it is visualized by lysines at 215 and 218. These two hot spots are at a distance of ~40 Å. (**B**) PGK. Hot spot I is completely in an insertion at position 71 of the homologous horse PGK. Its position is indicated by positive side chains at position 71 and 72. Hot spot II is visualized by lysines at position 254 and 256 of horse PGK. The distance between these two hot spots is ~40 Å. (**C**) GAPDH. Hot spot I starts at residue 60 of bacterial GAPDH and continues as an insertion between residues 60 and 61. Its position is indicated by lysines at position 59 and 60. Hot spot II corresponds with residues 261–266 of bacterial GAPDH. Lysines at 262 and 265 illustrate its position. The distance between hot spot I and hot spot II is ~48 Å.

gPGK	HS-I :	74 G(K)I(R)S	(at insertion)	Q = 4
	HS-II:	255 G(K)S(K)C		
gTIM	HS-I :	151 (K)(K)L(K)(K)	(at insertion)	Q = 6
	HS-II:	215 G(K)N A(R)		
gGAPDH	HS-I :	65 T(K)S(K)P	(at insertion)	Q = 5
	HS-II:	277 (K)(R)A S(K)		
gALDO	HS-I :	15 N(R)L(K)T	(at insertion)	Q = 6
	HS-II:	142 (K)(R)A(K)(K)		

Fig. 5. The hot spot sequences (HS-I and HS-II) of four glycosomal enzymes. The residue numbers refer to the corresponding sequences in Figure 1. The hot spot which is closest to the N terminus, hot spot I, is always associated with an insertion. Residues with positively charged side chains are in circles. There are 16 lysines and five arginines in the eight tabulated hot spot sequences, indicating a preference for lysines. Q = total charge of hot spot I + hot spot II.

Two charges ~7 Å apart could therefore be a characteristic feature of both hot spots.

Discussion

The approach we have followed is based on two unproven but plausible premises. One is that all glycosomal enzymes share a common topogenic configuration that is both necessary and sufficient to single them out for post-translational transfer into the glycosomes. On the strength of this assumption, we have denied a possible topogenic role to two major differences that distinguish the glycosomal PGK from its cytosolic homologue and close evolutionary sibling in the same organism: possession of a supplementary, fairly hydrophobic, 20-residue-long C-terminal extension, and the addition of a substantial number of net positive charges in other areas than regions B and D in Figure 3B (Osinga *et al.*, 1985). The most remarkable of these is region A in Figure 3B with the addition of two positive charges and removal of two negative charges over a four-amino acid stretch (region A in Figure 3B) (Osinga *et al.*, 1985). This may be an over-simplification (as discussed by Borst, 1986) and it remains possible that these important features do play a significant role in the targetting of PGK, even though they may not exist in other glycosomal enzymes. On the other hand they could play a role in the internal organization of the particles. A hydrophobic tail

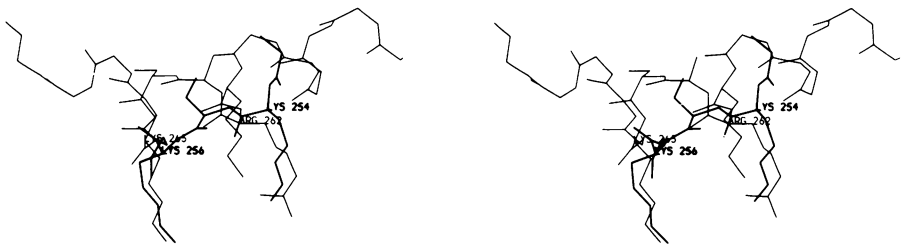

Fig. 6. Superposition of the three-dimensional structures of hot spot II as observed in the homologous structures of horse PGK (Merritt, 1981; Rice, 1981; Osinga *et al.*, 1985) and bacterial GAPDH (Walker *et al.*, 1980; Leslie and Wonacott, 1984; Michels *et al.*, 1986). The positive side chains were built in an extended conformation, while keeping intact the main chain geometry of the homologous structures. The g-PGK hot spot II sequence (thick lines) (GKSKC) is homologous to the stretch 253−257 in horse PGK, which is in an extended conformation. The two positive charges of Lys 254 and Lys 256 are at a distance of 7 Å. The g-GAPDH hot spot II sequence (thin lines) (KRASK) is homologous to the stretch 261−265 in *B. stearothermophilus* GAPDH, which is in an α-helical conformation. The two positive charges of Arg-262 and Lys-265 are at a distance of 7 Å. Four atoms of HS-II (PGK), Cβ-254, Cγ-254, Cβ-256, Cγ-256 were superimposed on respectively Cβ-262, Cγ-262, Cβ-265 and Cγ-265 of HS-II (GAPDH). The root mean square difference in position for these four atoms is 0.8 Å.

may serve to anchor the enzyme in the particle membrane, other elements may be involved in complex formation. Glycosomal enzymes of *T. brucei* differ from other glycolytic enzymes in that they have a high tendency to form multi-enzyme complexes. Such complexes are stable and can easily be isolated in the absence of a surrounding membrane (Opperdoes and Nwagwu, 1980; Oduro *et al.*, 1980; Aman *et al.*, 1985). Therefore, at least some of the elements unique to the glycosomal enzymes might be involved in protein−protein interactions of the constituent enzymes of such complexes.

The second assumption in our approach is that the many deletions, insertions and amino acid substitutions that distinguish glycosomal enzymes from their cytosolic counterparts in other organisms do not significantly affect the three-dimensional structures of these enzymes. Because the sequence homology is as high as 50%, this, however, is a realistic assumption, as can be deduced from the structural and sequential knowledge of several well-studied families of homologous proteins (Chothia and Lesk, 1986; Mori *et al.*, 1986). Moreover, in the case of TIM this assumption has already been proven to be valid, since the close structural homology allowed us to resolve the g-TIM structure by molecular replacement. Also there is a high degree of conservation of residues in functionally important regions of the *T. brucei* enzymes (Michels *et al.*, 1986; Misset and Opperdoes, 1987), while the kinetic properties of several trypanosome enzymes are remarkably similar to those of other glycolytic enzymes (Misset and Opperdoes, 1984, 1987).

Despite the above assumptions, it remains of interest that a surface configuration consisting of two hot spots, each made of two positive residues 7 Å apart, separated by a distance of ~40 Å is a conserved feature of the four glycosomal enzymes. The absence of these hot spot peptides in the known sequence of trypanosomal cytosolic enzymes is consistent with their importance for import into the glycosome. It may be pointed out here that these glycosomes are organelles which occur solely in the Trypanosomatidae and most likely have import signals which are unique as well. Hence, the presence of the proposed pair of positive charges on the surface of proteins in other organisms has no effect on the destination of these proteins and may well occur quite frequently. However, for cytosolic enzymes of Trypanosomatidae, the presence of the import signal would have immediate consequences. As no accurate structures of these cytosolic enzymes are available as yet, a test for the occurrence of the putative glycosome import signal on these enzymes cannot be carried out. Moreover, due to the presence of insertions in, or near, hot spot I, refined structures of glycosomal enzymes are necessary to be able to formulate the precise details of the proposed import signal. For g-TIM this refinement is under way, while only very recently have promising crystals of g-GAPDH been obtained.

Obviously the proposed signal should be present on all glycosomal proteins. It is suggestive in this respect that all other glycosomal enzymes studied, with the exception of glucosephosphate isomerase, are highly basic proteins with a pI between 8.8 and 10.2, which is between 1 and 6 units higher than in any other organism studied (Misset *et al.*, 1986), and also higher than that of the corresponding cytosolic enzymes in *T. brucei* (Misset and Opperdoes, 1987; Misset *et al.*, 1987). The fact that there is an exception to this rule may indicate that overall charge is not decisive in itself. On the other hand, glucosephosphate isomerase could very well, in spite of its relatively low pI (7.5), exhibit the surface configuration of positive charges that we believe to be critical for intraglycosomal transfer. We can also not exclude that the primary translation product of glucosephosphate isomerase mRNA, taken up by the glycosome, has a higher pI than the mature enzyme.

Whether insertions, especially the positively charged insertion bearing hot spot I, will turn out to be another significant feature of glycosomal enzymes remains to be seen. It is of interest, however, that, like g-PGK, g-GAPDH and g-ALDO, the glycosomal glucosephosphate isomerase and glycerol-3-phosphate dehydrogenase are also several kilodaltons larger in size than their counterparts from other organisms (Misset *et al.*, 1986). In the case of two other glycosomal enzymes, hexokinase and phosphofructokinase, no such comparison can be made since these two enzymes seem to be unrelated to their counterparts from other organisms (Nwagwu and Opperdoes, 1980; Misset and Opperdoes, 1984; Misset *et al.*, 1986).

Irrespective of their topogenic functions, common features specific of glycosomal enzymes may be of major interest for the design of drugs, not only against the agents of sleeping sickness, but against other pathogenic Trypanosomatidae, such as *Trypanosoma cruzi* and *Leishmania*, which all possess characteristic glycosomes (Taylor *et al.*, 1980; Hart and Opperdoes, 1984). It is intriguing, and possibly relevant, in this respect, that the potent trypanocide Suramin is a symmetrical molecule containing two clusters of negative charges separated by a distance of ~40 Å.

Materials and methods

Organism

All analyses were carried out on material derived from a cloned strain of the bloodstream form of *T. brucei* 427.

Amino acid sequences

The amino acid sequences of the *T. brucei* enzymes used in Figure 1 have all been published (Osinga *et al.*, 1985; Clayton *et al.*, 1985; Michels *et al.*, 1986; Swinkels *et al.*, 1986) and have been compared with the homologous human sequences (for references see legend to Figure 1) and that of all other organisms for which sequences have been published (for references see the above cited papers and the references cited therein).

Charge-difference profiles

For the construction of charge-difference profiles the sequences were aligned with the corresponding homologous sequences. A deletion in either of the sequences was treated as having a neutral charge. A probe with the length of five residue positions was moved along the sequence and the total charge at each position calculated for both sequences. The difference in total charge divided by five was plotted at each position.

Three-dimensional structure of T. brucei triosephosphate isomerase

Refined chicken TIM coordinates (Banner *et al.*, 1975; Alber *et al.*, 1981) were obtained from Drs D.C.Phillips and P.Artymiuk. A data set to 2.9 Å of the *T. brucei* TIM (Wierenga *et al.*, 1984), together with the molecular replacement method, allowed the structure to be resolved to such a degree that the *T. brucei* polypeptide chain could be built into the electron-density map.

Three-dimensional modelling

Amino acid residues of the putative import signals for *T. brucei* GAPDH, PGK and TIM have been modelled with the interactive computergraphics program 'GUIDE' (Brandenburg *et al.*, 1981) into the homologous structures of *B. stearothermophilus* GAPDH (Leslie and Wonacott, 1984), horse PGK (Rice, 1981) and chicken TIM (Banner *et al.*, 1975; Alber *et al.*, 1981), respectively.

Acknowledgements

We are grateful to Drs D.C.Phillips and P.Artymiuk for providing us with the coordinates of chicken TIM, to Dr C.C.F.Blake for the coordinates of refined horse PGK and to Dr A.Wonacott for the refined coordinates of bacterial GAPDH. We thank Professor C.de Duve for his interest and for critical reading of the manuscript. It is a pleasure to thank Ms R.A.Hogenkamp and Ms F.Van de Calseyde-Mylle for excellent secretarial assistance. This investigation was supported by the United Nations Development Programme/World Bank/World Health Organization Special Programme for Research and Training in Tropical Diseases and the Science and Technology for development Programme of the Commission of the European Communities (to F.R.O.). The Netherlands Foundation for Chemical Research (SON) with financial aid from The Netherlands Organization for the Advancement of Pure Research (ZWO) (to P.B. and W.G.J.H.) and a fellowship of the Wellcome Trust (to W.C.G.).

References

Alber,T., Banner,D.W., Bloomer,A.C., Petsko,G.A., Phillips,D.C., Rivers,P.S. and Wilson,I.A. (1981) *Phil. Trans. R. Soc.*, **B293**, 159–171.

Aman,R.A., Kenyon,G.L. and Wang,C.C. (1985) *J. Biol. Chem.*, **260**, 6966–6973.

Banner,D.W., Bloomer,A.C., Petsko,G.A., Phillips,D.C., Pogson,C.I., Wilson,I.A., Corran,P.H., Furht,A.J., Milman,J.D., Offord,R.E., Priddle,J.D. and Waley,S.G. (1975) *Nature*, **255**, 609–614.

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

Brandenburg,N.P., Dempsey,S., Dijkstra,B.W., Lijk,L.J. and Hol,W.G.J. (1981) *J. Appl. Cryst.*, **14**, 274.

Brenner-Holzach,O. and Smit,J.D.G. (1982) *J. Biol. Chem.*, **257**, 11747–11749.

Chothia,C. and Lesk,A.M. (1986) *EMBO J.*, **5**, 823–826.

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

Fairlamb,A.H., Opperdoes,F.R. and Borst,P. (1977) *Nature*, **265**, 270–271.

Hart,D.T. and Opperdoes,F.R. (1984) *Mol. Biochem. Parasitol.*, **13**, 159–172.

Leslie,A.G.W. and Wonacott,A.J. (1984) *J. Mol. Biol.*, **178**, 743–772.

Maquat,L.E., Chilcote,R. and Ryan,P.M. (1985) *J. Biol. Chem.*, **260**, 3748–3753.

Merrett,M. (1981) *J. Biol. Chem.*, **256**, 10293–10305.

Michels,P.A.M., Poliszczak,A., Osinga,K.A., Misset,O., Van Beeumen,J., Wierenga,R.K., Borst,P. and Opperdoes,F.R. (1986) *EMBO J.*, **4**, 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. and Opperdoes,F.R. (1987) *Eur. J. Biochem.*, in press.

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.*, in press.

Molyneux,D.H. and Ashford,R.W. (1983) *The Biology of Trypanosomes and Leishmania, Parasites of Man and Domestic Animals*. Taylor and Francis, London.

Mori,N., Singer-Sam,J. and Riggs,A.D. (1986) *FEBS Lett.*, **204**, 314–317.

Nwagwu,M. and Opperdoes,F.R. (1980) *Acta Tropica*, **39**, 61–72.

Opperdoes,F.R. (1985) *Br. Med. Bull.*, **41**, 130–136.

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

Opperdoes,F.R. and Nwagwu,M. (1980) In Van den Bossche,H. (ed.), *The Host–Invader Interplay*. Elsevier/North-Holland Biomedical Press, Amsterdam, pp. 683–686.

Opperdoes,F.R., Michels,P.A.M., Misset,O., Hart,D.T. and Van Beeumen,J. (1986) In Agabian,N., Goodman,H. and Noguiera,N. (eds), *Molecular Strategies of Parasitic Invasion. UCLA Symposia on Molecular and Cellular Biology*. Alan R.Liss, New York, Vol. 42, in press.

Oduro,K.K., Bowman,I.B.R. and Flynn,I.W. (1980) *Exp. Parasitol.*, **50**, 240–250.

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.

Rice,D.W. (1981) *Acta Cryst.*, **A37**, 491.

Rottman,W.H., Tolan,D.R. and Penhoet,E.E. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2738–2742.

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

Sygush,J., Boulet,H. and Beadry,D. (1985) *J. Biol. Chem.*, **260**, 15286–15290.

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

Tso,J.Y., Sun,X., Kao,T.-H., Reece,K. and Wu,R. (1985) *Nucleic Acids Res.*, **13**, 2485–2502.

Walker,J.E., Carne,A.F., Runswick,M.J., Bridgen,J. and Harris,J.I. (1980) *Eur. J. Biochem.*, **108**, 549–565.

Wierenga,R.K., Hol,W.G.J., Misset,O. and Opperdoes,F.R. (1984) *J. Mol. Biol.*, **178**, 487–490.

Received on 19 September 1986; revised on 21 October 1986