Review

Author for correspondence:
Wilfredo Quiñones
e-mail: wilqui@ula.ve

Phosphoglycerate kinase: structural aspects and functions, with special emphasis on the enzyme from Kinetoplastea

Maura Rojas-Pirela[1], Diego Andrade-Alviárez[2], Verónica Rojas[1], Ulrike Kemmerling[3], Ana J. Cáceres[2], Paul A. Michels[4,5], Juan Luis Concepción[2] and Wilfredo Quiñones[2]

[1]Instituto de Biología, Facultad de Ciencias, Pontificia Universidad Católica de Valparaiso, Valparaiso 2373223, Chile
[2]Laboratorio de Enzimología de Parásitos, Departamento de Biología, Facultad de Ciencias, Universidad de Los Andes, Mérida 5101, Venezuela
[3]Instituto de Ciencias Biomédicas, Universidad de Chile, Facultad de Medicina, Santiago de Chile 8380453, Santigo de Chile
[4]Centre for Immunity, Infection and Evolution, and [5]Centre for Translational and Chemical Biology, School of Biological Sciences, The University of Edinburgh, The King's Buildings, Edinburgh EH9 3FL, UK

DA-A, 0000-0002-8353-9946; PAM, 0000-0003-3726-6104; WQ, 0000-0002-6574-3732

Phosphoglycerate kinase (PGK) is a glycolytic enzyme that is well conserved among the three domains of life. PGK is usually a monomeric enzyme of about 45 kDa that catalyses one of the two ATP-producing reactions in the glycolytic pathway, through the conversion of 1,3-bisphosphoglycerate (1,3BPGA) to 3-phosphoglycerate (3PGA). It also participates in gluconeogenesis, catalysing the opposite reaction to produce 1,3BPGA and ADP. Like most other glycolytic enzymes, PGK has also been catalogued as a moonlighting protein, due to its involvement in different functions not associated with energy metabolism, which include pathogenesis, interaction with nucleic acids, tumorigenesis progression, cell death and viral replication. In this review, we have highlighted the overall aspects of this enzyme, such as its structure, reaction kinetics, activity regulation and possible moonlighting functions in different protistan organisms, especially both free-living and parasitic Kinetoplastea. Our analysis of the genomes of different kinetoplastids revealed the presence of open-reading frames (ORFs) for multiple PGK isoforms in several species. Some of these ORFs code for unusually large PGKs. The products appear to contain additional structural domains fused to the PGK domain. A striking aspect is that some of these PGK isoforms are predicted to be catalytically inactive enzymes or 'dead' enzymes. The roles of PGKs in kinetoplastid parasites are analysed, and the apparent significance of the PGK gene duplication that gave rise to the different isoforms and their expression in *Trypanosoma cruzi* is discussed.

1. Introduction

Metabolism is a fundamental process in living organisms, consisting of a network of biochemical reactions catalysed and regulated by enzymes. The activities of these enzymes support a wide variety of processes such as cell growth and proliferation, synthesis of cellular components and generation of forms of energy to sustain them. Some enzymes such as kinases are critical in metabolism and other cellular processes necessary for homeostasis and cell
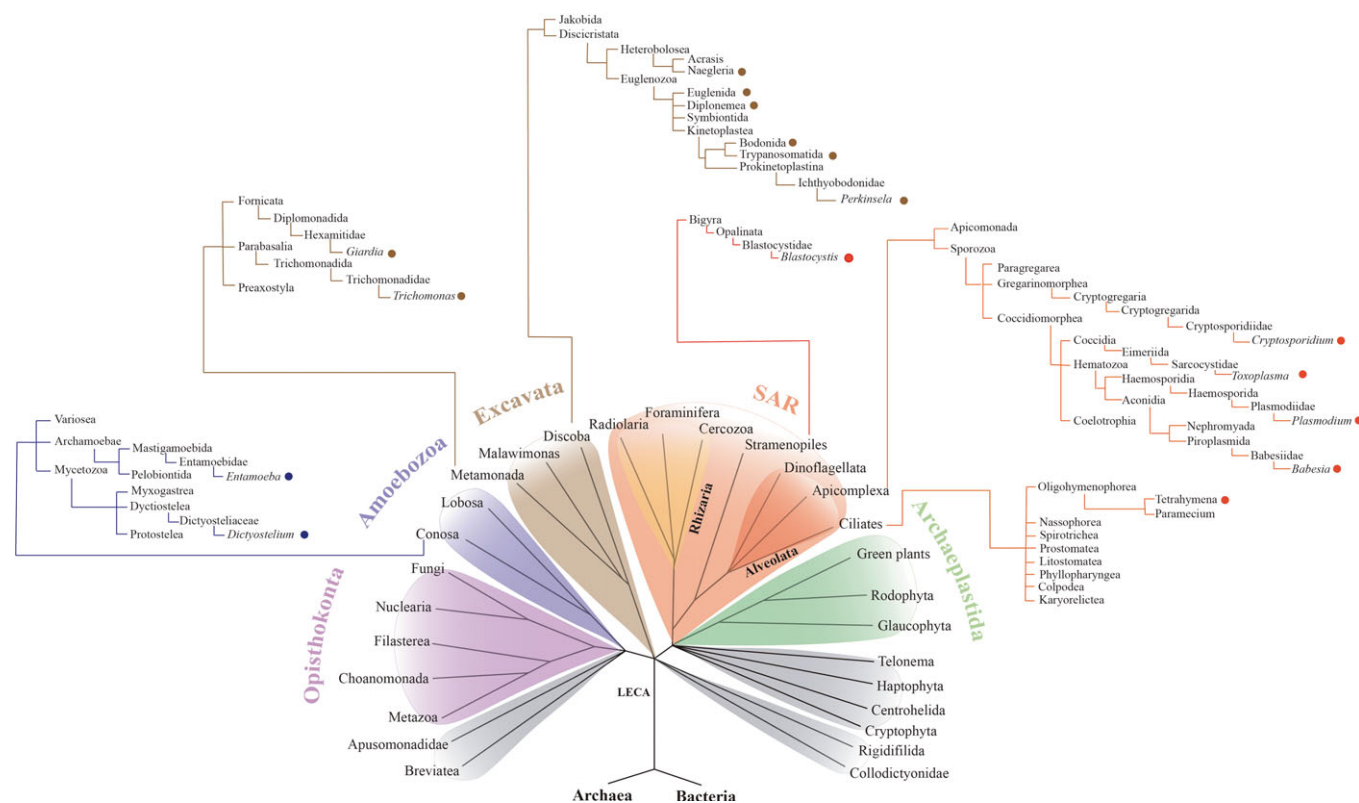
Figure 1. Phylogenetic relationships of the eukaryotes. LECA is the 'last eukaryotic common ancestor' from which different eukaryotic supergroups evolved [6]. The main protistan organisms discussed in this paper are indicated by coloured dots.

survival, although when homeostasis breaks down they can also be involved in the development of cancer and other diseases [1]. In many eukaryotes, kinases form one of the largest gene families, representing approximately 2% of all genes [2]. In parasites of clinical interest such as *Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania* spp., which in humans cause African trypanosomiasis, Chagas disease and leishmaniasis, respectively, protein kinases (PKs) represent approximately 2% of the proteins encoded in each genome, as they do in other eukaryotic organism [3]. The significant presence of PKs encoded in these parasites suggests a key role of these enzymes in their biology [4]. These parasites belong to the Kinetoplastea, a group of flagellated protists comprising members that parasitize many plant and animal species, and in humans are responsible for diseases with serious public health threats and socioeconomic effects [5].

Protists embrace many diverse unicellular eukaryotes, which in modern taxonomy are understood to be paraphyletic (i.e. not forming a single clade) and exclude unicellular fungi [6]. Thus, protists are found in all so-called supergroups, the recognized major eukaryotic subdivisions: Amorphea (Amoebozoa, Opisthokonta), Archaeplastidia, SAR (which includes Stramenopila, Alveolata and the relatively little studied Rhizaria) and Excavata [6]. Figure 1 shows the phylogenetic position of the Kinetoplastea (within the Excavata) with respect to mammals and fungi (within the Opisthokonta) and to various other parasitic and free-living protistan genera which figure in this review and belong to various supergroups: *Dictyostelium*, *Entamoeba*, *Trichomonas*, *Giardia*, *Blastocystis*, *Plasmodium*, *Toxoplasma*, *Naegleria*, *Euglena* and *Diplonema*.

Some of the many PKs identified in kinetoplastids have been shown to be involved in different cellular processes, including transducing signals from the surface of the cell to the nucleus [7–11]. Kinases are classified according to their amino acid sequence similarity and are grouped into subfamilies which share general functional properties [12]. In recent studies, some metabolic enzymes have been included in the protein kinome [13–15]. Attributed to this group of metabolic PKs are also some proteins related to the enzyme phosphoglycerate kinase (PGK) with demonstrated and/or predicted non-metabolic function, thus so-called moonlighting PGKs. Structural and functional aspects of the PGKs from kinetoplastid organisms will be discussed in detail in this review. Figure 2 highlights these organisms which include both parasitic and free-living organisms, the endosymbiotic kinetoplastid *Perkinsela* sp., as well as the free-living *Diplonema papillatum* and *Euglena gracilis,* belonging to separate taxonomic lineages which, together with the Kinetoplastea, are grouped within the Euglenozoa.

2. Phosphoglycerate kinase

PGK (E.C. 2.7.2.3), also known as ATP:3-phospho-D-glycerate 1-phosphotransferase, is considered an essential enzyme for many organisms. This enzyme is mainly associated with a metabolic function, when it catalyses the phospho transfer between two intermediates of carbohydrate metabolism [17]. In glycolysis, this enzyme catalyses the transfer of a phospho group from 1,3-bisphosphoglycerate (1,3BPGA) to ADP to produce 3-phosphoglycerate (3PGA) and ATP in which the high energy released from the carbon substrate is invested. The opposite reaction, in the direction of 1,3BPGA and ADP formation, takes place during gluconeogenesis [18,19]. The forward reaction is one of the two ATP-generating steps in the glycolytic pathway, hence its importance in the metabolism of many organisms, such as protists, both free-living and parasitic ones, the latter including *Plasmodium falciparum* [20],
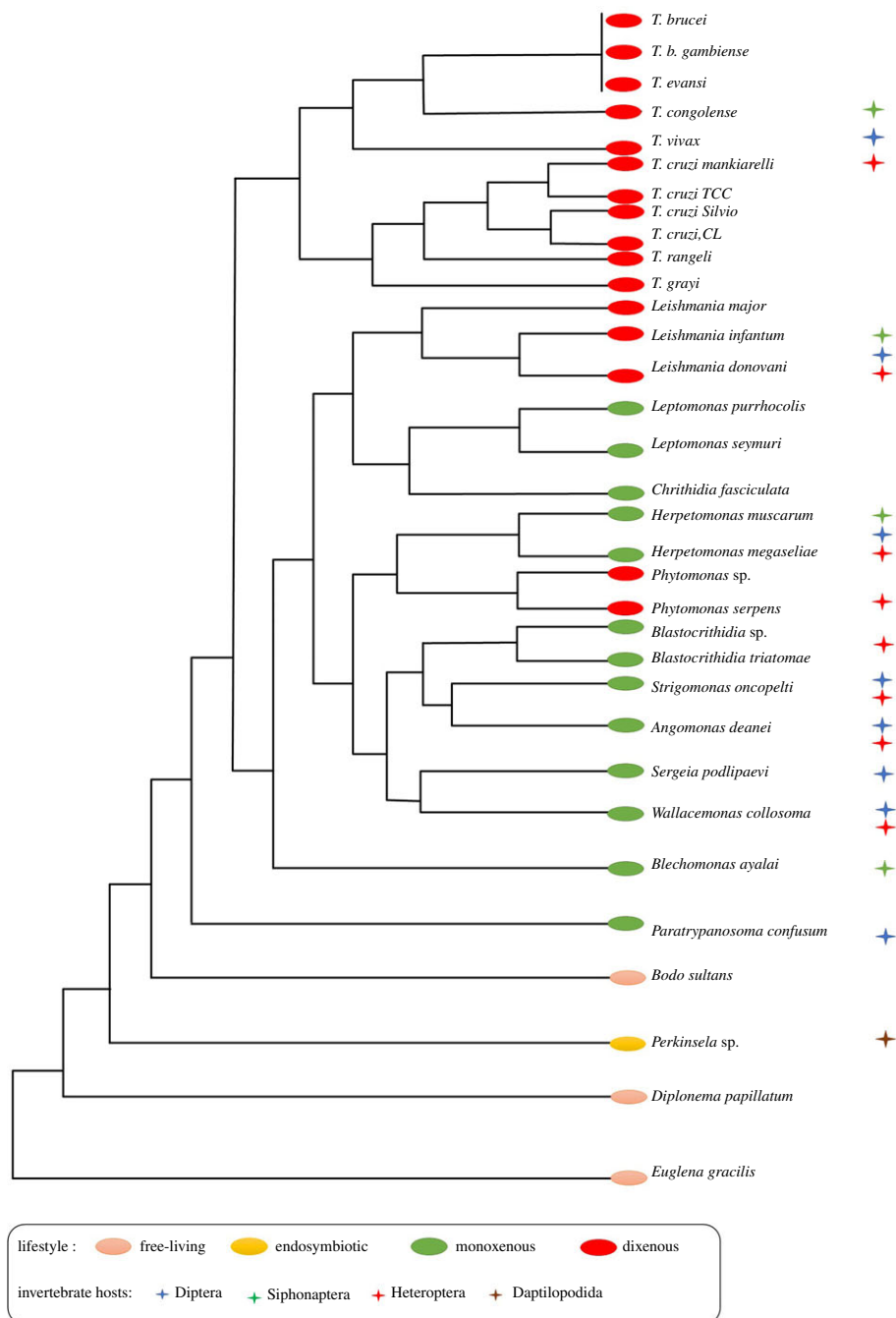
Figure 2. Evolutionary relationships among Kinetoplastea. Outgroups for the construction of the tree (based on small subunit ribosomal RNA sequences) are *D. papillatum* and *E. gracilis*, which, together with the Kinetoplastea, belong to the clade Euglenozoa (figure 1). Figure based on [16].

Entamoeba histolytica [21] and kinetoplastids of clinical interests, such as *T. brucei*, *T. cruzi* and *Leishmania* spp. [4,22–28].

The first attempts to elucidate the functional characteristics of this enzyme occurred in the 1940s, through research performed by Bücher [18,29] who isolated and crystallized a PGK from yeast extracts. After these initial studies, the analysis focused on PGK started to increase markedly, leading to the elucidation of the first low-resolution structures of PGKs from horse muscle [30] and yeast [31]. Later, the discovery of the autosomal *pgk-2* gene in human and mouse testis cells marked a great impulse in the research history of this protein [32]. Currently, the information available about PGK in the NCBI database is considerably extensive and the crystal structural data present in the RSCB Protein Data Bank comprise PGKs from a broad range of organisms, with both structures of the protein in the apo form and with bound substrates, products or molecules that mimic substrates, as well as in open and closed protein conformations [33] (table 1).

PGK is a highly conserved enzyme that has been found in members of each of the three domains of life: Bacteria, Archaea and Eukarya [50–52]. In-depth studies of the structure, function and general properties of PGKs have often been focused on seeking chemotherapeutic agents since this glycolytic enzyme has been considered a target for treatment of some diseases in humans [53]. In addition to its role in glycolysis and gluconeogenesis, and a related metabolic role in the carbon-fixation reactions in autotrophic organisms [54,55], other functions have been attributed to this enzyme. Its non-canonical functions are related to processes such as DNA replication and repair [56], angiogenesis and tumour growth [57], plasminogen binding [58], cell invasion [59], being a constituent of a flagellar axoneme [60] and viral replication [61], among others, which will be highlighted in the following sections of this paper.

Table 1. A sample of PGKs available at the Protein Data Bank. 3PGA, 3-phosphoglyceric acid; AMP, adenosine 5′-monophosphate; ADP, adenosine 5′-diphosphate; ATP, adenosine 5′-triphosphate; BTB, 2-[Bis-(2-hydroxy-ethyl)-amino]-2-hydroxymethyl-propane-1,3-diol; Cl⁻, chloride ion; K⁺, potassium ion; Mg²⁺, magnesium ion; FMT, formic acid; PGE, triethylene glycol; TZ, [4-(4-amino-6,7-dimethoxyquinazolin-2-yl)piperazin-1-yl][(2R)-tetrahydrofuran-2-yl]methanone; SO₄²⁻, sulfate ion; ANP, phosphoaminophosphonic acid-adenylate ester; GOL, glycerol; MPD (4S)-2-methyl-2,4-pentanediol; AMP-PNP, adenylyl-imidodiphosphate; NA, not available.

organism	domain	enzyme	stoichiometry	conformation	ligands	PDB	reference
Pyrococcus horikoshii	archaea	PGK	homodimer		3PGA, GOL, MPD, Cl⁻	2CUN	Mizutano & Kunishima (NA)
Francisella tularensis	bacteria	PGK	monomer	open	ADP	4FEY	Brunzelle *et al.* (NA)
Thermotoga maritima	bacteria	PGK	monomer	closed	3PGA, ANP, Mg	1VPE	[34]
Staphylococcus aureus	bacteria	PGK	monomer	—	—	4DG5	Roychowdhury *et al.* (NA)
Streptococcus pneumoniae	bacteria	PGK	monomer		3PGA, AMP-PNP	3ZLB	[35]
Campylobacter jejuni	bacteria	PGK	monomer	open	K⁺, SO₄, FMT, PGE	3Q3 V	[36]
Thermus caldophilus	bacteria	PGK	monomer	open	—	2IE8	[37]
Bacillus anthracis	bacteria	PGK	monomer	open	Cl⁻, Mg²⁺, BTB	3UWD	[36]
Thermus thermophilus	bacteria	PGK	monomer	—	Na⁺, GOL	1V6S	Mizutani *et al.* (NA)
Escherichia coli	bacteria	PGK	monomer	—	Ca⁺	1ZMR	[38]
Pseudomonas sp.	bacteria	PGK	monomer	open	3PGA	6HXE	[39]
′TAC II 18	bacteria	PGK	monomer	open	—	6I06	[39]
Acinetobacter baumannii	bacteria	PGK	monomer	—	—	5BT8	Fairman *et al.* (NA)
Coxiella burnetii	bacteria	PGK	monomer	—	ADP, Mg²⁺	4NG4	[40]
Geobacillus stearothermophilus	bacteria	PGK	monomer	open	ADP, Mg²⁺	1PHP	[41]
Homo sapiens	eukarya	PGK-1	monomer	open	3PGA, ADP	2XE7	[42]
Saccharomyces cerevisiae	eukarya	PGK	monomer	closed	3PGA, ATP, Mg²⁺	3PGK	[43]
Mus musculus	eukarya	PGK-1	monomer	open/closed	3PGA, ATP, TZ	4O3F	[44]
		PGK-2	monomer	open/closed	3PGA, ATP	2PAA	[45]
Sus scrofa	eukarya	PGK	monomer	open/closed	3PGA, AMP, Mg²⁺	1HDI	[46]
Plasmodium falciparum	eukarya	PGK	monomer	open	SO₄²⁻	3OZ7	[47]
Plasmodium vivax	eukarya	PGK	monomer	—	K⁺, Br⁻	6Y3A	Bilsland *et al.* (NA)
Trypanosoma brucei	eukarya	PGK	monomer	closed	3PGA, ADP, Mg²⁺	13PK	[48]
Equus caballus	eukarya	PGK	monomer	open	—	2PGK	[49]

In this review, several aspects related to the structure and function of PGKs in various protists will be discussed, with emphasis on kinetoplastids. This analysis could serve as a prelude to consider additional functions of this enzyme in the biology of trypanosomatid parasites, since most PGK studies of these parasites so far focused mainly on its glycolytic function.

3. Structural aspects of phosphoglycerate kinase

PGK enzymes have been isolated from a wide variety of organisms. It is a typical hinge-bending monomeric enzyme with a molecular weight of approximately 45 kDa [62]

(table 2). The enzyme is composed of a single folded polypeptide chain that forms two domains of almost equal size, which are separated by a deep cleft, and linked by two α-helices (α-helices 7 and 14) endowing it with its characteristic bilobed structure [73]. Both the N- and C-terminal domain contain a typical Rossmann fold with a core of six parallel strands of β-sheets surrounded by α-helices. The β-turns and irregular structure segments connect both the β-sheets and α-helices [19,36,43]. Both domains are involved in substrate binding, with the N-terminal domain binding 3PGA or 1,3BPGA, whereas the C-terminal domain binds MgADP or MgATP (figure 3a).

Since the 1970s, numerous crystallographic studies of this metabolic enzyme have been performed and the results

Table 2. General physico-chemical and kinetic parameters of some characterized PGKs. ND, not determined; NS, not specified.

organism	domain	enzyme	stoichiometry	enzyme type	MW (kDa)	optimal pH	pI	ATP (mM)	3PGA (mM)	reference
Geobacillus stearothermophilus	bacteria	PGK	monomer	wild-type	42	5.5–8.5	4.9	2.9	2.2	[63]
Corynebacterium glutamicum	bacteria	PGK	homodimer	recombinant	104	7–7.4	ND	0.11	0.26	[64]
Pseudomonas sp. 'TAC II 18'	bacteria	PGK	monomer	wild-type	40.1	8	4.9	0.21	0.53	[65]
				native	40.1	ND	ND	0.24	0.48	[65]
				His-tagged	ND	ND	ND	0.38	0.89	[65]
Synechocystis sp.	bacteria	PGK	monomer	recombinant	ND	ND	ND	0.19	0.18	[55]
Thermus thermophilus	bacteria	PGK	monomer	wild-type	44.6	6–8.5	5	0.28	1.79	[66]
Zymomonas mobilis	bacteria	PGK	monomer	wild-type	44	ND	ND	1.1	1.5	[67]
Entamoeba histolytica[b]	eukarya	PGK	ND	wild-type	ND	ND	ND	5	NS	[21]
			monomer	recombinant	48.4	5–7	ND	1.439	0.547	[21]
Phaeodactylum tricornutum	eukarya	PGKase-1	monomer	recombinant	45	ND	ND	0.89	1.55	[54]
Leishmania major[c]	eukarya	PAS-PGK	monomer	recombinant	62	5.5	ND	0.118	0.661	[28]
Leishmania mexicana mexicana	eukarya	PGKB	monomer	recombinant	47	ND	ND	0.24	1.45	[68]
		PGKC	monomer	recombinant	53	ND	ND	0.13	2.75	[68]
Trypanosoma rangeli	eukarya	PGKB	monomer	wild-type	44	ND	ND	0.13	0.28	[27]
			monomer	recombinant	44	8–10	ND	0.5	0.71	[27]
Trypanosoma brucei	eukarya	PGKB (cytosol)	monomer	wild-type	45	6–9	6.3	0.46	2.04	[22]
			monomer	recombinant	45	ND	ND	0.21	1.25	[69]
		PGKC (glycosome)	monomer	wild-type	47	ND	9.3	0.28	1.55	[69]
			monomer	recombinant	47	ND	ND	0.29	2.02	[69]
Trypanosoma cruzi	eukarya	PGKA	monomer	wild-type	56	ND	6.98[a]	0.217	0.174	[24]
			monomer	recombinant	ND	8	ND	0.236	0.850	[25]
		PGKB	monomer	wild-type	47	5.5	6.6	0.09	0.62	our unpublished results
		PGKC	monomer	wild-type	48	ND	ND	0.10	0.192	[25]
Spirulina geitleri	eukarya	PGK	monomer	wild-type	ND	ND	4.8	0.48	1.5	[70]
Spinacea oleracea (spinach)	eukarya	PGK	monomer	wild-type	46	6.5–9.5	4.3	0.3	1.1	[70]
yeast	eukarya	PGK	monomer	wild-type	47.1	6–9.2	7.2	0.48	1.28	[71]
rabbit muscle	eukarya	PGK	monomer	wild-type	47	6–9.2	7	0.42	1.37	[71]
Homo sapiens	eukarya	PGK-1	monomer	wild-type	ND	ND	ND	0.12	0.22	[72]
		PGK-2	Monomer	wild-type	ND	ND	ND	0.34	0.16	[72]

[a]Predicted.

[b]K_m also calculated for GTP (wild-type: 0.016 mM; recombinant: 0.151 mM).

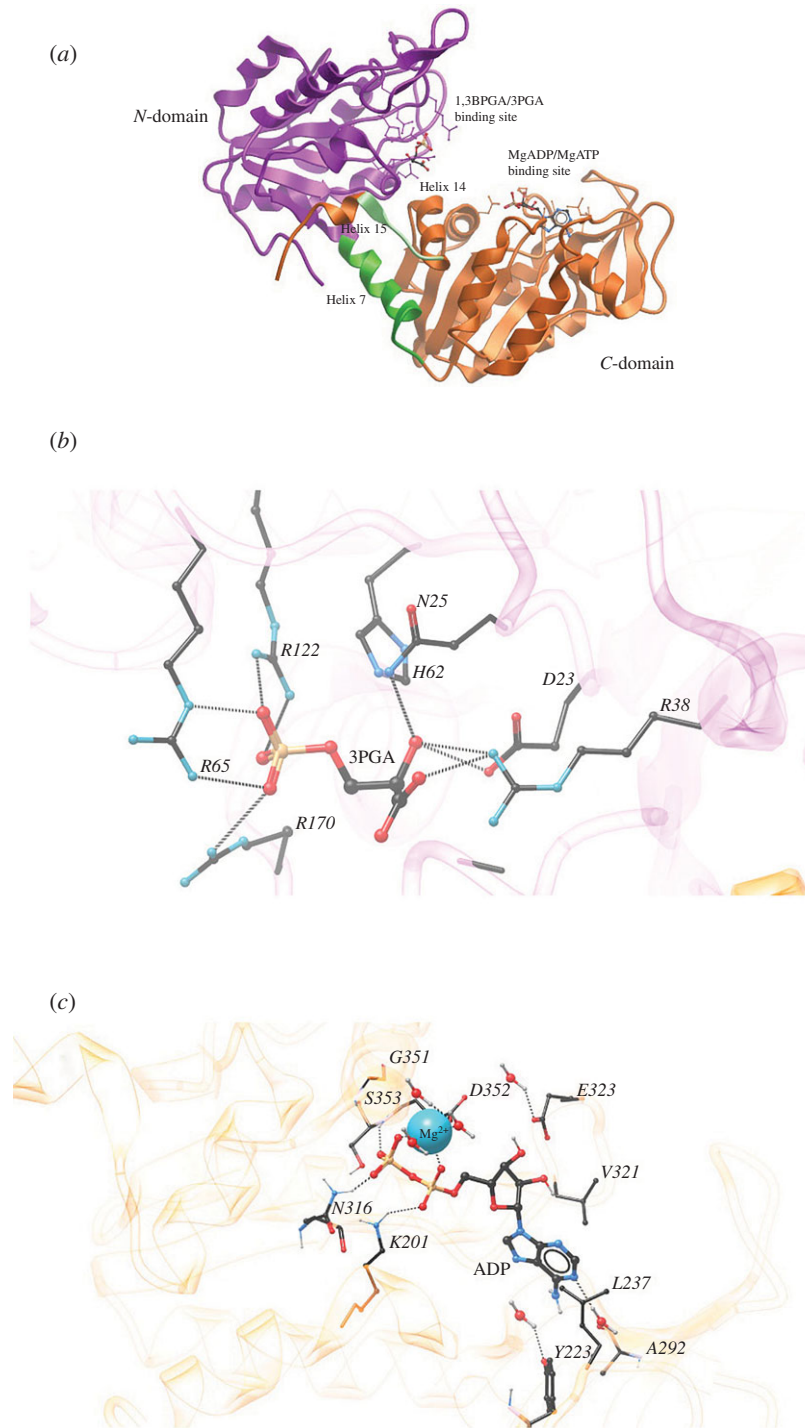[c]Kinetic parameters at pH 7.5 were also calculated (ATP 0.150 mM; 3PGA 0.540 mM).

Figure 3. Three-dimensional structure of phosphoglycerate kinase. (*a*) Ribbon representation of the overall structure of pig muscle PGK (PDB: 1HDI). In colour is highlighted the N-domain (violet); helix 7 or interdomain helix (green); the link between helix 14 and 15; amino acids 404–408 (light green), and the C-domain (orange). The substrate binding sites for 1,3BPGA/3PGA and MgADP/MgATP are indicated in both the N- and C-domain, respectively. The pig muscle ternary complex shown here exhibits an open conformation in comparison with the ternary complex of other PGK structures [46]. (*b*) Substrate 3PGA binding site at the N-terminal domain of pig muscle PGK. (*c*) The MgADP/MgATP binding site of *B. stearothermophilus* (bacterium later renamed to *Geobacillus stearothermophilus*) for the ligand MgADP (PDB: 1PHP). In both (*b,c*), interactions between amino acid residues and substrate through hydrogen bonds (dashed lines) are shown. Atom colour code: black (carbon); white (hydrogen); red (oxygen); blue (nitrogen), orange (phosphorus). Ion colour code: cyan (magnesium, Mg^{2+}).

deposited in the Protein Data Bank (PDB) (table 1). These structural studies of PGKs from different organisms, both prokaryotes and eukaryotes, such as horse [74], pig muscle [75], *Bacillus stearothermophilus* [41], *Thermotoga marítima* [34], *T. brucei* [48], *P. falciparum* [47] and *Bacillus anthracis* [36], have demonstrated that the overall structure of this enzyme is highly conserved. However, these studies also revealed the presence of subtle differences in the PGK structure that must have evolved in response to the environments where this enzyme works. This was most clearly evidenced by crystallographic analysis of PGKs isolated from organisms that live at different temperature ranges. In response to the environmental temperature, PGK has undergone multiple specific adaptations (global and local) [34,39]. Some of these adaptations involved an alteration in the content of hydrophobic and polar amino acids in the N-terminal core. In PGK from isolates of *Pseudomonas psychrophiles*, the N-terminal core is enriched in polar amino acids, a clear difference

compared with its mesophilic and thermophilic counterparts in which predominantly hydrophobic interactions take place. The presence of these polar amino acids could be responsible for protein flexibility, a property that allows the enzyme to proceed with catalysis under low-temperature conditions. Another difference observed at the N-terminal domain of the psychrophilic PGK is an unstructured portion between the sheets β5 and β6. In the same location, the mesophilic PGK has a one-turn helix while the thermophilic and hyperthermophilic PGKs have a two-turn helix [39]. It has been suggested that an increase in flexibility in psychrophilic enzymes allows a better coupling of their substrates and undergoing rapid conformational changes necessary for catalysis, at low energy cost [76]. On the other hand, some determinants of the stability of this enzyme at extremely high temperatures have been identified in hyperthermophilic organisms such as *Thermotoga maritima*. This PGK is one of the most stable enzymes; it is characterized by having a half-life of irreversible thermal inactivation of 2 h at 100°C. The stability of this enzyme is related to a drastic decrease in its flexibility. Its thermostability is attributed to loop stabilization and a depletion phenomenon, which results in a shortening of the loops, which in turn induces a stabilization of the folded state. A reduction in flexibility in protein loop regions by stabilization possibly occurs by the formation of additional hydrogen bonds, as well as shortening or elimination of loops. The stiffness of *T. maritima* PGK is further augmented by an increase in salt bridges, especially between some residues exposed on the surface of the protein [34].

3.1. The substrate binding regions

The location of the substrates on PGK as well as their conformation and nature of interactions with the enzyme have been elucidated by X-ray diffraction studies, through the comparison of electron density maps [19]. The 3PGA or 1,3BPGA binding site at the N-terminal domain is characterized by the presence of a 'basic patch' which comprises arginines and histidines that are highly conserved among PGKs. The study of the PGK from pig muscle identified three arginine residues (at positions 65, 122 and 170) which interact through hydrogen bonds with the oxygen atoms of the substrate's 3-phosphate group (figure 3b). Additionally, a protonated His62 residue interacts through a hydrogen bond with the oxygen in the bridge between the phosphate and carbon skeleton of 3PGA. On the other hand, amino acids Asp23 and Asn25 are responsible for forming hydrogen bonds with the 2-oxygen of 1,3BPGA, acting as the hydrogen receptor and hydrogen donor, respectively (figure 3b) [77].

In addition to being the binding site of 3PGA, the basic patch seems to be a site for regulation of the catalytic activity of PGK. The inhibitory or stimulatory effect of some ions on the activity of some PGKs [47,69,78] has led to the proposal that this basic patch, besides being part of the 3PGA interaction site, can be a place of regulation of the catalytic activity, through binding an inorganic anion. However, only for *P. falciparum* is there PGK structural evidence that demonstrates the binding of ions to this 'basic patch' [47].

As illustrated in figure 3c for *B. stearothermophilus*, the C-terminal domain contains the binding site of ATP or ADP through a 'hydrophobic pocket' present on its surface within the cleft. The structure of this pocket is highly conserved in all PGKs studied so far. The adenine ring is flanked by residues Va1321 and Leu237 [41]. The $-NH_2$ group associated with the

adenine ring forms weak hydrogen bonds with the protein, mainly via the carbonyl oxygen atom of Ala292; however, it also has an additional interaction through a water molecule to the hydroxyl group of Tyr223. The pentose in the nucleotide forms hydrogen bonds with Asp323 through the ribose 2'- and 3'-hydroxyl groups (figure 3c). This molecule adopts an unusual conformation due to the location of the adenine ring and the phosphate groups in the active site. The α- and β-phosphate groups form coordinate bonds with the Mg^{2+} which also establishes coordinate contacts with Asp352 and two water molecules. Additionally, Lys201 forms a hydrogen bond specifically with the β-phosphate, while Asn316 and Ser353 interact through hydrogen bonds with the α-phosphate group. Finally, the β-phosphate lies close to the amino end of the α-helix (helix 12 in the yeast nomenclature) to interact with the characteristic Gly349–Gly350–Gly351 motif. This motif is presumably stabilized by the α-helix dipole and via hydrogen bonding interactions with the main-chain amide H atoms of residues Gly351 and Asp352 [41]. The phosphoryl-group transition state is stabilized by the required divalent Mg^{2+}. This metal ion exhibits a tetragonally distorted octahedral coordination. Axial bonds are through a single water molecule and one of the carboxylate oxygen atoms of Asp352. In equatorial bonds intervene two water molecules and two oxygen atoms of the phosphate groups, one of each of the phosphate groups α and β. Except for the interaction with the Asp, most of the coordinating atoms are at a distance between 2.27 and 2.29 Å from the Mg^{2+}. The oxygen atom from the aspartate carboxylate is only at a distance of 1.96 Å from the Mg^{2+}, which reflects the charged nature of this interaction [41]. It has been suggested that the presence Mg^{2+} is key to PGK activity, because this divalent ion, by forming a complex with the nucleotide's α- and β-phosphates, and thus shielding their negative charges, allows the nucleophilic attack to occur when the ligands are bound to the enzyme. This charge-stabilization phenomenon is a distinctive characteristic of phosphoryl transfer reactions [79].

The amino acid composition of the substrate-binding sites of PGKs appears to be conserved between the different organisms, including those of the kinetoplastids studied so far [4,27,51,80,81].

3.2. The PGK hinge-bending motion and its catalytic mechanism

In the absence of substrate, PGK exhibits an 'open' conformation. The simultaneous binding of both substrates, 3PGA/1,3BPGA and Mg-ADP/Mg-ATP, in the N- and C-terminal domain, respectively, induces an extensive hinge-bending motion, which leads to a 'closed' conformation by the domains approaching each other and so bringing their bound substrates into proximity [82]. In the case of the glycolytic reaction, this close proximity favours the nucleophilic attack by the ADP-β-phosphate oxygen atom at the 1-phosphate of 1,3BPGA. This reaction involves a change in the configuration in the γ-phosphate group, followed by a unique displacement mechanism, in which a direct transfer of the phosphoryl group between the bound substrates occurs through a charged transition state. The additional negative charge, which evolves in the SN2-reaction at the pentagonal transition state phosphate, is stabilized by the enzyme [34]. During its catalytic cycle, PGK seems to spend most of its time in a completely open

conformation with short periods of closure and catalysis; this allows rapid diffusion of substrates and products into and out of the binding sites. It is important to note that PGK has a propensity for the open form, and only the simultaneous binding of both substrates affects this conformation [42,83].

Studies of the human PGK have revealed that the binding of 3PGA induces the alignment and approach of two residues, Arg65 and Arg170 (both present in the N-terminal domain), allowing them to form essential salt bridges with a reoriented Asp218 (present in the C-terminal domain) only in the closed conformation. These salt bridges are the main contributors to the stabilization of the catalytically active closed conformation. In turn, the binding of ADP promotes the displacement of catalytically essential residues in a 'catalytic loop', extending helix-8 and prepares the C-terminal domain to stabilize the closed conformation. The presence of a hydrophobic region or hydrophobic 'patch' encrypted in the enzyme seems to favour the open conformation. This region, comprising several residues in the hinge helix-7 as well as other residues in the protein's core, becomes exposed during domain closure, and is responsible for returning the enzyme to the open conformation by acting as a 'spring-loaded' release mechanism and makes the open conformation the thermodynamically most stable one. On the other hand, ionic interactions act to keep the conformation closed to allow catalysis [42].

The catalysis-associated mechanism of hinge movement, documented for human PGK, is very similar to that reported for PGKs from other organisms, including the PGK isoenzyme of *T. brucei* that is located in peroxisome-related organelles called glycosomes which, in kinetoplastids, contain the major part of the glycolytic pathway [34,36,48]. Although the *T. brucei* PGK is a hinge flexion enzyme, its catalysis activation mechanism is very peculiar, since the effects induced by the binding of the two substrates are synergistically combined to induce important conformational changes in the enzyme [48].

4. PGK functional forms

With regard to its structural organization, PGK is generally monomeric [62], except in some extremophilic microorganisms where it adopts a dimeric or tetrameric conformation [51,52], while in others, it has been observed as a bifunctional enzyme [52,84] (table 3). In Archaea such as *Pyrococcus woesei* and *Methanothermus fervidus*, the PGKs are homodimers in their native state, formed by monomers of approximately 46 kDa [51]. In other archaean organisms such as *Sulfolobus solfataricus*, the PGK is particularly unusual, since it consists of a tetramer of four identical subunits, each of them with an approximate molecular weight of 45 kDa, like the monomeric PGK form found in almost all other organisms [126].

The structure of the enzyme may not only vary in the number of PGK subunits that catalyse the glycolytic reaction, but also in the presence of additional domains with other catalytic activity than PGK, as the enzymes that have been identified in some extremophiles [84]. In the thermophilic bacterium *T. maritima*, the existence of two functional forms of this enzyme has been documented, a monomeric PGK and a PGK–triose-phosphate isomerase (TIM) bifunctional complex. In this bifunctional enzyme, both a PGK domain and a TIM domain are covalently associated into a single structure. This 70 kDa PGK-TIM fusion protein is encoded by the *fus*

gene. Similar to some other archaeal PGKs, this fusion protein forms tetrameric complexes [84]. Although the formation of macromolecular complexes of PGK is not a specific feature of Archaea, it has been suggested that this may be related to thermo-adaptation. Probably, dimerization contributes to greater stability of the protein by favouring hydrophobic interactions through subunit contacts and by reducing the surface area exposed to the solvent [51,127].

When analysing the *Leishmania* genome, something similar to what is observed in these extremophilic organisms was found: an open-reading frame (ORF) that showed indications for a fusion of fragments of PGK- and TIM-related genes (table 3). This will be further discussed in §8 of this review. Furthermore, proteins with additional regulatory domains associated with the PGK structure have been identified in several organisms including kinetoplastids as documented later in this review and various recent publications. The presence of some of these regulatory modules, such as the PAS-domain (Per-ARNT-Sim), calmodulin-binding domain (CaMBD), cyclic nucleotide binding domains (CNB) and transmembrane domains (TMD), possibly endow the PGK with non-canonical functions (table 3) [4,28,128]. An additional discussion of the possible implications of the presence of these regulatory domains in isoenzyme function will be discussed in the following sections.

On the other hand, in *Escherichia coli*, the PGK has been catalogued as an enzyme that is part of a complexome, associated with proteins involved in glycolysis and stress response [129]. The PGK is a component of the GmpI (phosphoglyceromutase)–Pgk–AhpC (alkyl hydroperoxide reductase) (GmpI–Pgk–AhpC) complex. The formation of such complexes could indicate some linkage of intermediate metabolism with the response to stress conditions. Furthermore, in pancreatic β cells exposed to endoplasmic reticulum (ER) stress, the integrated stress response is coupled to metabolic alternations triggered by sulfhydration of key enzymes in intermediary metabolism [130]. Additionally, among archaeal members, including extreme halophiles and some methanogens, PGK appears as a couple with glyceraldehyde-3-phosphate dehydrogenase (GAPDH/PGK) that is functionally active in the glycolytic direction, while the same coupled enzymes operate in the opposite direction (gluconeogenis) in anaerobic and hyperthermophilic Archaea [131]. A GAPDH/PGK interaction has also been reported for human erythrocytes as a pH-dependent phenomenon. This specific protein–protein interaction between GAPDH and PGK may play a role in determining the fate of 1,3BPGA produced in the GAPDH-catalysed reaction [132].

5. PGK activity regulation

5.1. Regulation by ions
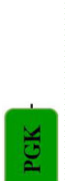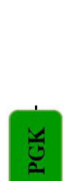
PGK activity is influenced by the presence of mono- and divalent ions [51,63,64,133]. Some ions, including Ni^{2+}, Co^{2+}, Mn^{2+}, Cd^{2+}, Ca^{2+} or Zn^{2+}, could stimulate or inhibit its activity. Mg^{2+} ions are necessary during catalysis by this enzyme, as reported for different PGKs studied so far (see §3.1) [63,64,134]. In *Corynebacterium glutamicum*, Zn^{2+} ions strongly inhibit the PGK activity (Ki = 0.45 mM) [64]. However, such inhibition should not be considered as a generality, since other PGKs are activated in the presence of metal ions other than Mg^{2+} [63,133]. Kinetic studies of *B. stearothermophilus* PGK have

Table 3. PGK enzymes and PGK-related proteins identified by genomic analyses in different protists. Sequences of different PGKs were retrieved from the following databases: Tetrahymena Genome Database (TGD) [85], National Center for Biotechnology Information (NCBI) [86], The UniProt Consortium (UniProt) [87], Database for *Dictyostelium discoideum* (DictyBase) [88], Ensembl Protists, The Amoeba Genomics Resource (AmoebaDB) [89], The Giardia Genomics Resource (GiardiaDB) [90], The Trichomonas Genomics Resource (TrichDB) [90], The Toxoplasma Genomics Resource (ToxoDB) [91], The Plasmodium Genomics Resource (PlasmoDB) [92], The Cryptosporidium Genomics Resource (CryptoDB) [93] and The Kinetoplastid Genomic Resource (TriTrypDB) [94]. Alignments and determination of identity percentages reported in the text were performed using *Clustal Omega* and *Muscle* (EMBL-EBI) [95]; query sequences used were Tb927.1.710, Tb927.1.720, Tb927.1.700, Tb927.11.2380, TcCLB.511419.40 and TcCLB.511419.50. Identification of domains and motifs present in PGK sequences was done using the following bioinformatics servers: *PAS domain*: Protein Blast (NCBI) [96], InterProScan (EMBL-EBI) [97] and Simple Modular Architecture Research Tool (SMART) [98]; *e*-values ranges between 1 e^{-6} and 8.9 e^{-9}. *CNB domain*: NCBI, EMBL-EBI and SMART; *e*-values ranges between 2.6 e^{-6} and 4 e^{-36}. *HTH (Helix–Turn–Helix)* motif: ExPaSy [99], GYM [100] and iDNA-Prot [101]; *e*-values range between −0.08 and 1.45. For the PGK of *T. congolense*, the HTH domain was identified with a score of 29. *Membrane helix*: SMART [98], Predicting transmembrane protein topology with a hidden Markov model (TMHMM prediction) [102] and Phobius [103]; *e*-values ranges between 0.8 and 1. A specific *e*-value for each identified protein is given in electronic supplementary material, table SIII. The subcellular localization prediction of identified PGKs was based on either recognition of a PTS1 consensus sequence for glycosomal localization as described by Acosta *et al.* [104] or using the following servers: DeepLoc [105], WoLF PSORT [106] and Cell-Ploc [107]. Subcellular localization symbology: *, predicted location through bioinformatics tools; ◆, established localization through experimental studies; ◇, established localization through experimental studies in a non-sequenced isolate; ▲, based on consensus sequence proposed in other studies [104]. Localization symbology: Cytosol; Cytosol; Mitoch, Mitochondrion; P. memb, Plasma membrane; Lys, lysosome; nud, Nudeus. Domain symbology: MYCBP-associated protein domain — MYCBP, calmodulin-binding domain —, AAA-ATPases domain (Walker A/B) , PAS domain PAS, PGK domain PGK, PGK domain having lost residues involved in binding substrate (3PGA) PGK, PGK domain having lost residues involved in binding ATP PGK, PGK domain having lost residues involved in binding substrate (3PGA) and residues binding ATP PGK, triosephosphate isomerase domain TIM, HTH domain , cyclic nucleotide-binding domain CNB.

organism	number of pgk genes	ID gene	name	PM (kDa)	associated domains	pI	protein length a.a	localization	Database
free-living protists									
Tetrahymena thermophila	1	TTHERM_00856750	hypothetical protein	149	MYCBP—PGK	5.21	1284	Cytosol /nucl*	TGD Database [85,108]
		TTHERM_00929450	phosphoglycerate kinase (PGK1)	46	PGK	8.85	420	Cytosol *	
Euglena gracilis	2	AAU11484.1	phosphoglycerate kinase cytosolic ($_C$PGK)	45	PGK	5.75	414	Cytosol ◆	NCBI [109]; UniProt [87,110].
		AAU11483	phosphoglycerate kinase chloroplast precursor polyproteic (qpPGK)	105	PGK PGK PGK	4.97/5.79/5.79	139/420/420	Chloroplast ◆	
Dictyostelium discoideum	1	DDB_G0287595	phosphoglycerate kinase A (PGKA)	46	PGK	6.27	420	Cytosol◆	DictyBase [88,111]
Naegleria gruberi NEG-M	1	NAEGRDRAFT_81218	phosphoglycerate kinase	44	PGK	6.14	406	Cytosol *Flagellum ◆	Ensembl Protists [112,113].
parasitic protists									
amitochondriate protists									
Entamoeba histolytica HM-1:IMSS-A	1	EHI7A_022660	phosphoglycerate kinase, putative	45	PGK	6.77	415	Cytosol*	[21,89]
Giardia lamblia 2DH	1	DHA2_90872	phosphoglycerate kinase	44	PGK	7.35	409	Cytosol *	GiardiaDB [90]

(Continued.)

Table 3. (Continued.)

organism	number of pgk genes	ID gene	name	PM (kDa)	associated domains	pI	protein length a.a	localization	Database
Trichomonas vaginalis G3	2	TVAG_268050	phosphoglycerate kinase	45	PGK	7.07	419	Cytosol *, ◆	TrichDB [90,114]
		TVAG_383940	phosphoglycerate kinase	45	PGK	7.33	419	Cytosol *, ◆	
Apicomplexa									
Toxoplasma gondii ME49	2	641.m00193	phosphoglycerate kinase PGKI	45	PGK	6.57	417	Cytosol *, ◆	ToxoDB [91,115]
		41.m01331	phosphoglycerate kinase PGKII (ApiPl)	63	PGK	9.00	593	Apicoplast◆	
Plasmodium falciparum 3D7	1	PF3D7_0922500	phosphoglycerate kinase	45	PGK	7.20	416	Cytosol *P. membrane ◆	PlasmoDB [92,116]
Cryptosporidium parvum IOWA-ATCC	2	CPATCC_0006130	phosphoglycerate kinase	45	PGK	8.50	385	Cytosol/nucl*	CryptoDB [93]
		CPATCC_0006650	PGK-family protein	45	PGK	6.40	419	Cytosol/nucl*	
Babesia microti Strain R1	1	BMR1_03g00625	phosphoglycerate kinase	44	PGK	6.12	403	Cytosol*◆	UniProt [87,117]
Stramenopiles									
Blastocystis hominis Str. *Singapore Isolated B*	1	GSBLH_T00001089001	phosphoglycerate kinase	46	PGK	9.05	420	Mitoch*	Ensembl Protists [112]
Diplonemea and Kinetoplastea									
free-living									
Diplonema papillatum	1	BAQ25443	phosphoglycerate kinase	45	PGK	9.26	421	Glycosome ◆▲	[118]
Bodo saltans Lake Konstanz strain	3	BSAL_14730	phosphoglycerate kinase,putative	56	PGK	6.52	513	Cytosol*	TriTrypDB [94]
		BSAL_29930	phosphoglycerate kinase, putative	57	PAS PGK	9.06	527	Cytosol*	
		BSAL_70390	phosphoglycerate kinase, putative	89	PGK CNB	9.28	829	Glycosome (Mem)*	

(*Continued.*)

Table 3. (Continued.)

organism	number of pgk genes	ID gene	name	PM (kDa)	associated domains	pI	protein length a.a	localization	Database
endosymbiotic									
Perkinsela sp. CCA 156/4	1	XU18_4992	phosphoglycerate kinase, putative	58		9.35	540	Glycosome ▲	Ensembl Protists [112]
obligate parasites									
Paratrypanosoma									
P. confusum CUL13	2	PCON_0085140	phosphoglycerate kinase	28		5.34	272	Cytosol*	TriTrypDB [94]
		PCON_0003620	phosphoglycerate kinase	104		9.31	956	Glycosome*	
salivarian trypanosomes									
T. brucei TREU927	4	Tb927.1.710	phosphoglycerate kinase (PGKB)	45		7.55	420	Cytosol◆	[22,60,119–121]
		Tb927.1.700	phosohoglycerate kinase (PGKC)	47		9.67	440	Glycosome◆	
		Tb927.1.720	phosphoglycerate kinase (PGKA)	56		9.00	509	Glycosome◆	
		Tb927.11.2380	phosphoglycerate kinase, putative	101		8.34	913	Flagellum◆	
T. congolense IL3000	6	TcIL3000_1_220.1	phosphoglycerate kinase	45		7.19	420	Cytosol◆	[122]; TriTrypDB [94]; GeneDB [123]
		TcIL3000_1_230	phosphoglycerate kinase, putative	45		7.19	420	Cytosol◆	
		TcIL3000_1_240	phosphoglycerate kinase, putative	56		7.69	509	Glycosome◆	
		TcIL3000.11.2180.1	phosphoglycerate kinase, putative (fragment)	66		8.01	602	Cytosol*	
		TcIL3000_0_37920	phosphoglycerate kinase	45		6.93	420	Cytosol*	
		TcIL3000_0_37930	phosphoglycerate kinase	45		7.19	420	Cytosol*	

Table 3. (*Continued*.)

organism	number of *pgk* genes	ID gene	name	PM (kDa)	associated domains	pI	protein length a.a	localization	Database
T. evansi STIB805	4	TevSTIB805.1.640	phosphoglycerate kinase	47	PGK	9.69	440	Glycosome ◆	[26]; TriTrypDB [94]
		TevSTIB805.1.650	phosphoglycerate kinase	45	PGK	8.54	420	Cytosol *	
		TevSTIB805.1.660	phosphoglycerate kinase	56	PGK	9.00	509	Cytosol / Mitoch*	
		TevSTIB805.11_01.2430	phosphoglycerate kinase, putative	101	PGK CNB	8.30	913	Glycosome (Mem)*	
T. vivax Y486	4	TvY486_0604100	phosphoglycerate kinase, putative	41	PGK	10.19	371	Glycosome▲	TriTrypDB [94]
		TvY486_0100140	phosphoglycerate kinase, putative	45	PGK	6.52	420	Cytosol *	
		TvY486_1102390	phosphoglycerate kinase, putative	86	PGK CNB	6.84	795	Glycosome*	
		TvY486_0100150	phosphoglycerate kinase, putative (Fragment)	32	PGK	8.36	301	Cytosol / Mitoch*	
stercorarian and reptilian trypanosomes									
T. cruzi CL-Brener Esmeraldo like	6	TcCLB.506127.9	phosphoglycerate kinase, putative	13	PGK	10.47	115	Cytosol *	[4,24,25]; TriTrypDB [94,124]
		TcCLB.506125.30	phosphoglycerate kinase, putative	22		11.25	201	Mitoch*	
		TcCLB.511419.40	phosphoglycerate kinase, putative(PGKB/C)	47	PGK	6.6	419	Cytosol ◇	
		TcCLB.511419.50	3-phosphoglycerate kinase, glycosomal (PGKA)	56	PGK	16.91	503	Glycosome◇	
		TcCLB.506945.20	PAS-domain containing phosphoglycerate kinase, putative	58	PAS PGK	9.26	527	Glycosome◇	
		TcCLB.504153.20	phosphoglycerate kinase, putative	100	PGK CNB	9.03	906	Glycosome (Mem)*	
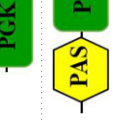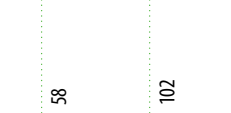
(*Continued*.)

Table 3. (Continued.)

organism	number of pgk genes	ID gene	name	PM (kDa)	associated domains	pI	protein length a.a	localization	Database
T. rangeli SC58	4	TRSC58_02767	phosphoglycerate kinase	45	PGK	6.02	419	Cytosol ◇	[27]; TriTrypDB [94]
		TRSC58_00739	phosphoglycerate kinase	55	PGK	6.74	503	Glycosome ◇	
		TRSC58_04456	PAS-domain containing Phosphoglycerate kinase, putative (PAS-PGK)	58	PAS-PGK	9.52	527	Glycosome ◇	
		TRSC58_00037	phosphoglycerate kinase, putative	95	PGK-CNB	8.43	864	Glycosome ◇	
T. grayi ANR4	4	D004_04781080	phosphoglycerate kinase	48	PGK	6.80	448	Cytosol *	TriTrypDB [94]
		D004_00061260	PAS-domain containing phosphoglycerate kinase, putative (PAS-PGK)	58	PAS-PGK	9.42	527	Glycosome* ▲	
		D004_02181000	phosphoglycerate kinase	80	PGK-CNB	8.22	730	Cytosol *	
		D004_13681000	3-phosphoglycerate kinase glycosomal	29	PGK	8.74	262	Cytosol *	
Leishmania									
Leishmania major Strain Friedlin	4	LmjF.20.0110	phosphoglycerate kinase B, cytosolic	45	PGK	8.77	446	Cytosol ◆	[28,81]; TriTrypDB [94]
		LmjF.20.0100	phosphoglycerate kinase C, glycosomal	51	PGK	9.12	479	Glycosome ◆	
		LmjF.27.1720	phosphoglycerate kinase (pseudogene) putative PAS-domain containing	29	PGK-TIM	9.57	270	Cytosol*	
		LmjF.30.3380	phosphoglycerate kinase, putative(PAS-PGK)	58	PAS-PGK	8.75	527	Glycosome/Lys ◆	

(*Continued.*)

Table 3. (Continued.)

organism	number of pgk genes	ID gene	name	PM (kDa)	associated domains	pI	protein length a.a	localization	Database
Leptomonas									
Leptomonas pyrrhocoris H10	4	LpyrH10_25_1870	phosphoglycerate kinase C, glycosomal (PGKC), putative mitochondrial	51		8.84	408	Glycosome/Mitoch*	TriTrypDB [94]
		LpyrH10_25_1860	3-phosphoglycerate kinase, glycosomal, putative mitochondrial	55		6.83	506	Glycosome/Mitoch*	
		LpyrH10_04_6440	PAS-domain containing phosphoglycerate kinase, putative, putative mitochondria	58		8.78	527	Glycosome * ▲	
		LpyrH10_07_2090	phosphoglycerate kinase, putative	102		8.36	923	Glycosome (Mem) *	
Crithidia									
Crithidia fasciculata Cf-Cl	5	CFAC1_180007000	phosphoglycerate kinase B	45		7.42	417	Cytosol ◇	[80,125]; TriTrypDB [94]
		CFAC1_180006900	phosphoglycerate kinase C	50		9.12	481	Glycosome◇	
		CFAC1_180007100	phosphoglycerate kinase APAS-domain containing	55		7.34	506	Glycosome◇	
		CFAC1_26059100	phosphoglycerate kinase, putative (PAS-PGK)	58		9.60	527	Glycosome▲	
		CFAC1_23031900	phosphoglycerate kinase/cyclic nudeotide-binding domain containing protein, putative	101		8.32	922	Glycosome (Mem)*	

(Continued.)

15

Table 3. (Continued.)

organism	number of *pgk* genes	ID gene	name	PM (kDa)	associated domains	pI	protein length a.a	localization	Database
parasitic and host to endosymbiotic bacterium									
Angomonas									
Angomonas deanei (GCA_000442575.2)	12	AGDE_06248	phosphoglycerate kinase	29		5.39	270	Cytosol*	Ensembl Protists [112]
		AGDE_05604	phosphoglycerate kinase	55		6.98	502	Cytosol*	
		AGDE_10081	phosphoglycerate kinase	55		6.98	502	Cytosol*	
		AGDE_10148	phosphoglycerate kinase	55		6.98	502	Cytosol *	
		AGDE_05595	phosphoglycerate kinase	44		8.55	418	Cytosol*	
		AGDE_08002	phosphoglycerate kinase	42		8.64	398	Cytosol*	
		AGDE_07220	phosphoglycerate kinase	103		8.94	924	Glycosome (Mem)*	
		AGDE_10178	PAS-domain containing phosphoglycerate kinase, putative (PAS-PGK)	57		9.74	527	Glycosome▲	
		AGDE_05077	3-phosphoglycerate kinase, glycosomal	102		8.56	919	Glycosome (Mem)*	
		AGDE_04297	3-phosphoglycerate kinase, glycosomal	103		8.44	924	Glycosome (Mem)*	
		AGDE_02770	3-phosphoglycerate kinase, glycosomal	103		8.44	924	Glycosome (Mem)*	
		AGDE_05968	phosphoglycerate kinase	310		9.20	310	Glycosome (Mem)*	

(Continued.)

Table 3. (Continued.)

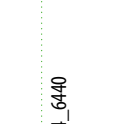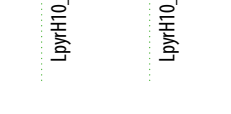| organism	number of *pgk* genes	ID gene	name	PM (kDa)	associated domains	pI	protein length a.a	localization	Database
Strigomonas									Ensembl Protists
Strigomonas culicis (GCA_000442495.1)	10	STCU_09535	phosphoglycerate kinase	28		4.52	270	Cytosol*	[112]
		STCU_09108	phosphoglycerate kinase	36		4.65	417	Cytosol*	
		STCU_07788	phosphoglycerate kinase	48		8.26	455	Glycosome (Mem)*	
		STCU_01151	phosphoglycerate kinase C, glycosomal	48		8.39	455	Glycosome (Mem)*	
		STCU_05393	phosphoglycerate kinase C, glycosomal	48		8.39	455	Glycosome (Mem)*	
		STCU_04883	phosphoglycerate kinase	55		7.60	500	Cytosol	
		STCU_04975	PAS-domain containing phosphoglycerate kinase, putative (PAS-PGK)	57		9.57	527	Glycosome▲	
		STCU_02304	phosphoglycerate kinase	103		8.31	935	Glycosome (Mem)*	
		STCU_08802	phosphoglycerate kinase	103		8.31	935	Glycosome (Mem)*	
		STCU_06143	phosphoglycerate kinase	103		8.31	935	Glycosome (Mem)*	

shown that Mg^{2+} and Mn^{2+} have the same contribution to its activity, while in the presence of other cations such as Co^{2+} and Ca^{2+}, the enzyme showed only 58% and 15% activity, respectively [63]. On the other hand, the PGK present in pea seed tissue is apparently activated in the presence of Mn^{2+} and Co^{2+}, ions that appear to be as effective as Mg^{2+} at equivalent concentrations. On the contrary, Ca^{2+} and Fe^{3+} induce a decrease of 52 and 30% of enzymatic activity, respectively. The activity of this PGK is completely inhibited by ions such as Cu^{2+}, Zn^{2+} and Hg^{2+} [133]. As in other kinases, such as pyruvate kinase (PYK), divalent cations form metal–nucleotide complexes that become the substrate of the catalysed chemical reaction and lead to the activation of the enzyme (see §3.1) [135]. With regard to monovalent ions such as Na^+ and K^+, the influence of these ions on the activation of the enzyme can vary, even for PGKs from organisms belonging to the same kingdom. In the case of *C. glutamicum* PGK, the presence of these ions significantly increases the activity of this enzyme [64], while in *Thermus thermophilus*, PGK is slightly inhibited by Na^+ [66]. On the other hand, in a hyperthermophilic organism such as *M. fervidus*, where PGK is a homodimeric enzyme, the activity increases in the presence of K^+ ions. In this organism, the activity of this enzyme appears to be affected differently by Na^+ and K^+ salts. Apparently, there are several non-cooperative binding sites with different specificity and affinity for these ions. Hill coefficient analysis suggested that the enzyme could have at least six cooperative binding sites per dimer for K^+ ions [51]. In the psychrophilic *Pseudomonas* sp., PGK activation is influenced by sulfate ions in a concentration-dependent manner. At lower concentrations (below 20 mM), an activation of the enzyme occurs, while at higher concentrations (50 mM), it induces a significant inhibition of enzyme activity [65]. Studies with some protists revealed similar results to those found in kinetic studies of prokaryotic PGKs. In *T. brucei*, the cytosolic and glycosomal PGK isoenzymes (recombinant and native) are negatively affected by sodium salts (NaCl and NaSO₄) [69]. The activity of the corresponding PGK isoenzymes of *T. cruzi* appears also to be negatively affected by NaCl (at concentrations above 150 mM). In the case of *T. cruzi*, this inhibition is probably due to reduction in the solvation capacity that leads to the formation of oligomeric forms other than monomers and loss of PGK activity (W. Q. *et al.* 2006, unpublished results). It is important to note that, contrary to what has been observed in trypanosomatids, *P. falciparum* PGK is stimulated by high concentrations of Na^+ and K^+ salts (up to 200 mM). In addition, it has been proposed that the basic patch in the N-terminal domain of this enzyme could be involved in catalytic regulation by divalent anions, similar to yeast PGK (see §3.1) [47,78].

5.2. Regulation by nucleotides

PGK activity is regulated by various nucleotides (AMP, ATP, ADP), as has been documented for different organisms [25,64,71]. Rabbit muscle PGK inhibition occurs by the presence of AMP, ADP, GDP, GMP, IDP and IMP. Apparently, the inhibitory action is of a mixed type [71]. This purine nucleotide-mediated regulation has also been observed in plants. The activity of cytosolic and chloroplast PGKs of *Pisum sativum* in the direction of ATP generation is regulated by AMP and ATP, while in the opposite direction (use of ATP), regulation occurs through AMP and ADP [136]. For yeast, there are references for the inhibition of PGK by ADP and AMP. Apparently,

this enzyme has two nucleotide-binding sites, one binding site for the substrate ATP/ADP and another regulatory site for ADP separate from that for the nucleotide substrate. AMP also binds to the enzyme, probably at the same site as the substrate ATP. In addition, ADP is a competitive inhibitor, while AMP is a non-competitive inhibitor of 3PGA binding. This mechanism of inhibition might reflect an interesting way to regulate the direction of the enzyme's reversible reaction [137].

In bacteria such as *C. glutamicum*, inhibition of the ATP/3PGA-dependent reaction of PGK by ADP with a low Ki value (0.1 mM) has been documented [64]. The apparent mixed-type inhibition by ADP reveals the key role of PGK in gluconeogenesis. In trypanosomatids, regulation of PGK isoenzymes by nucleotides has also been reported [25]. *Trypanosoma cruzi* PGKC, a glycosomal enzyme, is inhibited at concentrations higher than 230 µM ATP, with a Ki of 270 µM. The inhibition of this isoenzyme by ATP has led to the proposal of a key role of this PGK in the regulation of glucose metabolism, specifically in gluconeogenesis. It is likely that this enzyme is optimized to work in the gluconeogenic direction even at low ATP concentrations by having a high affinity for this substrate. So far, this PGKC from *T. cruzi* is the only report that is available about a PGK of kinetoplastids regulated by ATP.

5.3. Regulation by redox state

In photosynthetic organisms, thioredoxin-dependent redox regulation of enzyme activities serves as a mechanism to control a large number of cellular processes, including the Calvin–Benson cycle. Most of the enzymes involved in this cycle are activated by light through dithiol/disulfide exchanges controlled by a chloroplastic thioredoxin. Several studies have shown that chloroplast PGK (chlPGK) is one of the enzymes undergoing such redox regulation [55,138]. Also in the cyanobacterium *Synechocystis* sp. (PCC6803), PGK activity is under redox control. A chloroplast-type thioredoxin (Trx) is responsible for maintaining this PGK in its reduced and active form, even under conditions of oxidative stress. Critical amino acids for this redox regulation are a cysteine pair, Cys314 and Cys340, located in the C-terminal domain of the *Synechocystis* PGK. The role of these two Cys residues for such regulation is linked to their proximity to the catalytic site and the associated conformational change that promotes their oxidized state [55]. In the unicellular green alga *Chlamydomonas reinhardtii*, a similar redox regulatory mechanism for chlPGK1 has been reported, mediated primarily by a Trx enzyme that is active during the light phase. However, structural analyses showed that the formation of disulfide bonds, mediated by two cysteine residues (Cys227 and Cys361), is a phenomenon that does not directly affect the affinity for the substrates; however, it has an obvious impact on the turnover and catalytic efficiency of the enzyme [139]. In addition to this redox regulation of the PGK, dependent on Trx, the chlPGK of *C. reinhardtii* was identified as a potential candidate for other post-translational redox modifications such as *S*-thiolation, glutathionylation and *S*-nitrosylation. It was found that some Cys residues, such as Cys361, may be subject to modification by glutathione. This means that these residues can also play a regulatory role under certain conditions or in the presence of specific other proteins [139–141]. It is possible that chlPGK can be regulated by multiple redox mechanisms. The complex interaction between the variable environmental conditions and

the intracellular redox state perhaps determine the type and extent of each redox modification to which it is subjected. Additionally, these mechanisms could contribute to the fine adjustment of carbon fixation in photosynthetic organisms [139].

5.4. Regulation by non-coding RNAs

For tumour cells also, a regulation of PGK1 function by specific non-coding RNAs, microRNAs (miRNAs) and long non-coding RNAs (LncRNAs), has been reported [142–144]. In hepatocellular carcinoma (HCC), the dysregulation of miR-450b-3p induces inhibition of cell viability, colony formation and cell-cycle progression *in vitro*. This effect is attributed to the inhibition of PGK1 expression. Apparently, the 3′-UTR of the PGK1 gene contains a complementary sequence of miR-450b-3p; therefore, PGK1 probably is a direct target of miR-450b-3p. This regulation of the PGK not only has an effect on its expression, but also on its function through an inhibition of the PGK1-mediated AKT phosphorylation in HCC cells [142]. AKT is a kinase that acts as a key regulator of cell growth, cell-cycle progression and apoptosis [145]. In colorectal cancer cells, another miRNA, miR-548c-5p, suppresses proliferation by recognizing the 3′-UTR of *PGK1* to decrease the expression of this gene [143]. Unlike miRNAs, LncRNAs have an opposite effect on the function of PGK1 by acting at the post-translational regulation level [144,146]. Through RNA pull-down assays and immunoblotting, it was confirmed that MetaLnc9 interacts directly with the PGK1 in lung cancer NSCLC cells. This interaction prevents PGK1 degradation by ubiquitination and stimulates the PGK1-activated oncogenic AKT/ mTOR signalling pathway, eventually promoting metastasis [144]. This mechanism of PGK regulation through direct interaction with an LncRNA has also been documented for gall bladder carcinoma (GBC). The direct interaction of LncRNA GBCDRlnc1 with PGK1 also prevents its ubiquitin-mediated degradation; however, in this case, the consequence of such interaction is the induction of chemoresistance of GBC cancer cells by activating autophagy. This activation could be mediated through the regulation of the autophagy-associated proteins ATG5 and ATG12 [146].

6. PGK as a moonlighting protein

Moonlighting proteins comprise a heterogeneous collection of proteins from different classes that can perform multiple physiologically relevant biochemical or biophysical functions. Currently, more than 300 moonlighting proteins have been identified. These proteins are expressed in organisms throughout the 'tree of life', with their primary functions being attributed to different biochemical processes. Some of these proteins have the ability to perform their primary and secondary—moonlighting—functions simultaneously, while others change their function in response to environmental changes [147,148]. Moonlighting proteins can provide many potential benefits to an organism, such as the coordination of cellular activities. Proteins with moonlighting functions include receptors, transcription factors, adhesins and metabolic enzymes [149]. Among the metabolic enzymes, seven glycolytic enzymes have been identified as moonlighting proteins, one of them being PGK [150–152]. In various organisms, other functions in various cellular processes are attributed to PGK in addition to exerting its canonical

function in glycolysis and gluconeogenesis, including roles in parasite–host relations (figure 4).

6.1. Cell invasion

In bacteria, PGK has been generally reported as a soluble, cytosolic protein, but a surface-associated form has also been described for the enzyme of several species. In some pathogenic bacteria, PGK has been identified as a major outer surface protein [35,153–156]. In isolates belonging to the genus *Streptococcus*, the enzyme is expressed on the cell surface, where it seems to be involved in pathogen–host interaction through binding to host proteins such as plasminogen and actin. In the case of plasminogen, the binding to streptococcal PGK occurs specifically through its angiostatin domain (kringle domains 1–4) [35]. It has been suggested that this interaction induces a conformational change in plasminogen that leads to its opening and facilitated conversion to plasmin [157,158]. Also, phytopathogenic organisms such as *Spiroplasma citri* contain a membrane-associated PGK that interacts with actin of the *Circulifer haematoceps* insect vector during the invasion of the bacterium into the insect's salivary glands [155]. Peptide mapping and site-directed mutagenesis studies indicated that the binding of the streptococcal PGK to both plasminogen and actin is mediated through three lysine-rich regions in the PGK structure [159]. This interaction is a key event for tissue degradation, cell adhesion and immune evasion during the infection process [35,59,156].

6.2. Nuclear roles

During evolution, cells have developed several strategies to detect and adapt to environmental changes, including those that cause changes in cellular energy levels. Studies have demonstrated the existence of a link between the metabolic flux and the regulation of gene expression. This relationship involves also moonlighting metabolic enzymes in the nucleus. Apparently, the existence of metabolic enzymes in the nucleus allows to provide information about the cell's nutritional and/ or energy status by sensing the levels of metabolites and cofactors such as ADP, ATP and NAD(H) which can freely permeate from the cytosol to the nucleus through the nuclear pores. By their acquisition of also a role in the regulation of gene expression, such enzymes may link the supply and demand of energy and nutrients to gene transcription and so perhaps provide a way by which the cell can establish a quick and efficient adaptative response [160].

A role for the regulation of nuclear processes by the energy/redox state has been shown in various studies. For example, the redox state controls the activity of transcription factors such as NPAS2 and CtBP that are involved in the maintenance of mammalian circadian rhythms and in cell differentiation and development, respectively. This is exerted by differential binding of NAD^+ and NADH to lactate dehydrogenase, acting as a cofactor with NPAS2, or directly to CtBP [161,162]. With regard to ATP, chromatin remodelling and several forms of hormonal gene regulation are known to require this cofactor [163].

There are several reports about the presence of glycolytic enzymes in the nucleus. To some of these enzymes, functions have been attributed associated with processes such as DNA replication and repair, and histone methylation and acetylation [160,164–171].
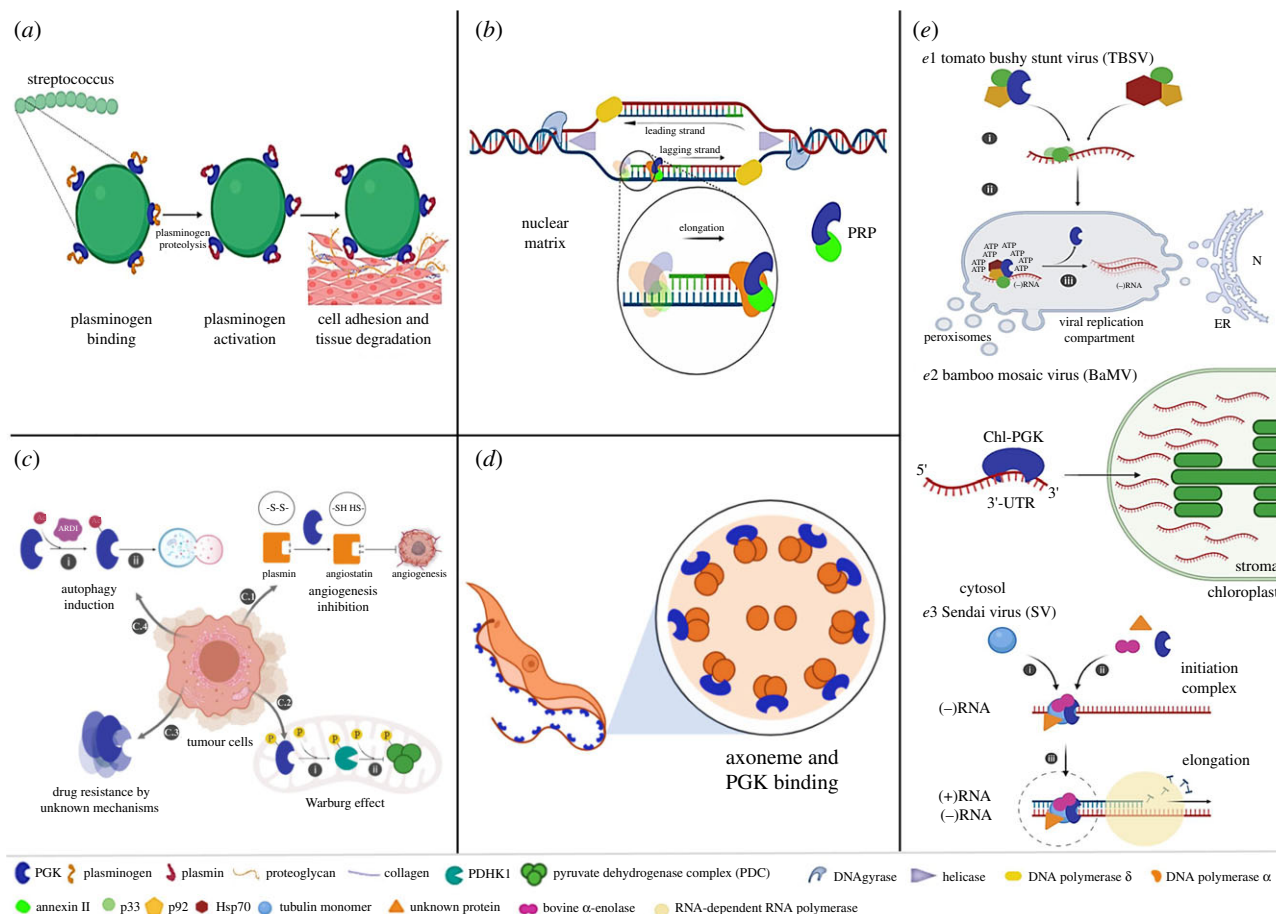
Figure 4. PGK moonlighting functions. (*a*) Cell invasion: PGK is located at the surface of some pathogenic bacteria where the processing of plasminogen to plasmin is promoted. (*b*) DNA replication: Together PGK and Annexin II protein constitute the Primer Recognition Particle (PRP) localized in the nuclear matrix. (*c*) Tumour growth and cancer progression: C.1. PGK is secreted from cancer cells. In the extracellular space, PGK allows the conversion of plasmin into angiostatin and the subsequent inhibition of angiogenesis. C.2. PGK phosphorylated at residue S203 is translocated into mitochondria to activate PDHK1 by phosphorylation (I). PDHK1 phosphorylates and inhibits the catalytic activity of the pyruvate dehydrogenase complex (PDC) and downstream metabolic reactions in mitochondria (II). C.3. Overexpression of PGK involved in drug resistance. C.4. PGK posttranslational modification (acetylation of residue K388) by ARD1 (I) induces PGK activation and subsequent phosphorylation of intermediate proteins that finally switch on the autophagy mechanism (II). Acetylation of other K residues contributes to increasing the glycolytic flux of tumour cells. (*d*) Functions associated with the flagellum. (*e*) Viral replication in TBSV: PGK and other host proteins are involved in viral mechanisms and able to bind to viral RNA. PGK also provides local ATP required in viral replication. (*e*.2) The bamboo mosaic virus (BaMV): chloroplast PGK (chlPGK) interacts with 3′-UTR of Viral RNA to direct the translocation of viral RNA from the cytosol to the chloroplast and allow its accumulation in the stroma. (*e*.3). In Sendai virus (SV): PGK participates in the initiation complex and stimulates viral gene transcription.

PGK has also been detected in the nucleus of both animal and plant cells [56,169,172,173]. In the pea *P. sativum*, both chloroplast and cytosolic PGK are also present in nuclei of leaf and root cells, where they were located primarily in the euchromatin. This location in the euchromatin might suggest participation with DNA polymerase α in replication. Both enzymes could act as nuclear receptors for cytosolic metabolites (e.g. ATP, ADP and/or phosphoglycerates) signalling the metabolic status of the cell [169]. Additionally, there are reports about mammalian cells where DNA synthesis catalysed by DNA polymerase α and ε on single-stranded DNA is stimulated by PGK. This activation might be due to a primer recognition capability of PGK. In this context, PGK, in complex with annexin II in the nuclear matrix, would act as a Primer Recognition Protein (PRP), with both proteins acting together as cofactors of DNA polymerase α, fulfilling a role in lagging strand DNA replication. [56,169,174]. An alternative suggestion is that PGK, in the presence of ATP, could regulate the replication complex through the phosphorylation of some proteins (such as DNA polymerase α-Primase B Subunit) [169] forming this complex

that undergoes a cell cycle-dependent phosphorylation and dephosphorylation [175].

Together with other glycolytic enzymes in the nucleus, PGK could have a leading role in the modulation of repair and the replicative synthesis of DNA. By retaining their binding capacity for cofactors, substrates and inhibitors (which are mainly those of energy metabolism), one can imagine they establish a regulatory link between the energy status of a cell, and its DNA replication and repair functions [169].

6.3. Angiogenesis, tumour growth and drug resistance

There is growing evidence demonstrating that PGK performs, in some cases, a role in mediating tumour suppression mechanisms while in other cases promoting cancer initiation and progression, and drug resistance [53,57,176–178]. In different tumour cell lines, PGK participates in angiogenesis by acting as a disulfide reductase [57,176]. The PGK is secreted into the environment where it reduces plasmin that will generate angiostatin, a tumour angiogenesis inhibitor [57,179]. The formation of angiostatin can occur through mechanisms

that involve the autoproteolysis of plasmin in alkaline conditions or due to proteases, specifically serine and metalloproteases [57,179]. However, when PGK is involved, the process seems to be a thiol-independent phenomenon which facilitates reduction in the disulfide bond(s) in plasmin [176]. This is a non-conventional mechanism due to the lack of the typical sequence Cys–Gly–X–Cys normally present in the active sites of protein reductants, but in this case involving predominantly hydrophobic interactions. Apparently, when PGK binds to plasmin, it induces a conformational change in kringle 5, by which the target disulfide bond (Cys-512–Cys-536) becomes accessible for nucleophilic attack. These events allow the exposure and proteolysis of target site residues in the plasmin's kringle 5 [176].

With regard to the initiation and progression of cancer, PGK can function as an oncogene, through its participation in oncogenic signalling pathways (such as AKT/Mtor, MYC, Nocth, CXCR4/β-catenin) [142,180–182]. Importantly, this enzyme and its post-translational modifications (PTMs), such as phosphorylation, acetylation, ubiquitination and succinylation, play a key role in regulating tumour metabolism. Depending on the PTMs to which this enzyme is subjected, there is an inhibition or induction of cell growth and progression [53]. PGK in tumour cells can behave like a canonical kinase, phosphorylating other proteins. Li *et al.* [183] discovered that in tumour cells, under certain conditions, translocation of PGK1 to mitochondria is induced. Once within mitochondria, this PGK1 functions as a PK, phosphorylating pyruvate dehydrogenase kinase (PDHK1). The subsequent phosphorylation of the pyruvate dehydrogenase complex (PDH) by this PDHK1 induces inhibition of the complex, subsequently causing suppression of the mitochondrial metabolism of pyruvate, thus promotion of the Warburg effect with an increase of lactate production [183].

On the other hand, drug resistance of tumour cells is associated with several factors, such as an enhancement of the cellular anti-apoptotic capacity, improved ability to repair DNA damage and an upregulation of the expression of the ATP-dependent drug transport protein. A significant amount of clinical and experimental evidence has shown that overexpression of intracellular PGK1 plays also an important role in resistance to a wide variety of medications and radiation [53]. The precise role of PGK1 in drug resistance mechanisms remains so far unknown. Some proposed suggestions are that it may be linked to increasing the glycolytic flux in response to drug-induced tumour starvation and intratumoral hypoxia, upregulating autophagy, triggering the HSP90/ERK pathway mediated DNA repair and methylation and/or scavenging ROS and inhibiting ROS-stimulated apoptosis [53].

In liver cancer cells, SIRT7 (sirtuin type 7) regulates the acetylation of PGK1 residue Lys323, a mechanism that appears to enhance the efficiency of the glycolytic pathway during the Warburg effect, and consequently promoting cancer proliferation and tumorigenesis [184].

6.4. Autophagy induction

A moonlighting role of PGK has also been found in its involvement in autophagy induction in human tumour cells, a process that requires its prior activation via lysine acetylation. Specifically, Lys388 is acetylated by the acetyltransferase ARD1 under glutamine deprivation and hypoxia conditions, when mTOR cannot inhibit ARD1 by phosphorylating its Ser228

residue. Subsequent to the ARD1-dependent PGK1 acetylation are the interaction with Beclin1 and the activation of the latter by its phosphorylation of residue Ser30, inducing activation of the Beclin1–VPS34 complex through conformational changes that increase the affinity of VPS34 to phosphatidylinositol. Hence, the generation of increased phosphatidylinositol 3-phosphate (PI3P) finally promotes the occurrence of autophagy [185].

6.5. Functions associated with the flagellum

The presence of a large (110 kDa) PGK-like protein (designated PGKL) and other proteins related to glycolytic enzymes associated with the flagellum of trypanosomatids suggests the possibility of the presence of a flagellar-associated glycolytic pathway that disappeared during early evolution of these parasites [60]. Studies involving a null mutant of *T. brucei* corroborate that PGKL's presence and function in the flagellum is not essential, at least in the procyclic, insect-stage form of the parasite and under the growth conditions evaluated. The PGKL was found to be tightly associated with the axoneme and most likely catalytically inactive since sequence comparison and structural analysis with functional PGKs demonstrated no conservation of key catalytical residues, similarly as observed for a homologous PGKL protein identified by genomic analysis of several other trypanosomatids and the distantly related kinetoplastid *Bodo saltans*. Brown *et al.* [60] suggested that glycolytic enzymes may have relocated from the flagellum to other cell compartments of the trypanosomatids (peroxisomes, giving rise to glycosomes), possibly as a response to environmental changes in the niche. Further details about the occurrence of this PGK-like proteins in kinetoplastid protists are discussed in §8 of this review.

6.6. Viral replication

PGK with moonlighting activity is also implicated in viral replication mechanisms [61]. Recruitment of host cellular components is promoted to allow the formation of replication particles of tomato bushy stunt virus (TBSV). TBSV replication requires the aggregation of peroxisome and ER compartments as well as the recruitment of several host proteins (Hsp70, Vps4p AAA+ ATPase, DEAD-box helicases) which participate in the assembly of the viral replicase complex, and in the activation of polymerases, as was evidenced for Hsp70. PGK is also recruited into this viral replication assembly. Its likely role was demonstrated to be the local supply of energy in the form of ATP for use by the molecular chaperones that promote efficient replicase assembly and activation of proteins involved in viral replication [61]. Interestingly, the supply of ATP during viral assembly is not the only function reported for PGK in viral mechanisms. In other studies, the enzyme was found to be involved in bamboo mosaic virus (BaMV) infection, interacting with the 3′-UTR of the viral RNA. Specifically, chloroplast PGK (chlPGK) was shown to be involved in the accumulation of BaMV in the *Nicotiana* cells, targeting the viral RNA to chloroplasts in the early stage of infection, to allow the viral replication within these organelles [186]. This chlPGK, together with the host heat shock protein Hsp90, translocate the viral ribonucleoprotein complex (RNP) across chloroplast membranes to the stroma. Subsequently, assembly of viral replication complexes

(VRC) in association with the thylakoid membranes is stimulated [187].

Furthermore, it has been shown that bovine PGK, other mammalian PGKs and yeast PGK can successfully stimulate the *in vitro* synthesis and elongation of Sendai virus RNA. However, PGK catalytic activity appeared not required to stimulate this transcription, but in association with tubulin, the transcript elongation rate was drastically increased, as was also demonstrated for α-enolase [188,189]. Therefore, the experimental evidence summarized above demonstrates that PGK participates as a host protein factor implicated in different steps of the replication process of some viral entities.

7. Phosphoglycerate kinase in protists

PGK has been studied in detail in several protist organisms, both parasitic and free-living ones. In these organisms, the enzyme has been found in different subcellular compartments and different physiological functions have been attributed to it. Additionally, as we will see in the following sections, some of these organisms have different isoenzymes (table 3) and some of these PGKs undergo PTMs necessary to carry out their function.

7.1. Free-living protists

7.1.1. *Tetrahymena thermophila*

Tetrahymena thermophila is a species of free-living ciliates, closely related to dinoflagellate and apicomplexan protists [190]. Like other ciliated protists, this organism is characterized by exhibiting nuclear dimorphism. It has a large non-germinal nucleus and a small nucleus which are cytologically and functionally different. Each of these nuclei is specifically used during the two different stages of the life cycle. The micronucleus (MIC) is transcriptionally silent and only plays a role during sexual life stages. It has a diploid genome consisting of five pairs of chromosomes, while the macronucleus (MAC) contains 200 different chromosomes that encode inheritable information passed from one sexual generation to the next [190–192]. The two most important functions attributed to the MIC are the generation of competitive sexual progeny and to provide genetic variation for improving the chances of survival of the progeny during periods of environmental changes [193]. For its part, the MAC is transcriptionally active, and its products maintain and control somatic cell functions during vegetative growth. The MAC chromosomes of *T. thermophila* encode multiple genes and exist, on average, in approximately 45 copies of each chromosome per MAC. The MAC genome directly determines the phenotypic characteristics of the cells [194]. Most genomic studies of *T. thermophila* have focused on the MAC genome [195–197]. When analysing the genome of this organism [85,108] we found two candidate genes that code for PGK enzymes (table 3). One of these isoenzymes is a canonical PGK [85,108,198], while the other is a multidomain PGK containing also a sequence related to the MYCBPAP protein family (i.e. proteins associated with binding the metazoan transcriptional activator Myc, so far not described for *T. thermophila*; table 3). Some of these MYCBPAPs, such as AMAP-1, associated with AMY-1 (amylase-alpha 1A) bound to Myc, may have a role in spermatogenesis in human testis cells [199].

Transcriptomic studies showed that both PGK genes are expressed in different phases of *Tetrahymena*'s cell cycle, during each of the three major physiological/developmental stages: growth, starvation and conjugation [108]. During the last stage, the gene TTHERM_00856750 (MYCBP-PGK) is differentially expressed in time, between 6 and 18 h after having initiated the sexual process. Its expression pattern is similar to that of other metabolic enzymes, and proteins such as kinesins, alkyrin, heterochromatin-binding proteins and TATA-binding proteins (TetraFGD, http://tfgd.ihb.ac.cn). In turn, the gene TTHERM_00929450 for the canonical PGK is only expressed 10 h after starting the conjugation. TTHERM_00929450 has an expression pattern very similar to other enzymes involved in glycolysis, the pentose-phosphate pathway and Krebs cycle. Notably, there are reports about the importance of these proteins in different cellular processes, especially chromosome segregation during mitosis and meiosis, programmed DNA elimination and MAC development [200–203]. One can imagine that this MYCB-PAP-PGK is involved in any of the processes that occur during conjugation, when gene expression correlates with the stages of meiosis, nuclear differentiation and DNA elimination [204]. The expression levels of MYCBPAP-PGK increase near 6 h post-conjugation, coinciding with the time of new MAC development [201]. It is possible that this MYCBPAP domain allows us to activate the transcription of specific regions of the MIC genome related to the formation of the *T. thermophila* MAC. Therefore, this PGK could act as a moonlighting protein, fulfilling additional functions besides its metabolic one.

7.1.2. *Euglena gracilis*

Euglena gracilis is a photosynthetic protist in which two isoforms of PGK with different subcellular locations were identified. One was located in the chloroplast (cpPGK) and the other in the cytosol (cPGK) (table 3) [110]. PGK was included in the 1345 candidate plastid proteins of a proteomic analysis [205]. The cpPGK isoform is synthesized as a precursor polyprotein, which yields two mature protein units (cpPGK) by its processing, after having been imported into the chloroplast. The processing is regulated by a thiol protease, located in the chloroplast stroma [206–208]. The result of this processing is two almost identical cpPGK proteins of 423 amino acids that differ in only one residue, Asp422 versus an Asn [110]. The strong similarity of *E. gracilis* cpPGK with its cytosolic homologues from other protists, animals and fungi suggests that this isoform is a molecular relic, in the sense that it is the only representative of the original cytosolic PGK found in chloroplasts among photosynthetic eukaryotes to date. With regard to the *E. gracilis* cytosolic PGK isoform, it is an orthologue of the cytosolic and glycosomal PGKs in the Kinetoplastea (see §8). Apparently, after an ancestor of the euglenid lineage within the Euglenozoa (figure 1) engulfed a chlorophytic alga, a non-endosymbiotic PGK replacement occurred in the *E. gracilis* plastid [110].

In addition to the cpPGK isoenzyme, other proteins such as the small subunit of RuBisCO (SSU) and the light-harvesting chlorophyll proteins (LHC) are synthesized as polyproteins in *E. gracilis* [209,210]. These proteins are part of a group of nuclear-encoded plastid-targeted proteins that are translated as polyprotein precursors. Once synthesized, they are routed to the chloroplast via complex N-terminal

targeting sequences. Inside the chloroplast, the individual proteins are liberated through proteolytic cleavage of conserved decapeptide linkers [210,211]. Like for other chloroplast proteins, the expression, transport and processing of the SSU and LHC proteins are controlled by light and temperature [210,212]. In the case of cpPGK, some similar process is likely to happen. Studies based on Y_9 mutants of *E. gracilis* have shown that the expression of this enzyme is influenced by light [213], although little is known about the molecular basis and function of this expression. Intriguingly, *Chlorella* chloroplast PGK exhibits circadian expression and appears to be an essential clock protein [214]. It would be interesting to study the physiological significance of a possible circadian regulation of PGK in *E. gracilis*. It could provide understanding of circadian regulation of glycolytic genes in trypanosomatids such as *T. brucei* [215].

7.1.3. Dictyostelium discoideum

Dictyostelium discoideum is a free-living amoebozoan. It is a so-called social amoeba where, during its life cycle, individual protists can aggregate into a multicellular slug and then into a fruiting body, from which spores are formed that in turn differentiate back into amoebae. The single PGK present in this protist—called PGKA—is located in the cytosol as in most other eukaryotic organisms and has additional functions to its canonical ones. Sequence analysis revealed the presence of a putative calmodulin-binding domain in its structure, as reported by Myre & O'Day [128] and Catalano & O'Day [216]. The calmodulin-binding domain is situated next to the hinge and within the ATP-binding domain [128]. The activity of this PGK may be negatively regulated by calcium via calmodulin-mediated signalling (Ca^{2+}-CAM-mediated signalling) [128]. Until now, the functionality and physiological significance of such putative regulatory mechanism in *D. discoideum* PGK are still unknown.

Calmodulin (CAM), being a sensor protein of intracellular calcium fluxes, controls metabolism in rabbit skeletal muscle through regulation of glycolysis and mitochondrial oxidative metabolism [217,218]. There are reports of glycolytic enzymes regulated by CAM binding [219]. Muscle phosphofructose kinase (PFK-M) is a CAM-binding protein. CAM atypically acts as a Ca^{2+}-dependent inhibitor of PFK-M polymerization. The active tetrameric PFK is stabilized in an inactive dimeric form, upon inhibitor binding in the central region linking the two PFK dimers. This regulation by Ca^{2+}-CAM-mediated signalling would lead to the existence of two PFK pools, a highly active pool of enzyme (A-conformers), and a pool of partially polymerized enzyme of very low catalytic activity in the D conformation [219,220]. In the case of PGKA of *D. discoideum*, it would be valid to think that the activity of this enzyme is probably modulated in response to Ca^{2+} signalling, similar to that has been reported for other glycolytic enzymes such as GAPDH associated with the sarcoplasmic reticulum membrane in skeletal muscle [221]. During the life cycle of this protist, Ca^{2+}-CAM signalling plays important roles in processes such as chemotaxis, aggregation and development [222,223].

7.1.4. Naegleria gruberi

Naegleria gruberi is a ubiquitous freshwater flagellated amoeba, characterized by drastic morphological changes, from having a true amoeboid form (which lacks a cytoplasmic microtubule cytoskeleton) to a flagellate (which has an elaborate microtubule cytoskeleton) [113]. Despite the few studies of some metabolic enzymes, so far little knowledge is available about the metabolism of *N. gruberi*. However, the sequencing of its genome and some transcriptomic studies allowed a broader view of the metabolic capacities of this protist [224,225]. Analysis of its genome suggested that this organism can use a wide variety of carbohydrates [225]. Experimental studies have shown that glucose is an important source of carbon for *Naegleria*; its presence stimulates growth and reduces the generation time [226]. Additionally, transcriptomic analysis revealed that genes involved in metabolism (including glycolysis, Krebs cycle and pyruvate–acetate metabolism) are differentially regulated during the differentiation process [224]. Although most regular glycolytic enzymes are present in *Naegleria*, this route apparently differs from that found in most other eukaryotic organisms [224]. The differences concern two aspects: (i) it lacks a hexokinase (HK), instead it has a glucose-specific glucokinase that is very similar to that found in trypanosomatids [225,227–229]; and (ii) the second phosphorylation step of glycolysis is catalysed by an inorganic pyrophosphate (PPi)-dependent PFK, instead of an ATP-dependent one [230]. Regarding the other glycolytic enzymes detected in the genome of this organism, a single *pgk* gene was identified [224]. It codes for a canonical PGK that is similar to those previously described for other protists [113]; table 3. It is noteworthy that *N. gruberi*, although it passes through significant morphological and physiological changes, contains in its genome only one gene for a PGK with a canonical structure. This enzyme is present in the cytosol [225], but *Naegleria* differentiation studies identified PGK, together with other glycolytic enzymes, also as a protein associated with the flagellum [113]. The expression of this enzyme occurs during the 60 min after the onset of the differentiation process, which coincides with the morphological transition from cysts to flagellates [113]. Although the function of PGK in the flagellum of *Naegleria* is unknown, the presence of glycolytic enzymes in flagella of other organisms (from protists to mammals) has led some authors to suggest that these enzymes may provide a localized source of ATP that is essential for cellular motility [60,231].

Naegleria catabolizes glucose through glycolysis and the Krebs cycle, with the electrons of NADH and $FADH_2$ fuelling the branched respiratory chain, while fatty acid catabolism occurs through β-oxidation followed by respiratory chain activity [232]. Though *N. gruberi* contains all genes necessary for the metabolism of amino acids, fatty acids and carbohydrates [225], this protist, during its growth, exhibits a notable preference for fatty acid oxidation as the main source for ATP synthesis. Like in other organisms, fatty acid oxidation occurs in both peroxisomes and mitochondria [232]. This preference of nutritional source is a distinctive feature of *Naegleria*, not documented for other pathogenic protists [232], although the role that glycolysis plays in covering the metabolic needs of the related pathogenic species *N. fowleri* is poorly understood. It has been shown that glucose can be an important source of carbon for this protist during human infection [233,234]. Trophozoites cultured in medium without glucose leads to a reduction in growth and induces encysting. Additionally, the addition of 3-bromopyruvate, a well-known inhibitor of glycolysis in tumours, to standard glucose-containing medium, also induces a dramatic decrease in the growth of this protist [234], in a similar way as in trypanosomes [235].

The PGK enzymes from the free-living protists described in §§7.1.1–7.1.4 have very similar characteristics to each other, except for the presence of an additional domain in the case of one of the PGKs of *T. thermophila* (MYCBPAP) and *D. discoideum* (calmodulin-binding domain). They have a similar number of amino acids and molecular weight (between 45 and 46 kDa, except for the *E. gracilis* cpPGK, which has a weight of 105 kDa before being processed), and apparently, most have only the proper PGK domain associated with its structure (table 3). In some cases, it has a dual subcellular localization, in the cytosol and essential organelles such as the *E. gracilis* chloroplast and *N. gruberi* flagellum (table 3). In these free-living organisms, the PGK seems often to be linked to functions beyond glycolytic/gluconeogenic metabolism, such as catalysis of the reverse PGK reaction as part of the Calvin cycle and with a regulatory role (for PGK functioning as a glycolytic/gluconeogenic enzyme) controlled by Ca^{2+}-CAM signalling. Through its pattern of expression, it may also be involved in processes like the circadian cycle in some of these free-living protists.

7.2. Parasitic protists

7.2.1. Amitochondriate protists

Entamoeba histolytica, the aetiological agent of amebiasis, depends exclusively on glycolysis for the generation of ATP, because it lacks the Krebs cycle and oxidative phosphorylation [236–239], a metabolic adaptation shared with two other pathogenic amitochondriate protists, *Giardia lamblia* and *Trichomonas vaginalis* [238]. Strikingly, these parasites, living in anaerobic/microaerophilic environments, use PPi instead of ATP as the high-energy phospho group donor in glycolysis [230]. *Entamoeba histolytica* has an unusual PGK. This monomeric PGK with a canonical structure (table 3) is selective for guanine nucleotides rather than adenine nucleotides [21]. The extent of the preference for GTP varies with the direction of the reaction that this enzyme catalyses. In the glycolytic direction, the selectivity for GTP over ADP is 150-fold, whereas in the gluconeogenic one, it is about 50-fold [240]. The presence of a guanine nucleotide-dependent PGK, responsible for the first substrate-level phosphorylation reaction of glycolysis, could most likely be indicative of changes in its interplay with other metabolic pathways and cellular processes in the parasite [21]. In addition, this peculiar nucleotide preference may have physiological relevance. The simultaneous presence of a GTP-dependent PGK and a nucleoside diphosphate kinase, previously identified [241], should be able to make ATP readily available from the GTP pool [21]. Alternatively, since *E. histolytica* lacks a de novo purine synthesis pathway [242], this PGK could serve to provide a source of GTP for DNA synthesis.

The PGKs of *G. lamblia* and *T. vaginalis* have been less well studied. *Giardia lamblia* has a PGK which is more alike to those found in most other eukaryotes [243]. For *T. vaginalis*, there have been several molecular and biochemical studies related to PGK. Two candidate *pgk* genes have been reported [114,244]. The encoded enzymes are 98% identical, have the same predicted molecular weight, a similar pI and both are apparently catalytically active (table 3). Probably, they are the products of a recent *pgk* gene duplication. In previous studies, it has been documented that a distinctive feature of the *T. vaginalis* genome is the retention of multiple paralogous copies of some genes, including those encoding glycolytic enzymes [244]. The expression of these isoenzymes is influenced by environmental factors such as temperature and nutritional condition [114,244,245]. PGK is one of the enzymes with a relatively high activity in *T. vaginalis* and, unlike other glycolytic enzymes (HK, aldolase (ALD), enolase (ENO), phosphoglycerate mutase (PGAM)), its activity is not affected by the presence or absence of oxygen [246]. Additionally, transcriptomic studies focused on regulation mechanisms of metabolic pathways have shown that the expression of the two PGK isoenzymes is downregulated under limited availability of glucose [114]. Additionally, the glycolytic pathway is apparently one of the major iron-regulated pathways in *T. vaginalis*. Transcription of at least one (but usually several) of the multiple gene copies encoding enzymes involved in glucose metabolism showed significant iron-dependent regulation. Especially those enzymes that supply substrates to the pathways involved (such as glucokinase, glycogen phosphorylase and phosphoglucomutase), as well as some other enzymes such as PGK, are upregulated [244]. This could be related to the fact that establishment of *T. vaginalis* infections is dependent on the efficient acquisition of essential nutrients such as iron from the host environment. The iron ion performs important functions in the cytoadherence of *Trichomonas* to host target cells [247], the expression of diverse proteinases [248], and it increases resistance to complement-mediated lysis [249].

7.2.2. Apicomplexan protists

In the genome of the apicomplexan organism *Toxoplasma gondii*, genes for all glycolytic enzymes from PFK down to PYK are present as duplicates. They code for proteins with different subcellular localization, with one set of enzymes constituting a complete pathway in the cytosol [115]. The expression of the two PGK isoforms (table 3), PGKI and PGKII, shows no significant difference between the two life cycle stages living in the warm-blooded vertebrate host, the rapidly proliferating tachyzoites and the slowly proliferating, encysted bradyzoites. Interestingly, PGK II (and isoenzymes of some other glycolytic enzymes) has a peculiar localization, within the apicoplast, a non-photosynthetic plastid relict characteristic of this group of organisms (except in *Cryptosporidium*—see below—that has lost it) [115]. PGK II serves to convert 1,3BPGA to 3PGA that is subsequently transferred to the cytosol through a system (pPTS) that transports phosphorylated glycolytic intermediates in antiport with inorganic phosphate (Pi) across the inner envelope membrane [250]. In the cytosol, this 3PGA can be converted to PEP that could be directed to the apicoplast through a PEP/Pi translocator (PPT) where it is converted to pyruvate by PYK. The pyruvate is then used as a substrate for the pyruvate dehydrogenase (PDH) to produce acetyl-CoA, necessary for the synthesis of fatty acids [250].

In *P. falciparum*, the agent that causes malaria, anaerobic glycolysis is known already for long time as the main source of ATP in the erythrocytic stage, while oxidative phosphorylation may play a more important role in the mammalian liver and mosquito vector stages [251,252]. Thus, the parasite's enzymes of the glycolytic pathway are essential for the development of its asexual blood stages [20]. The PGK has been characterized at the molecular, structural and biochemical levels [20,253,254]. Our genome analysis demonstrated the presence of a single *pgk* gene on chromosome 9 of this parasite,

in agreement with the result previously reported by [253]. The crystal structure of this glycolytic enzyme in the open conformation is similar to the structure of the monomeric PGKs from other organisms [47] (table 3). However, the kinetic properties of the parasite enzyme differ from those reported for its human counterpart. This has led to consider it as a potential target for chemotherapy. It has a greater affinity for its substrates than the human PGK. Additionally, it can use different divalent cations (Ca^{2+}/Mg^{2+}/Mn^{2+}) for its activity, showing a preference for Mg^{2+} [254]. One of the most distinctive features of this *P. falciparum* PGK is its high temperature sensitivity that has been attributed to a change in the proportion of basic amino acids in its structure [253]. Unlike the PGK of some trypanosomatids such as *T. brucei*, the *P. falciparum* PGK is activated in the presence of sulfate anions (as Na or K salts) [254].

Cryptosporidium parvum, an obligate intracellular parasite, can be a major cause of diarrhoea in human and mammals [255]. This parasite has a 9.4 Mb genome with 3807 protein coding genes [256]. Species of the genus *Cryptosporidium* have lost the apicoplast (and metabolic pathways associated with this organelle) as well as many of their mitochondrial metabolic capabilities, which causes that they depend even more on host cells to get some basic nutrients and rely heavily on glycolysis for their ATP generation [256–258]. However, this is not a general rule for all *Cryptosporidium* species. *Cryptosporidium marinus* and *C. andersoni* contain all enzymes of the Krebs cycle and a conventional respiratory chain [14,258].

Cryptosporidium possess genes for all 10 enzymes of glycolysis in the genome [259] and all enzymatic activities of this pathway, except for hexokinase, were detected in the cytosolic cell fraction [257]. PGK of this organism has been little studied. However, the databases (e.g. http://CryptoDB.org) list, in addition to a canonical PGK (CPATCC_0006650) with molecular weight 42 kDa and all key residues for activity, at the same chromosome 7, a very different 45 kDa PGK-family protein (CPATCC_0006130), which contains the Walker A and B motifs for ATP binding (table 3), but seems to lack any features specific for PGK activity. The metabolic function of this latter protein remains unknown. However, transcriptomic studies suggested that both enzymes may have a role in survival and resistance to environmental stress [260].

Studies of other apicomplexans of economic and clinical interest such as *Babesia*, the causative agent of babesiosis in many vertebrates including human and transmitted by ticks [261], have revealed important data about this protist. *Babesia bovis* has a genome with a size of 9.4 Mb that is organized in four chromosomes [262,263]. It is one of the smallest apicomplexan genomes found so far. Several genes associated with metabolic pathways (such as glycolysis, the Krebs cycle and oxidative phosphorylation) have been identified [264,265]. However, unlike in other protists, many pathways, such as fatty acid oxidation and de novo purine, haem and amino acid biosyntheses, are apparently absent [264]. In addition, proteomic analysis of erythrocytic stage cells of *B. microti* has revealed a significant reduction in its metabolic potential. During this life cycle stage, *B. bovis* and other species depend on glucose fermentation for their energy production and redox balance [117,265–267]. Some enzymes involved in the metabolism of *Babesia* have been studied [268]; however, so far there are only a few reports on its PGK. It is one of the glycolytic enzymes identified in *B. microti* through transcriptomic studies [265]. The enzyme is encoded by a gene located on chromosome 3 and has characteristics of a canonical PGK, with no other domains associated with its structure (table 3). Therefore, its function is most likely limited to generating ATP in the glycolytic pathway. Studies of *B. gibsoni* showed that diminazene aceturate (DA)-resistant isolates have an increased glucose consumption and intracellular ATP level compared to wild-type isolates [269]. Although the mechanism by which the glycolytic activity and ATP concentration are increased is still unknown, it would be interesting to analyse candidate enzymes that confer this characteristic to DA-resistant *B. gibsoni* isolates.

7.2.3. Stramenopile protists

Blastocystis belongs to the non-photosynthetic stramenopile group of protists containing also free-living unicellular algae and diatoms. But *Blastocystis* is an anaerobic parasite responsible for gastrointestinal infections in humans and a large variety of animals worldwide. Some studies have been conducted on the subcellular localization of its glycolytic enzymes, where both cytosolic and mitochondrial forms of some of these enzymes have been found [270]. When analysing the genome of this parasite, we found a single gene that codes for a PGK with a canonical structure, with a possible localization in its mitochondrial remnants (mitochondrial-related organelles, MROs) (table 3). These results concur with previous reports by Nakayama *et al.* [271] and Abrahamian *et al.* [272]. A bioinformatics analysis of 72 genomes covering 13 eukaryotic groups about the subcellular distribution of all glycolytic enzymes predicted that not only cytosolic but also mitochondrial forms of glycolytic enzymes are present in several representatives of the stramenopiles [271,272]. However, only the 'payoff-enzymes', i.e. enzymes of the second part of the pathway responsible for producing ATP and intermediates for lipid and amino acid biosynthesis, were located in both compartments of the stramenopiles [272]. In *Blastocystis* sp., this second half of the glycolytic pathway is even solely localized in its MROs [273].

The PGK of *Blastocystis* has a cyanobacterial origin [274] and is present in its MROs, along with the other enzymes (GADPH, PGAM, ENO, PYK) of the 'payoff phase' of glycolysis. In most stramenopiles, the compartmentalization of these glycolytic enzymes within mitochondria goes along with that of enzymes involved in the synthesis of serine (phosphoglycerate dehydrogenase, phosphoserine aminotransferase and phosphoserine phosphatase). This co-compartmentalization may facilitate the transfer of intermediates between both pathways, since they share a common intermediate, 3PGA. Whereas in most eukaryotes, the enzymes of serine biosynthesis, like those of glycolysis, are present in the cytosol, such cytosolic localization of the amino acid biosynthetic enzymes is missing from oomycetes and most other stramenopiles [272]. However, *Blastocystis* does not possess a serine biosynthesis pathway; its ancestors may have lost it with when adapting to an anaerobic parasitic lifestyle and reducing their mitochondria to MROs, while retaining the glycolytic 'payoff' enzymes in the organelles.

Most of the PGKs identified in the parasitic protists described above in this section have a canonical PGK structure. These enzymes have molecular weights ranging between 45 and 63 kDa. Most of these enzymes are present in the cytosol, with some exceptions having a localization in the mitochondria or MROs and the apicoplast. In the

case of these parasitic protists, these glycolytic enzymes seem to be much more often linked to functions beyond metabolic ones, such as de novo purine synthesis, virulence, protein homeostasis, functions that have not been observed for these enzymes in free-living protists.

8. Diplonemid and Kinetoplastid protists

8.1. Diplonemids

Diplonemids form a group of biflagellate, heterotrophic protists, most of them inhabiting marine ecosystems. These organisms, together with the euglenids (§7.1.2) and kinetoplastids, constitute the Euglenozoa phylum (figure 1). Diplonemids, which also include some pathogenic taxa [275], are considered metabolic versatile organisms. Genomic studies have shown that these organisms are characterized by having the most protein-encoding-rich transcriptomes when compared with euglenids and kinetoplastids. Of its 43 107 predicted proteins, approximately 2.55% correspond to metabolic enzymes, of which 246 are unique to diplonemids [276]. In D. papillatum, the presence and localization of glycolytic enzymes have been documented [118,277]. Several glycolytic enzymes of this organism, such as glucokinase, glucose-6-phosphate isomerase, TIM and PGK, have been associated with peroxisome-related microbodies called glycosomes, and their sequences contain a canonical type 1 peroxisomal-targeting signal (PTS1) at their C-terminus. Glycosomes (i.e. peroxisomes containing glycolytic enzymes) have previously been found in all Kinetoplastea studied (see next sections). The finding of these enzymes also in glycosomes of diplonemids, while not in E. gracilis, provides evidence for the notion that such compartmentalization originated in the common ancestor of diplonemids and kinetoplastids [118]. Whereas D. papillatum seemed to lack a fully functional glycolytic pathway (HK and PFK were not expressed), this protist is able to perform gluconeogenesis [118]. The gluconeogenic flux would be sustained by the use of amino acids as sources of energy and carbon. These are catabolized via the Krebs cycle, and the metabolites subsequently used for anabolic purposes in the gluconeogenic and pentose–phosphate pathways [118]. Retention of a full spectrum of genes encoding biosynthetic pathways allows Diplonema to obtain its amino acids both from nutritional sources and through biosynthesis, a property that could be an important factor for the ecological success of the diplonemids, since amino acids are the main nutritional component in the phagocytic diet of these heterotrophic organisms [118,276].

When analysing genes for glycolytic enzymes of D. papillatum, we obtained results similar to those previously reported. Some of these genes, such as for PFK and GAPDH, are duplicated [118]. Three ORFs were found for PGK, one of them (BAQ25443) coding for a fully functional PGK with a canonical PTS1 (a C-terminal tripeptide -SKL) (table 3). This PGK was located in glycosomes where it can perform the catalytic reaction in the gluconeogenic direction [118]. In addition, at least two ORFs that encode PGK-like proteins have been identified. These ORFs are located in different parts of the genome and encode proteins with molecular weights of 42 and 26 kDa. Both products appear to be catalytically 'dead' PGKs and lack a canonical PTS sequence (data not shown). The absence of activity of these PGKs may be due to redundancy associated with the adaptation to the heterotrophic lifestyle and development of the gluconeogenic state of D. papillatum.

8.2. Kinetoplastids

PGK is one of the most studied enzymes of kinetoplastids. Not only because it is one of the enzymes responsible for generating ATP in the glycolytic pathway that for its major part is compartmentalized within glycosomes, but also because multiple genes and PGK isoenzymes have been detected in these organisms [24,25,122,278]. Additionally, in many of the Kinetoplastea species studied, most of the genes for these isoenzymes are organized in tandem [23,80,279]. However, these duplicated, tandemly arranged genes show variations in their sequences. Such tandem arrays have the capacity to harbour important variation for enzyme function and subcellular localization [280]. Variations may occur in the subcellular localization of the PGKs in different species and in their expression in the successive developmental stages of these kinetoplastids [60,119,125,279,281]. The characteristics of these PGK genes and their products are summarized in table 3.

8.2.1. Free-living kinetoplastids

8.2.1.1. Bodo saltans
Bodo saltans is a free-living kinetoplastid that is evolutionarily related to trypanosomatid parasites. A comparative study of the genomes of B. saltans and parasitic trypanosomatids revealed that the transition from a free-living to a parasitic lifestyle has resulted in the loss of approximately 50% of protein-coding genes [282]. This protist has genes for several PGKs, like the parasitic kinetoplastids [24,25,60,122,278]. However, its three pgk genes are present in three separate loci [282]. Of these three genes, one (SAL_14730) codes for a potentially active enzyme of 56 kDa that is predicted to be present in glycosomes. The other two ORFs code for isoenzymes that both exhibit an unusual modular architecture, being multidomain PGKs in which, respectively, two and three different domains are combined (table 3). The gene SAL_29930 codes for a 57 kDa PGK with, at its N-terminal side, an extra PAS domain, similar to the enzyme that previously has been reported for T. cruzi and Leishmania major (see also sections below). In T. cruzi, a function for sensing environmental conditions has been attributed to this PAS-PGK protein, while in Leishmania, it is associated with adaptation to pH change, virulence and cell survival through autophagy [4,28]. The PAS-PGK seems to have different subcellular localizations in B. saltans and the two parasitic kinetoplastids; in B. saltans, bioinformatics analyses predict PAS-PGK to be a cytosolic protein, while in T. cruzi, it is associated with glycosomes, and in Leismania, it appears to have a dual localization, in glycosomes and the lysosome [4,28,104]. In turn, gene BSAL_70390 encodes an 89 kDa PGK-like protein with a cAMP-binding domain (CNB) and a transmembrane domain. It seems orthologous with the PGKL enzyme reported for T. brucei [60], as discussed in §6.5 and below in §8.2.3.2.1, although it lacks the seemingly relict DNA-binding domain present in the protein of the latter and some other trypanosomatid species. In both protists, the protein has as distinctive

feature that the PGK domain lacks key catalytic residues; in *B. saltans*, the PGK domain has specifically lost residues necessary for 3PGA binding. In addition, in *B. saltans*, this PGK is predicted to be targeted to a membrane (table 3), while in *T. brucei*, this enzyme has been located in the flagellum [60], along with some other glycolytic enzymes, such as a GAPDH isoform. The finding of these similar multidomain PGKs in bodonids and trypanosomatids indicates that their presence pre-dates the divergence of the parasitic trypanosomatid family. Although the presence of orthologues of PGK with a CNB and one or more membrane-spanning domains is not limited to *T. brucei* and *B. saltans*, it seems to be absent in several other kinetoplastids such as *Perkinsela* sp. and *Leishmania* spp. Moreover, its subcellular location seems to have undergone changes during the divergence of these parasites (table 3; electronic supplementary material, table SI). Although a possible explanation previously proposed for the presence of this flagellar degenerated PGK enzyme in *T. brucei* is a change to glycolysis regulation and compartmentalization during kinetoplastid evolution [60], we suggest that it may rather be related with a protein that acts as a 'functionally flexible' enzyme, fulfilling different cellular tasks not related to the proper PGK function, but adapted to the environmental and nutritional requirements of the different groups of kinetoplastids. This would be the reason why it is conserved in most kinetoplastids, although it is catalytically not functional. This type of enzyme can fulfil a wide variety of functions, without depending on his catalytic properties [283–285].

8.2.2 Endosymbiotic kinetoplastids

8.2.2.1. *Perkinsela* sp.

Perkinsela sp. is a kinetoplastid living as an endosymbiont within the amoebozoan *Paramoeba pemaquidensis* that infects marine animals. Its nuclear genome is approximately 9.5 Mb [286], much smaller than that of *B. saltans* [5], *Phytomonas* sp. [287], *Leptomonas pyrrhocoris* [288] and human parasites such as *T. brucei* [289], *Leishmania* spp. [290] and *T. cruzi* [291]. *Perkinsela* sp. is a sister group of *Ichthyobodo*, both of which make up the group of the Prokinetoplastea within the Euglenozoa phylum (figures 1 and 2). They represent an early branching lineage within the Kinetoplastea. By the robust branching of the clade base, the stable intracellularity of *Perkinsela* can be considered as an ancestral form of parasitism [292]. Although genomic analyses showed that *Perkinsela* sp. has lost the ability to develop a flagellum, as do other kinetoplastids, it retains hallmark features of kinetoplastid biology, included glycosome-like organelles. While the repertoire of its genes and proteome is markedly reduced, *Perkinsela* sp. possesses kinetoplastid-specific biochemical pathways [286]. With regard to the presence of glycosome-like organelles, various genes coding for PEX proteins involved in glycosome biogenesis, including both membrane proteins and cytosolic ones (such as the receptors for import of matrix proteins with a type 1 PTS (at the C-terminus) or type 2 (close to the N-terminus) PTS), were identified in the genome of this endosymbiont. Additionally, putative PTS1 or -2 motifs were identified in the first seven enzymes of the glycolytic pathway encoded in its genome [286]. Like in other kinetoplastids, in addition to glycolysis, other biochemical processes are predicted to occur in the putative glycosomes of *Perkinsela* sp., which include amino acid, nucleotide and sterol/isoprenoid

metabolism. PGK is one of the glycolytic enzymes identified in *Perkinsela* sp. Apparently, this organism has only one gene coding for a PGK [286], which could have two explanations. First, PGK isoenzymes present in the other kinetoplastids were lost in *Perkinsela* sp. as a result of its adaptation to an endosymbiotic lifestyle. Alternatively, *Perkinsela* represents the ancestral situation and the multiple isoenzymes were acquired by free-living kinetoplastids and parasites with complex life cycles. The *Perkinsela* enzyme has a canonical structure of 58 kDa, with only a PGK domain associated with its structure; it has all residues required for the catalysis and contains a PTS1 (-GKL).

Perkinsela sp. also lacks the PGK isoforms with possible capacity of sensing environmental conditions domains, such as the PAS-PGK (table 3). Intriguingly, this PGK isoenzyme with the domain related to the sensing of certain environmental conditions is also absent in *T. brucei*. Although this parasite alternates during his life cycle between two hosts, it is permanently extracellular, while other kinetoplastids, such as *Leishmania* spp. and *T. cruzi*, not only have an alternation of host but also, during certain stages live intracellularly.

8.2.3. Parasitic kinetoplastids

8.2.3.1. *Paratrypanosoma confusum*

Paratrypanosoma confusum is a monoxenous kinetoplastid isolated from the gut of *Culex pipiens* mosquitoes [293]. Evolutionarily, it represents the most basal trypanosomatid lineage branching between the free-living *B. saltans* and the parasitic trypanosomatids [293,294], and may thus provide insight into the emergence of parasitism. Preliminary studies of its genome showed that it shares at least 114 genes with 15 species of trypanosomatids, *B. saltans* and the heterolobosean *N. gruberi* and 129 protein genes with the endosymbiotic *Perkinsela* sp. [293]. *Paratrypanosoma* and stercorarian trypanosomes, including *T. cruzi* and *Trypanosoma grayi*, retain more ancestral genes than other trypanosomatid clades [294].

When analysing the *P. confusum* genome, we found two genes (PCON_0085140 and PCON_0003620) encoding PGK-like proteins, with sizes of 28 and 104 kDa, respectively (table 3). Both PGKs are apparently catalytically inactive, due to substitutions and deletions of residues essential for substrate binding. The predicted 104 kDa protein has the same domain organization as the 89 kDa PGKL of *B. saltans*, although the overall amino acid sequence identity is only about 39%. Surprisingly, no gene for a conventional PGK was detected. Until now, there have been no metabolic studies of *P. confusum*. However, it is very likely that its network for intermediary metabolism is organized very similarly as that in other kinetoplastids, with sequestering of a major part of the glycolytic pathway within glycosomes, because homologous genes encoding enzymes like HK, ALD and GAPDH, each with a PTS, are present in the genome. The absence of a functional PGK is most likely to be due to incompleteness of the genome data. Less probable alternatives are that this parasite contains a very different PGK not recognized in our bioinformatics analysis or has some mechanism to bypass the catalytic passage of this enzyme. An active PGK or bypass for it should be necessary, because genes for seemingly functional downstream glycolytic enzymes (PGAM, ENO, PYK) are present in the *P. confusum* genome.
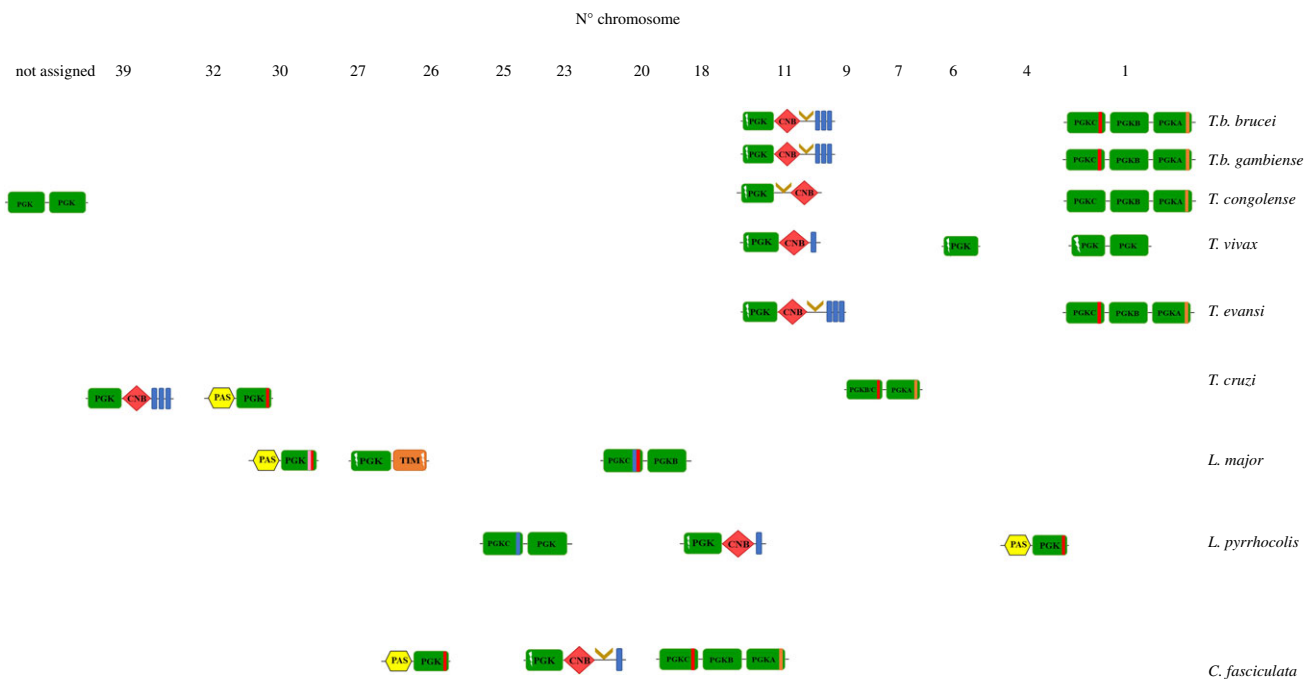
Figure 5. Genomic location of *pgk* genes in different species of Kinetoplastea. Determination of the chromosomal localization of identified *pgk* genes was performed using the Kinetoplastid Genomic Resource (TriTrypDB) [94]. Identification of domains and motifs present in PGK sequences was done using the following bioinformatics servers: *PAS domain*: Protein Blast (NCBI) [96], InterProScan (EMBL-EBI) [95,97] and Simple Modular Architecture Research Tool (SMART) [98]; *CNB domain*: NCBI, EMBL-EBI, SMART; *HTH (Helix–Turn–Helix)* motif: ExPaSy [99], GYM [100], iDNA-Prot [101]; *Membrane helix*: SMART [98]; Predicting transmembrane protein topology with a hidden Markov model (TMHMM prediction) [102] and Phobius [103]. Glycosomal localization prediction was based on recognition of a PTS1 consensus sequence as reported by Acosta *et al.* [104]. *Domain symbology:* PGK domain, PGK domain having lost substrate (3PGA)-binding residues, triosephosphate isomerase domain, HTH domain, transmembrane segment, PAS domain, cyclic-nucleotide binding-domain. Import sequences and inserts identified in PGK isoenzymes through experimental and theoretical studies: putative PTS1, putative lysosome, PGKA-like insert sequence.

A possible bypass that could be considered is the Rapoport–Luebering pathway, which is best known from its functioning in erythrocytes, and has also been documented in protists like *D. discoideum* [295]. This pathway comprises evolutionarily conserved reactions that generate and dephosphorylate 2,3-bisphosphoglycerate (2,3BPGA), which in turn is transformed to 3PGA. These two reactions are catalysed by the 2,3BPGA synthase/2-phosphatase (BPGM) enzymes and the multiple inositol polyphosphate phosphatase (MIIP1), respectively [295]. For *Dictyostelium*, it has been demonstrated that these enzymes provide a physiologically relevant regulation of cellular 2,3BPGA content. They also establish a molecular link between the turnover of phosphorylated inositol derivatives and glycolytic flux [296]. However, genomic searches did not reveal indications for its existence in this organism or any other kinetoplastid.

8.2.3.2. PGK in African trypanosomes

8.2.3.2.1. *Trypanosoma brucei*. *T. brucei* has three *PGK* genes (A, B and C) organized in tandem on chromosome 1 (table 3 and figure 5). The *pgkA* gene encodes a variant (PGKA) that is constitutively present at a low level within the glycosomes of both the mammalian bloodstream and insect procyclic forms of this parasite [120,297]. The second gene, *pgkB*, encodes the cytosolic enzyme PGKB, found at considerable concentration but only in procyclic forms, and the third gene, *pgkC*, encodes the major glycosomal enzyme PGKC, abundantly present only in bloodstream forms [23,119]. PGKC is directed to the glycosomes by a PTS1-like sequence present at the end of a 20-amino-acid

C-terminal extension. The gene cluster *pgkA–pgkB–pgkC*, located on chromosome 1, is part of a long multicistronic transcription unit. However, the different concentrations of the individual mRNAs during the life cycle are controlled post-transcriptionally [23,80].

The PGKB and PGKC gene products are monomeric proteins with molecular weights of 45 and 47 kDa, respectively. Their amino acid sequences are 95% identical, and they share approximately 45% identity with other, non-kinetoplastid eukaryotic PGKs [119]. Unlike PGKs from yeast and rabbit muscle, which can also use GTP and ITP, the *T. brucei* enzymes are strictly ATP dependent [22]. This specificity for ATP as substrate is related to substitution of amino acids in the structure of these isoenzymes, what probably prevents phosphoryl-group transfer during catalysis in the case of the other nucleotides. PGKA is monomeric protein with around 60% identity to the PGKB and C isoforms, and has the distinctive feature of an 80 amino acids long insertion in its N-terminal half. Part of the amino acid sequence (position 24–91) of this insertion has been shown necessary for glycosomal targeting of this isoenzyme, although so far, no specific sequence motif for it has been identified. It has been suggested that the glycosomal entry of PGKA is by virtue of its interaction with a protein that does contain an import signal [297]. This enzyme is not only less abundant than the PGKB and PGKC isoenzymes but has also negligent catalytic activity [121]. All these PGKA/B/C isoenzymes of *T. brucei* have very similar substrate affinities (table 2). The Km for ATP varies between 0.12 and 0.4 mM, while for 3PGA, it ranges between 1.62 and 2.40 mM.

The PGKs were the first glycolytic enzymes of *T. brucei* for which it was shown that their specific compartmentalization in glycosomes and cytosol has a key role in the proper execution of the parasite's energy metabolism and thus is essential for viability [298].

Whereas previous studies of PGK in *T. brucei* focused on the PGKA/B/C isoenzymes, recent genome analysis of this parasite by us and others revealed that it has an additional gene for a candidate PGK isoenzyme. This gene (Tb927.11.2380), located on chromosome 11, codes for a 101 kDa multidomain PGK-like protein (figure 5), which is constitutively expressed in the procyclic and bloodstream stages of the parasite and has a location in the flagellum, specifically in the axoneme. This flagellar PGK, designated PGKL, has an unusual structure with its PGK domain at the N-terminal side, followed by a cyclic nucleotide monophosphate-binding (CNB) homologous domain and a region matching a helix–turn–helix (HTH) domain of DNA-binding proteins (table 3). A distinctive feature of this PGKL is the apparent lack of catalytic PGK activity. Residues highly conserved in other PGKs and known to be involved in catalysis are absent, and its region homologous to DNA-binding proteins was predicted to have no longer DNA-binding ability [60]. Nonetheless, as will be discussed below (and is shown in table 3; electronic supplementary material, table SI), related PGKLs with CNB and HTH-like domains are also found in many other trypanosomatid species suggestive of a functional role. The combination of CNB and HTH domains is very common in transcription factors of Bacteria and Archaea, and has also been documented in eukaryotic proteins [299]. The HTH domain is frequently combined with a diverse set of catalytic domains, and several general functions are associated with such combinations. In proteins with also a catalytic module, the HTH domain functions as a substrate-recognition or localization domain and for recruitment of the enzyme to its site of action [299]. Sometimes, the combined HTH and enzymatic domains are used for feedback regulation of metabolic pathways, for example, when an HTH domain is linked to a domain of an enzyme catalysing a key step in a biosynthetic pathway to regulate the transcription of the genes for that pathway. This is what happens with the *E. coli* biotin operon repressor, BirA, which contains an N-terminal HTH domain fused to a C-terminal biotin ligase domain. In the presence of biotin, the enzymatic domain synthesizes the co-repressor, and the HTH domain represses the transcription of the biotin biosynthesis genes [300]. Other studies with this PGKL have shown that this isoenzyme may be necessary for 'social motility' during the *T. brucei* procyclic stage [301]. This term refers to group behaviour, in which parasites assemble into multicellular communities with emergent properties that are not evident in single cells [301]. Oberholzer *et al.* [302] and Lopez *et al.* [303] discovered that procyclic *T. brucei*, when cultured on semi-solid agarose, assemble into groups of cells that undergo collective movements, forming multicellular projections that radiate outward from the site of inoculation. Although individual parasites can move freely within a group, movement of the group is polarized such that it advances at a single, leading edge. Moreover, groups alter their movements when they sense other parasites nearby. This social motility requires sensing and responding to extracellular signals [302]. It may be regulated by a cAMP-dependent mechanism, in which cNMP-binding proteins (cNMP-BPs) are possibly involved [301,304,305]. PGKL is a candidate among the cNMP-BPs involved. Parasites in which PGKL gene expression was knocked down showed a slightly greater social motility than wild-type cells [301], although null mutants did not show morphological or growth phenotypes (see also §6.5) [60].

In *T. brucei,* social motility is probably regulated and depends on cAMP signalling systems in the parasite's flagellum. This flagellar cAMP signalling may control trypanosome progression through fly host tissues. cAMP-specific phosphodiesterases, localized in the flagellum and associated with social motility, are essential for virulence of bloodstream-form trypanosomes [306]. The social behaviour in this parasite seems to be important for differentiation and for the parasite to navigate through diverse environments encountered during their transmission and infection cycle [307,308].

8.2.3.2.2. *Trypanosoma brucei gambiense.* So far, PGK enzymes of *T. b. gambiense* have not been studied. However, in a preliminary analysis of the genome of strain *DAL972* of this subspecies that was originally isolated from a patient in West Africa, we identified genes for four different PGK isoenzymes (table 3 and figure 5). These enzymes have PGK-like characteristics very similar to those described above for the homologues in the non-human-infective subspecies *T. brucei brucei 427* (the strain for which most of the biochemical and molecular biological studies related to PGK isoenzymes have been performed) [22,23,80,119]. Besides genes for the isoenzymes PGKA, B and C, each 99–100% identical to its orthologue in *T. brucei*, there is also one coding for a protein identical to *T. brucei*'s flagellar PGKL, described by Brown *et al.* [60], but the expression and localization of this latter protein in *T. b. gambiense* remain to be established (electronic supplementary material, table SI; figure 5).

8.2.3.2.3. *Trypanosoma congolense.* In the genome of *T. congolense*, responsible for African animal trypanosomiasis, six genes encoding PGK isoenzymes have been identified (table 3 and figure 5). Of these six *pgk* genes, three (TcIL3000_1_220.1; TcIL3000_1_230; TcIL3000_1_240) were already reported by Parker *et al.* [122]. These genes have a genomic organization in tandem similar to that in *T. brucei* and code for two identical cytosolic proteins, C1PGK and C2PGKC (45 kDa), and a glycosomal membrane-associated PGK (56 kDa), which are each 83% identical to *T. brucei* PGKB/C and A, respectively. These isoenzymes are constitutively expressed during the life cycle of *T. congolense*; however, C1PGK is most abundant in procyclic forms while C2PGK is more abundant in the highly glycolytic bloodstream and metacyclic forms. This expression pattern of *T. congolense* PGKs is reminiscent of the expression of the PGKA/PGKB/PGKC isoenzymes of *T. brucei*. With regard to the PGK activity, more than 90% was detected in the cytosol, corresponding to both C1PGK and C2PGK. Although the 56 kDa PGK is co-expressed with the other two isoenzymes, its activity could be shown to be associated with the glycosomes due to its larger molecular weight [122].

Metabolic differences between *T. brucei* and *T. congolense* have not only been observed in the expression of their PGK isoenzymes, but also in those of many other glycolytic enzymes. Comparative transcriptomic studies of *T. brucei* and *T. congolense* showed differences for components of several processes between both parasites. The expression of most of the enzymes involved in glycolysis was negatively regulated in the so-called stumpy form of *T. brucei*, the non-proliferative but tsetse fly-infective forms that

accumulate in the blood at increasing parasite density, meanwhile no change between ascending and peak parasitaemia was seen for *T. congolense* [309].

Each of the isoenzymes has apparently a structure comprising only the PGK domain, except the 66 kDa protein encoded by TcIL3000.11.2180.1. This latter protein seems related to the 101 kDa PGKL of *T. brucei*. It has an incomplete PGK domain, lacking the 1,3BPGA binding site, but the residues involved in nucleotide binding and those that are part of the flexible hinge are well conserved. Like in *T. brucei* PGKL, the PGK domain is followed by CNB and HTH domains, but it differs from it (and the PGKL of some other *Trypanosoma* species—see below) by inversion of the order of these domains, suggesting that the *T. congolense* gene has undergone a recombination that changed its architecture. Moreover, the *T. congolense* protein lacks any transmembrane domains at the C-terminus (table 3). Genes for enzymes that participate in the glycolytic pathway are differentially expressed in *T. congolense*, being upregulated in parasites in the fly's proboscis. This is also the case for TcIL3000.11.2180.1. Differential expression of enzymes such as HK could be indicative of pre-adaptation of these parasites for the glucose-rich bloodstream environment in their mammalian host [310]; however, the significance of this expression pattern of TcIL3000.11.2180.1 is unclear, because this PGK isoenzyme is a catalytically inactive enzyme, or a 'dead enzyme'. A group of enzymes encoded in the genomes of other kinetoplastids such as *L. major*, *T. cruzi* and *T. rangeli* have also been identified as 'dead' enzymes [3,311]. In metazoan organisms, such 'dead' enzymes were found to represent a source of biological regulators [312,313]. For example, kinases that have lost catalytic activity can exhibit biological activities via regulation of protein–protein interactions or allosteric modulation of the activity of enzymes. Additionally, a 'dead' enzyme can be involved in influencing the availability of substrate for other active enzymes, immobilization of the substrate to a particular subcellular space, or protein–DNA interactions [314,315]. The presence of an HTH domain, as in the protein encoded by TcIL3000.11.2180.1, points to some function in the regulation of gene expression. Proteins with an HTH type domain have been documented in many eukaryotes and bacteria. They participate in ribonucleoprotein complexes in the cytoplasm where they specifically recognize various classes of double-stranded RNAs [316].

Another distinctive feature in *T. congolense* is that 4 (TcIL3000_1_220.1; TcIL3000_1_230; TcIL3000_0_37920; TcIL3000_0_37930) of the 6 PGK genes appear to code for identical protein sequences. Two of these genes (TcIL3000_1_220.1; TcIL3000_1_230) are located on chromosome 1, while the chromosomal localization of the other two genes (TcIL3000_0_37920; TcIL3000_0_37930) is still unknown (figure 5). These genes must have resulted from recent duplication events in contrast with the other two genes (TcIL3000_1_240 and TcIL3000.11.2180.1). It is possible that natural selection has been acting on the retention of these gene duplicates, similar to that has been proposed for other eukaryotic organisms with many gene copies of enzymes that catalyse the same reaction [317].

8.2.3.2.4. *Trypanosoma evansi*. *T. evansi* is another parasite that causes animal trypanosomiasis, also outside Africa because its transmission is not dependent on tsetse flies, which are endogenous only to sub-Saharan Africa, but occurs mechanically. Studies of this parasite's metabolism involved glycolysis, including its enzymes like PGK. Rivero

et al. [26] have shown that this parasite (strain TEVA 1), residing in the blood, has a more complex metabolism than has been postulated for bloodstream-form *T. brucei*. They showed that PGK activity occurs mainly within glycosomes, but it was not the only source for ATP synthesis within the organelles. Additional production of glycosomal ATP, necessary for other processess such as biosynthetic pathways compartmentalized in the glycosomes, was attributed to phosphoenolpyruvate carboxykinase (PEPCK) and pyruvate phosphate dikase (PPDK) [26]. Although there are so far no reports about the characterization of PGK isoenzymes in *T. evansi*, analysis of its genome showed that it has four *pgk* genes (table 3). Three genes code for the distinct PGKA/B/C isoenzymes (TevSTIB805.1.640; TevSTIB805.1.650; TevSTIB805.1.660, respectively), as in *T. brucei*. Two of them (TevSTIB805.1.640 and TevSTIB805.1.650) encode proteins that are 95% identical in sequence among them and to PGK isoenzymes B and C of *T. brucei*. This is in agreement with the notion that *T. evansi* evolved (independently, multiple times) from *T. brucei* after loss of critical parts of its mitochondrial DNA and can be considered as a subspecies of *T. brucei*; the two species exhibit overall a high identity of their genomes (approx. 95%) [318]. Through MS-based proteomics studies, more than 100 gene products, expressed by the parasite during infection in a rodent model, were identified [319]. Metabolic enzymes constitute approximately 23% of the identified proteins. Within that group of enzymes, the glycolytic enzymes were the most conspicuous, indicating their abundance and confirming the dependence of the parasite on this pathway as the main source of energy. Among the metabolic enzymes identified was PGK, along with other glycolytic enzymes (such as HK, GAPDH, PGAM, ALD, glucose-6-phosphate isomerase) as well as glycerol kinase. When performing an alignment of the 12 peptides obtained by the MS analysis from the isolated PGK protein [319] with the amino acid sequences predicted from the different *T. evansi* PGK genes we identified, 7/12 peptides appeared to match perfectly with the 45 kDa product of the TevSTIB805.1.650 gene (PGKB) and 12/12 peptides with the 47 kDa product (PGKC) encoded by the TevSTIB805.1.640 gene. In the case of the 56 kDa product (PGKA), only 3/12 peptides completely aligned with its sequence (these peptides also aligned with the sequences of the 45 and 47 kDa isoenzymes). However, none of the peptides aligned perfectly with the sequence of a 101 kDa PGK-like protein, encoded by the fourth gene that we also identified in our bioinformatics study and which is 100% identical to the *T. brucei* PGKL containing the CNB domain. Based on these analyses, we conclude that it is very likely that at least PGK isoenzymes of 47 and 45 kDa are expressed in the bloodstream form of *T. evansi*.

8.2.3.2.5. *Trypanosoma vivax*. *T. vivax* is another member of the African trypanosomes responsible for animal trypanosomiasis. However, it also occurs outside Africa because the parasite can be transmitted by different biting insects. *Trypanosoma vivax'* life cycle does not involve a procyclic stage in tsetse flies. When its bloodstream forms are ingested by an insect, they move directly to the proboscis, while differentiating via epimastigotes into metacyclic forms that can be introduced in new mammalian hosts. Only a few reports are available about PGK isoenzymes of *T. vivax*. Gene expression profiling through the complete, short life cycle of this parasite found that transcripts preferentially expressed in bloodstream forms

are associated with glycolysis and glycosomes, indicating substrate-level phosphorylation as the dominant process for ATP generation in this stage, a consistent feature of bloodstream-form African trypanosomes. Within these enzymes that undergo a difference in expression and that are remarkably abundant in the bloodstream form are two PGK isoenzymes of 45 and 41 kDa (TvY486_0100140 and TvY486_0604100) [320] and table 3, 82/83% and 46% identical to *T. brucei* PGK isoenzymes B and C, respectively. However, the 41 kDa protein has undergone substitutions in its 3PGA binding site and is probably not active. Similar to the different subspecies of *T. brucei* (including *T. evansi*), an orthologue of PGKL was also identified in *T. vivax*, having 66% identity with the *T. brucei* protein, although with the lower molecular weight of 86 kDa due to the absence of an HTH-like domain (as also described above for the *B. saltans* and *P. confusum* PGKL) and shortening of membrane-spanning region at the C-terminal side. Additionally, a gene (TvY486_0100150) coding for an unusually small 32 kDa PGK is present in the genome of this protist. Despite its size, this enzyme has an apparently functional active site. An aspect that caught attention when analysing the genes for PGK isoenzymes was that the amino acid sequence of this 32 kDa protein contains a mitochondrial-type N-terminal targeting sequence. This mitochondrial PGK may have a non-canonical function in the organelle, for example, comparable with PGK's function as a PK as has been reported for mitochondria in some tumour cells [183]. Similar findings in other kinetoplastids will be described later in this review.

The analysis carried out in this work showed that *pgk* genes in at least some African trypanosomes are distributed in more than two loci. In these parasites, these genes are concentrated mainly on chromosomes 1, 6 and 11 (figure 3). Some of these parasites, such as *T. congolense*, have up to six candidate PGK genes. These genes code for a wide diversity of proteins. Some have a canonical PGK structure, while others are multi-domain, 'dead' PGKs or truncated PGKs (table 3 and figure 5). One of the most notorious aspects of these *pgk* genes is that they can be variable in number and structure between different species of African trypanosomes. Nevertheless, some of these PGK isoenzymes are highly conserved among different species.

8.2.3.3. Stercorarian and reptilian trypanosomes

8.2.3.3.1. *Trypanosoma cruzi*. *T. cruzi* is one of the kinetoplastids of which PGK isoenzymes have been characterized in detail [4,24,25,124]. Until now, three genes encoding PGKs in *T. cruzi* have been reported. Two, *pgkA* and *pgkB/C*, are located in tandem on chromosome 9, while the third one, the *pas–pgk* gene, is found on chromosome 32 (table 3 and figure 5) [4,24]. However, as will be described in the next sections, bioinformatics analysis of its genome revealed that this parasite has at least 6 *pgk* genes, some of them encoding proteins with very peculiar characteristics (table 3).

Most studies of *T. cruzi* have focused on PGKA, B and C. Similar to the situation in *T. brucei*, these PGK isozymes of *T. cruzi* have a location within glycosomes (PGKA, 56 kDa/PGKC, 47 kDa) and in the cytosol (PGKB, 45 kDa). These three isoenzymes are expressed simultaneously in the different life cycle stages of this parasite [24]. However, PGKC and PGKB expression is increased when glucose in the culture medium is limiting and amino acid consumption increases. This differential expression of these PGK isoenzymes is highly indicative of their participation in the change of energy and carbon metabolism in the parasite, specifically in a transition from glycolysis to gluconeogenesis [124].

The PGKA isoenzyme is characterized by having an insertion of approximately 80 amino acids in its N-terminal half and is tightly associated with the inner surface of the glycosomal membrane, although so far no region responsible for this association has been identified. The protein's amino acid sequence has 75%, 77% and 59% identity with the PGKAs from *T. brucei*, *T. congolense* and *Crithidia fasciculata*, respectively [24]. However, the 80 residues long insertion, specific for all PGKAs, is less conserved. The insertion of the *T. cruzi* enzyme shows, respectively, 57%, 54% and 52% identity with the corresponding peptide of the three other trypanosomatid PGKAs. An additional striking feature of the PGKAs of the different trypanosomatids is the calculated high isoelectric point of the insertion (between 9.6 and 10.5) [24]. Until now, little is known about the possible function of this insertion within the protein's structure. Kinetic studies demonstrated that its deletion has no effect on the catalytic activity of the *T. cruzi* PGKA. It is thought that the insertion could have some function in the formation of multienzyme complexes [25]. Although the nature of PGKA's intimate association with the glycosomal membrane has not yet been determined, results from different experiments indicated that this isoenzyme plays an important role in maintaining the glycolytic flux in *T. cruzi* [124].

Trypanosoma cruzi PGKB and C are essentially identical and share about 80% identity with their *T. brucei* orthologues [24]. The initial finding of only two PGK genes (one of them for PGKA) and the fact that PGKB and PGKC could barely be distinguished by western blotting, except for their different subcellular localization, have led to the following hypothesis [24]. First, that the PGK B and C isoenzymes are encoded by the same gene (*pgkb*: TcCLB.511419.40), and second, that the product of this single gene has a dual location, in the cytosol and in the glycosomes [124]. The protein predicted from the *pgkb* gene does not have an apparent PTS sequence. A routing signal might be encrypted internally in the sequence of the protein and would be responsible for the partial compartmentalization of the enzyme within the glycosomes. Additionally, the slight differences between PGKB and PGKC as detected on western blots may be due to covalent posttranslational modification of the protein, that in turn then also contributes to the regulation of the PGKB and PGKC levels in their respective compartments in response to the available carbon source. Such PTM may also influence the catalytic activity of the PGKC isoenzyme, as will be discussed below [25].

The activity of PGKB accounts for approximately 80% of the total activity and is cytosolic, whereas PGKA and PGKC are together responsible for an estimated 20% of the total cellular activity [24]. PGKA is characterized by low catalytic activity; the 20% glycosomal PGK activity is for 23% attributed to this enzyme, while 77% corresponds to PGKC activity. Kinetic studies of recombinant PGKA yielded Km values for 3PGA and ATP of 850 and 236 µM, respectively [25,124]. On the other hand, the PGKB and PGKC isoenzymes, despite having very similar sequences and molecular weight, have very different affinities for both substrates. PGKB's Km values for 3PGA and ATP are 625 µM and 99 µM, respectively (W. Q. *et al.* 2006, unpublished results), with the Km value for ATP being 2.4 times lower than that reported for the

recombinant PGKA of this parasite. The PGKC is partially inhibited by ATP at concentrations higher than 130 μM, showing a Ki value (ATP) of 270 μM. The Km for ATP was 10 μM, an order of magnitude below that of PGKA and PGB. However, the Km of PGKC for 3PGA was 192 μM, similar to that reported for isoenzymes PGKA and PGKB. This was the first report of a PGK enzyme inhibited by ATP [25]. From these studies, we have tried to understand the possible function of each of these isoenzymes and the physiological significance of their simultaneous expression during all developmental stages of *T. cruzi* studied so far. Data suggest that PGKA is responsible for the glycolytic reaction of converting 1,3BPGA to 3PGA, with concomitant formation of ATP, within the glycosomes. On the other hand, the inhibition of PGKC by ATP and its high affinity for this substrate would suggest that this enzyme works in the gluconeogenic direction, supporting this anabolic pathway through glycosomes even when the ATP/ADP ratio within the organelles is very low. Finally, the cytosolic PGKB may rather exert a glycolytic function [25,124].

Recently, we identified and characterized also the product of a gene, TcCLB.506945.20, that codes for a PAS–PGK in the *T. cruzi* strain CL-Brener Esmeraldo like. A PAS–PGK protein is also present in the genome of all other *T. cruzi* strains analysed (electronic supplementary material, table SI), and other kinetoplastids (table 3), except African trypanosomes [4,321]. The *T. cruzi pas–pgk* gene is on a different chromosome than the *pgkA–pgkB/C* gene array (figure 5). The encoded protein has a molecular weight of 58 kDa, exhibits 44–45% identity to PGKA and PGKB and has all residues involved in PGK substrates binding and catalysis (W. Q. *et al.* 2018, unpublished results). Structurally, this PAS–PGK is a multi-domain protein with, in addition to a PGK domain, a PAS domain at its N-terminal end [4]. Like the other PGK isoenzymes of *T. cruzi*, this PAS–PGK is expressed throughout the life cycle of the parasite [94].

The PAS–PGKs of *T. cruzi* and other trypanosomatids possess a PTS1, in the case of the *T. cruzi* PAS–PGK the C-terminal sequence -PRL [4]. Indeed, proteomic analysis of purified *T. cruzi* glycosomes has shown that this protein is present in glycosomes where it appears to be peripherally associated with the membrane. Like the other PGK isoenzymes, PAS–PGK is expressed during all phases of epimastigotes (life cycle stage in the triatomine insect vector) growth in culture; however, it is more abundant in the proteome of glycosomes of epimastigotes harvested in the stationary growth phase than in exponentially growing cells [104]. PAS domains are known to have a site with the ability to bind a wide diversity of ligands (FMN, FAD, haem, chromophores and fatty acids) but they can also function in signalling without binding a small molecule, through the dimerisation of PAS domains. The PAS domain of *T. cruzi* PAS–PGK contains amino acids predicted to be involved in the formation of a similar ligand-binding site as the PAS domain of other proteins from different organisms [4]. It is noteworthy that the *pas–pgk* gene in the genome of the *T. cruzi* CCC strain has apparently undergone a duplication. The products of both sibling genes are predicted to contain a catalytically active PGK domain (electronic supplementary material, table SI). Duplication of PAS–PGK genes has not been detected in any other *T. cruzi* strains or kinetoplastid species we analysed. The absence of this PAS–PGK in African trypanosomes (such as *T. brucei*), reported by us in previous studies [4], was also reported by Durrani *et al.* [321].

In *T. cruzi*, three *pgk*-related genes were identified in addition to those already previously documented (table 3). One of these genes, TcCLB.504153.20, codes for a 95 kDa PGK similar to the PGKL reported in *T. brucei*, a multidomain protein with PGK and CNB domains, but like *T. vivax* and *B. saltans* lacking a separate HTH domain. Target prediction studies suggest that this protein may be located at the glycosomal membrane (result not shown). Moreover, it has a pI value around 9, even higher than the PGKL of *T. brucei*. A high pI is characteristic of many glycosomal proteins [322]. Additionally, a gene (TcCLB.506127.9) encoding a small-sized (13 kDa) protein was found in the *T. cruzi* genome that was identical to the N-terminal part of the PGKL, indicating that it must have resulted from a partial gene duplication, or a gene duplication and subsequent truncation. Finally, the TcCLB.506125.30 gene codes for a fragment of a possible PGK protein. All residues involved in catalysis were completely lost; the remaining amino acids cover three segments of a transmembrane domain, showing 93% identity with the C-terminal end of the PGKL, thus also pointing to a partial duplication of the TcCLB.504153.20 gene. These last two 'pgk-like' genes probably do not code for functional proteins but are evolutionarily recent genomic junk.

The analysis of the *T. cruzi* genome not only revealed the existence of PGK isoenzymes and ORFs for potential PGK-related proteins additional to those already previously described for this parasite, but also that the number and nature of these possible PGK isoenzymes and related proteins may vary among different strains (electronic supplementary material, table SI). In some strains such as *Sylvio X10/1–202*, the number of 'dead' PGK isoenzymes exceeds that of catalytically active isoenzymes. On the other hand, analysis of the genome of the *T. cruzi* CCC strain revealed that its *pas–pgk* gene underwent a duplication process (electronic supplementary material, table SI). Together, these findings may be indicative of the high 'fluidity' of the genome of these organisms, resulting in frequent (partial) gene duplications and subsequent rapid evolution of redundant genes. Another important fact of this analysis is that, among all kinetoplastids analysed up to this point, *T. cruzi* is the one with the highest number of PGK isoenzymes located and expressed simultaneously in the glycosomes.

8.2.3.3.2. *Trypanosoma rangeli*. In *T. rangeli*, a protist closely related to *T. cruzi*, some *pgk* genes (and their products) have also been characterized. Our genome analysis of this kinetoplastid revealed four *pgk* genes (table 3), a result similar to that reported by Villafraz *et al.* [27]. The product of the TRSC58_02767 gene is a 45 kDa protein, lacking a typical PTS and indeed detected in *T. rangeli* epimastigotes. Its amino acid sequence has 85.7% identity with the cytosolic PGKB of *T. cruzi* [27]. The products of the TRSC58_00037 and TRSC58_04456 genes probably correspond to glycosomal proteins [27] (table 3). TRSC58_00037 codes for a multidomain 95 kDa PGK-like protein similar to the PGKL (with domains: PGK-CNB-HTH-TMD; table 3) reported for several African trypanosomes (*T. brucei* TREU927, *T. b. gambiense* and *T. evansi*) and also found in *Crithidia* (table 3). As in other trypanosomatids, this protein is predicted to be located in the membrane of glycosomes, except for the procyclic form *T. brucei*, where it was shown to be flagellar [60]. Regarding the product encoded by the TRSC58_00739 gene, previous studies have shown that this PGK is very similar to the 56 kDa PGKA of *T. cruzi*, with 82.1% sequence identity [27]. The similarity between these sequences has led us to propose that this enzyme probably has a location in the glycosomes

[27], although its sequence does not have a PTS1 motif. The TRSC58_04456 gene codes for a PAS–PGK, similar to that reported in previous studies of *T. cruzi* [4]. It shares 86% sequence identity with *T. cruzi* PAS-PGK. In addition to having a PAS domain in the N-terminal half, it contains a C-terminal tripeptide PKL (as its counterpart in *T. cruzi*), a PTS1, suggesting a glycosomal location of the protein [4,27].

The genomic organization of some of these genes in *T. rangeli* is similar to that reported for other trypanosomatids. The TRSC58_00739 and TRSC58_02767 genes are linked, while the other two genes, TRSC58_00037 and TRSC58_04456, are apparently located elsewhere in the genome [27]. This organization is reminiscent of that in *T. cruzi*, where the PGKA gene is followed by a PGKB/C gene. However, with regard to the subcellular localization of the expressed proteins, a difference was found between epimastigotes of the two species. PGKB of *T. rangeli* (product of the TRSC58_02767 gene) has an exclusively cytosolic location [323]. No PGK activity was detected in any other compartment than the cytosol [323], although the possibility that the products of genes TRSC58_00739 and TRSC58_04456 have also negligible activity in the glycosomes cannot be discarded, since the residues involved in substrate binding, catalysis and flexible hinge structure are highly conserved [27].

Regarding the kinetic characteristics of the PGKB of this parasite, both the natural and recombinant enzymes displayed standard Michaelis–Menten kinetics. The Km values for ATP are 0.13 mM and 0.5 mM, and for 3PGA 0.28 mM and 0.71 mM, in each case for the natural and recombinant enzyme, respectively (table 2) [27]. These values are comparable with previously reported Km values for the PGKB isoenzyme of *T. brucei* (Km^{ATP} 0.46 mM, Km^{3PGA} 2.04 mM) [69] and *T. cruzi* (Km^{ATP} 0.09 mM, Km^{3PGA} 0.62 mM; W. Q. *et al.* 2006, unpublished results). The similarity of these values reflects the high level of amino acids conservation in the active site between the PGKB of *T. brucei*, *T. cruzi* and this PGKB of *T. rangeli* [27]. The results obtained so far suggest the presence of a single PGK in the cytosol of the *T. rangeli* epimastigotes that acts in glycolysis when glucose is available in the environment. However, when this monosaccharide or other nutrients that may enter the glycolytic pathway is limited or absent in the vector insect, this PGK should participate in gluconeogenesis, when the parasite consumes lipids and/or amino acids [27,323].

8.2.3.3.3. *Trypanosoma grayi*. *T. grayi* is an African crocodile-infecting trypanosome; however, it is evolutionarily closer to *T. cruzi*, *T. c. marinkellei* and *T. rangeli* than to African trypanosomes [324,325]. Little is known about the metabolism of this trypanosome, but its genome has been sequenced. Like for other trypanosome species, several *pgk* genes are present in the *T. grayi* genome (table 3). The DQ04_00061260 and DQ04_02181000 genes apparently code for multidomain PGK-like proteins, while the DQ04_04781080 gene codes for a PGK protein consisting of a single domain.

Also a gene encoding a truncated PGK protein (DQ04_13681000) was observed in this trypanosome. It only retains the amino acids involved in the 3PGA substrate binding, but has completely lost key amino acids to carry out other functions such as catalysis and nucleotide binding, similar to what has been observed in some other trypanosomatids (see above and table 3). *Trypanosoma grayi* has a 58 kDa PAS-PGK (DQ04_00061260), like *T. cruzi*, with a possible location in the glycosomes, where it seems to be the only PGK isoenzyme. The multidomain PGK encoded by the

DQ04_02181000 gene is characterized by having also CNB and HTH domains, but not a transmembrane domain. Sequence analysis of this 80 kDa protein suggests that it is located in the cytosol, and lacks amino acids involved in substrate binding, thus it could be considered a catalytically 'dead' PGK enzyme, although it may exert other functions as discussed for PGKLs elsewhere in this paper (§8.2.3.2.1). By contrast, the 48 kDa PGK, encoded by DQ04_02181000, has all characteristics of a fully active enzyme and apparently has a location in the cytosol.

It is very interesting that half (2 of 4) of the genes in *T. grayi* discussed here are predicted to code for proteins that are probably catalytically inactive PGKs. This characteristic seems to be shared with *T. cruzi* and *T. congolense*. For several of these ORFs, information about their expression is still lacking. In some cases, the ORFs are clearly derived from partial gene duplications and can almost certainly be considered as silent, non-functional DNA. For some other ORFs, it is feasible that proteins are produced that exert other physiological functions than catalysing the PGK reaction. This aspect will be discussed in more detail, in §9.

8.2.3.4. *Angomonas* and *Strigomonas*

Angomonas deanei belongs to the trypanosomatid group of monoxenous parasites, including also *Crithidia*, *Blastocrithidia*, *Strigomonas*, *Herpetomonas* and *Leptomonas*, genera of organisms that together infect a wide range of insects [326,327]. *Angomonas*, along with the genera *Strigomonas* and *Kentomonas*, form a monophyletic clade within the Kinetoplastea, the subfamily Strigomonadinae (figures 1 and 2), that is characterized by the presence of a single β-proteobacterial endosymbiont in their cytoplasm [328]. Typically, there is only a single bacterium per trypanosomatid cell (*Candidatus* Kinetoplastibacterium or TPE (trypanosomatid proteobacterial endosymbiont)) that divides synchronously with the host protist; thus, at the end of the mitotic process, each daughter cell contains a single endosymbiotic microorganism [329].

Both organisms, the trypanosomatid and its endosymbiont, display a sophisticated interrelationship, in which the host is supplied with essential compounds including haem, nucleotides and essential amino acids. The endosymbiotic *Ca.* Kinetoplastibacterium spp. preferentially retain those genes necessary for the biosynthesis of compounds needed by their hosts [329,330]. Strigomonadinae can grow in defined media lacking haem and containing a reduced number of amino acids because many metabolites, important for the trypanosomatid's growth, can be synthesized by the endosymbiont and delivered to its host [330,331]. Besides the metabolic contribution, the endosymbiont affects the morphology and ultrastructure of the host cells [329,332,333]. Ultrastructural studies of *A. deanei* and *Strigomonas culicis* showed that many glycosomes are located in close association with the endosymbiont [333,334]. Based on studies of *A. deanei*, Motta *et al.* [333] postulated that ATP produced in the protist is used by the endosymbiont, since the bacterium's respiratory chain seemed to be inactive due to the absence of activity of cytochrome aa_3 oxidase and succinate dehydrogenase (nonetheless, it consumes O_2 once isolated from the host [335]). The ATP produced by the host cell can possibly be hydrolysed by the symbiont's ecto-ATPase, which may support an active transport system in the prokaryotic cell

envelope [334]. The observation that different endosymbionts in protists such as *A. deanei* and *S. culicis* maintain a close association with glycosomes suggests that metabolic products synthesized in these organelles are essential for the physiological maintenance of the bacterium and the symbiotic association. Besides ATP, other phosphorylated intermediates generated by the glycolytic pathway of the host, such as the 3PGA produced by PGK could be used by the endosymbiont [334].

A search in the database revealed in *A. deanei* 12 genes potentially coding for PGKs. Three of these PGKs share the characteristic length of 502 amino acids and are potentially active, because they seem to have all residues required for substrate binding and catalysis. Another PGK with 418 residues is also potentially active (table 3). None of these four PGKs has a consensus PTS, suggesting that these enzymes are located in the cytosol of *A. deanei*. A PGK with a PAS-domain with 526 amino acids was also detected, it could be active and located in the glycosomes, since it has at its C-terminus a PTS1 sequence (-PKL), which previously has been shown functional in targeting this protein to the glycosomes of *T. cruzi* and *Leishmania major* [28,104]. Seven other genes that code for PGK-like sequences are predicted to be inactive, because they lack residues involved in catalysis or substrate binding. One of them is a PGK of 270 amino acids; the other six, between 310 and 924 residues long, share the presence of between one to three transmembrane (TM) segments. Three of these latter proteins have also a CNB domain (but no HTH domain) located between the PGK and the TM domain (table 3). All of these 12 genes belong to *A. deanei*. A PGK gene of *Ca.* Kinetoplastibacterium spp. was also identified (WP_015238965.1), coding for a product with a molecular weight of around 40 kDa but predicted to be inactive.

A similar situation was found for *S. culicis* that has at least 10 genes coding for PGKs, seven of them for active PGKs. Three of these PGKs with length of 417, 455, and 500 amino acids are probably in the cytosol and we hypothesize that the protein of 455 residues with a TM is associated with the glycosomal membrane. A PAS–PGK of 527 amino acid possesses a PTS1 (-PKL) and thus will most likely be located in the glycosomes. Three proteins of 935 residues with three different domains, PGK-CNB-TMD, have been identified; two of them (STCU_08802; STCU_06143) are, in contrast with their orthologues in *A. deanei*, predicted to be active, while the inactive (STCU_02304) protein has one additional TM segment. Among the other inactive PGKs found is one of 270 residues, and another one of 458 amino acids with a CNB domain included (table 3). Proteomics analysis on wild-type *S. culicis* (with the endosymbiont) and a corresponding aposymbiotic (Apo) strain (i.e. without endosymbiont) showed that throughout growth in batch culture, the PGK, PAS-PGK (STCU_04975) and PGKc (STCU_01151) are constitutively expressed at similar levels in both strains [336].

Five different PGK sequences belonging to the endosymbiont (*Ca.* Kinetoplastibacterium blastocrithidii) were found. However, an alignment of these sequences showed that all of them code for the protein identified with the sequence WP_015237848.1 (table 3).

From this analysis of PGK sequences reported for *A. deanei* and *S. culicis*, we infer that the 3PGA can be generated within glycosomes as well as in the cytosol of these endosymbiont-containing hosts, in a similar way as other trypanosomatids like *T. cruzi* which contain PGKs located in the glycosomes and cytosol.

8.2.3.5. *Leishmania*

Different studies of PGK isoenzymes have been carried out for *Leishmania* spp. [28,81,279,281,337,338]. These studies have shown that these parasites have two tandemly linked PGK genes on chromosome 20 (figure 5), corresponding to the genes *pgk*B and *pgk*C in *Trypanosoma* and *Crithidia*. These genes code for a 45 and 51 kDa protein, respectively (table 3). The two polypeptides are 99.5% identical in their sequences. As in other trypanosomatids, the PGKB isoenzyme is in the cytosol, while PGKC is compartmentalized in glycosomes [81,279,281]. The isoenzyme PGKC is characterized by having a long C-terminal extension of 63 amino acids that is unique to PGKCs of the genus *Leishmania* [81,281,338]. This extension, being the only difference between PGKB and PGKC besides two residues substitutions elsewhere in the proteins, is most likely responsible for routing the isoenzyme to the glycosomes. No PTS1 is present in its C-terminal sequence, but an internal sequence acting as glycosomal import signal has not been ruled out. A notable feature of this 63 residues long C-terminal extension is that it has alternating stretches of hydrophobic and charged, mainly positive amino acids. These characteristics may be involved in a protein–protein interaction, allowing the association of the PGKC with other soluble or membrane-associated glycosomal proteins. Additionally, it could serve as a binding site for some ligand, allowing the modulation of the activity of this isoenzyme [281]. On the other hand, theoretical and *in vitro* studies of the PGKC of *L. mexicana* by Raghunathan and co-workers [68] demonstrated that this extension has a propensity to form a transmembrane helix located between residues 451–472, ending in a cluster of two Arg residues (R471/R472). Therefore, there is also a possibility that the extension serves to anchor the PGKC to the glycosomal membrane. This idea of a membrane association could be supported by results obtained by Quiñones *et al.* [339], who observed a 56 kDa protein that is not an integral membrane protein, but seems firmly attached to the inner face of the glycosomal membrane or is a peripheral component of a membrane protein complex. Based on further theoretical considerations, Raghunathan and co-workers [338] proposed that the C-terminal extension may form a metabolite or ion transporter within the glycosomal membrane. However, *in vitro* or *in vivo* experimental studies supporting this notion are lacking. Although the 63 residues C-terminal extension is present in all *Leishmania* species analysed, with the sequence being conserved at 65–96% identity pointing to a shared function, it is important to consider that such a long extension is not common in the orthologues of trypanosomatids; it is only found also in the *Leptomonas* spp. glycosomal PGK (with 40–53% sequence identity to the *L. mexicana* PGKC extension), whereas in all other trypanosomatids studied, it is much shorter or even absent (e.g. in *Crithidia* 38, *T. brucei* 20 and *T. cruzi* 0 residues). Moreover, there is strong evidence that in all peroxisomal, including glycosomal membranes non-selective channels are present that allow the translocation of ions and metabolites with molecular weight up to 300–400 Da [340], rendering unlikely the possibility of an additional, PGK-coupled transport system only in *Leishmania*.

One aspect of the *pgk* genes in which *Leishmania* differs from other trypanosomatids studied, (i.e. *T. brucei*, *T. congolense*, *C. fasciculata* and *T. cruzi*) is the absence of a third PGK gene in a tandem array, a *pgk A* gene coding for a functional PGK with a long insertion in its N-terminal half [81,279,281]. The consequence of this absence is still unknown; however, different explanations have been proposed. *Leishmania* is less dependent on glycolysis than other trypanosomatids such as *T. brucei*; previous studies with axenic amastigotes of *Leishmania* spp. and amastigotes in *in vitro* infected (non-activated) macrophages showed that the glucose utilization is markedly lower than fatty-acid or amino acid oxidation [341,342]. However, recent research demonstrated that amastigotes in activated macrophages and in animal models rely on glycolysis [343].

Although the PGK B and C enzymes of *L. mexicana* are very similar to their developmentally expressed B and C orthologues in *T. brucei* (to which they have around 73% identity for both isoenzymes), the expression and activity attributed to each of the enzymes are rather like those of the isoenzymes of *T. cruzi* (to which the identities are approx. 76%). Both isoenzymes are also expressed simultaneously in *Leishmania*, with 80% of the PGK activity found in the cytosol and 20% in the glycosomes of cultured *L. mexicana* promastigotes [281]. This difference is related to the differences in expression levels of these isoenzymes, with the transcripts and proteins of PGKB and PGKC both being expressed at a ratio of approximately 4 : 1. The protein levels are maintained as a result of multiple levels of control, both posttranscriptional and posttranslational [279]. Regarding the function of these enzymes, the 80/20 expression of the PGK isoenzymes could reflect how fluxes have to be routed through the different branches of carbon metabolic network that is distributed over the glycosomal and cytosolic compartments (the glucose catabolism to pyruvate or via PEPCK and malate dehydrogenase to succinate, and gluconeogenesis to glucose 6-phosphate), while the ATP consumption and production and NAD^+/NADH redox level within the glycosomes have to be kept in balance [281,344].

As for *T. cruzi*, a PAS–PGK, encoded by the LmjF.30.3380 gene, was recently described for *L. major* (LmPASPGK) [28], and is also present in other *Leishmania* species (electronic supplementary material, table SI). This enzyme has a molecular weight (58 kDa) and domain structure that are very similar to those of the *T. cruzi* PAS–PGK [4; table 3]. Little information is known about a possible ligand to its PAS domain. Its PGK activity appeared to be insensitive to FAD, FMN, 4-hydroxycinnamic acid and fatty acids. However, association of this enzyme with haem in *in vitro* assays increased its activity by 20%. Regulation of the activity is a distinctive feature of this LmPAS-PGK. It has maximal activity at pH 5.5 that decreased towards a pH close to 7.5. This effect of pH on PGK activity occurs indirectly through the PAS domain. Kinetic studies also demonstrated that the absence of the PAS domain has no effect on the affinity of the enzyme for its substrates (ATP/3PGA) [28]. It may be relevant to note that the Km values of the LmPAS-PGK for its substrates (118 μM to ATP and 661 μM) are very similar to those of PGKB of *T. cruzi* (W. Q. *et al.* 2006, unpublished results; table 2). Interestingly, a dual subcellular location has been reported for this *Leishmania* PGK isoenzyme, in glycosomes and lysosomes. The enzyme contains a PTS1 (the tripeptide -PKL) for import into glycosomes and seems to have, in that same C-terminal region before the glycosomal sorting signal, a post-Golgi-sorting motif (FLELL), which is probably responsible for its translocation to lysosomes [28]. This putative lysosomal translocation signal in LmPAS-PGK is very similar to the known dileucine-based sorting signals identified in lysosomal proteins [345,346] and appears to be conserved in the PAS-PGKs of other trypanosomatids. The lysosomal localization of LmPAS-PGK is intriguing and possibly provides a rationale for why the pH optimum for PGK activity at 5.5 has developed, a pH unlikely to be encountered within glycosomes [340], although the availability of PGK substrates (3PGA/1,3BPGA and ADP/ATP) in lysosomes and how this activity relates to a physiological role of the enzyme in these organelles remain to be established. Adhikari *et al.* [28] suggested that this LmPAS-PGK may exert different functions in the survival of the parasites, including regulation of the glycolytic flux and autophagosome formation in promastigotes, and probably has also a role in disease development in macrophages or mice. LmPAS-PGK, like other PAS proteins, seems to be an enzyme that links different cellular processes, similar to what we previously proposed for TcPAS-PGK [4]. This would not be the first report of a PAS protein with a possible function in autophagy. Nuclear receptor coactivator 4 (NCOA4) is a transcriptional co-regulatory protein that contains a basic helix–loop–helix domain, a PAS domain and seven LXXLL motifs [347]. The main function attributed to NCOA4 is to interact with several activated nuclear transcription factors, including the aryl hydrocarbon receptor and the androgen receptor [348]. However, NCOA4 can fulfil also other previously unknown functions, such as apparently acting as an autophagy cargo receptor [349].

Like in other kinetoplastids discussed in this review, a pseudogene encoding a part of a PGK-like protein was identified in *L. major* (table 3 and figure 5). However, the predicted protein is very peculiar. The software used for its sequence analysis predicted it to be a bifunctional PGK–TIM enzyme, having a part of a TIM-like protein fused at the C-terminal side of the PGK fragment. Bifunctional PGK–TIM enzymes have previously been identified in thermophilic bacteria, such as *Thermotoga maritima* [84]. In *T. maritima*, the 70 kDa PGK–TIM fusion protein is an elaborate bifunctional 'polyprotein' that carries both PGK and TIM activity. Unlike canonical PGKs, this PGK–TIM polyprotein forms a tetrameric complex that only exhibits significant catalytic activity at high temperatures [84]. Also *T. neapolitana* produces PGK as both an individual enzyme and a PGK–TIM fusion protein, but unlike *T. maritima*, which forms only one PGK–TIM fusion protein, *T. neapolitana* expresses two additional PGK–TIM fusion proteins, of 73 and 81 kDa [350]. A possible advantage for a cell to combine PGK and TIM in a single protein is enigmatic, because the two enzymes catalyse reactions which do not represent consecutive steps in the glycolytic pathway. Bifunctional enzymes usually catalyse two reactions involving a common metabolite. Such enzymes can also be found in parasitic protists like trypanosomatids. One example is the bifunctional enzyme dihydrofolate reductase/thymidylate synthase (DHFR-TS), encoded by a single copy gene (*dhfr-ts*), that plays a role in the reduction of folate, a key cofactor for DNA synthesis [351–353]. It has an N-terminal DHFR domain and a C-terminal TS domain, with the two domains separated by a junction peptide of variable size. Two sequential catalytic activities are contained on a single polypeptide chain, and the metabolite is channelled between both catalytic sites. Another, different kind of example is the 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase, in which the two linked enzymes are, respectively, responsible

for the synthesis and hydrolysis of the potent glycolysis allosteric regulator fructose 2,6-bisphosphate [354]. However, whether PGK, or any other glycolytic enzyme is involved in the formation of such a bifunctional enzyme in kinetoplastid protists would require further studies.

8.2.3.6. Leptomonas

In the monoxenous insect parasite *Leptomonas pyrrhocoris*, the presence of *pgk* genes is similar to that observed in other trypanosomatids analysed so far; its genome contains several genes encoding PGK isoenzymes. Two of these genes are arranged in tandem on chromosome 25 and the other two are located on chromosomes 4 and 7 (figure 5). These isoenzymes have molecular weights ranging from 51 and 101 kDa (table 3). Two of these PGK isoenzymes are part of multidomain proteins, with the PGK domain being associated with a CNB domain (in LpyrH10_07_2090) and a PAS domain (in LpyrH10_04_6440). The other two (LpyrH10_25_1860 and LpyrH10_25_1870), which are approximately 56% identical, are structurally canonical PGKs (table 3). It is important to highlight that analysis of the sequences of these latter two proteins revealed for both a probability for a glycosomal localization, despite the absence of consensus PTS motifs, as well as a possible localization in the mitochondrion, in agreement with the presence of mitochondrial-like amino-terminal targeting sequences in these isoenzymes (results not shown). Although there is no experimental evidence yet for the presence of glycosomes in *Leptomonas*, previous genomic analysis suggested that the initial part of the glycolytic pathway (from HK to PGK) as well as several other enzymes such as PPDK, which are commonly found in glycosomes of trypanosomatids, are similarly compartmentalized in *Leptomonas* [288]. In addition, genes for Krebs cycle enzymes and all the subunits of the pyruvate dehydrogenase complex were identified, also indicating the presence a functional mitochondrion in *L. pyrrhocoris* [288]. Based on a possible mitochondrial localization of some PGK isoenzymes, it would be valid to wonder what the meaning of the presence of these enzymes in the mitochondrion of *L. pyrrhocoris* could be. There are reports of tumour cells and protists for which the presence of a PGK in mitochondria has been documented [183,273]. As for *T. vivax*, we suggest that these isoenzymes may have a non-canonical function or an additional one to their canonical one.

Although the number of reports related to sugar and polysaccharide metabolism in *L. pyrrhocoris* is limited, this protist can use a wide variety of molecules as carbon sources. Based on genome analysis, this protist is predicted to have the ability to consume hexose sugars (glucose, fructose, mannose and galactose), pentose sugars (ribose, xylulose and ribulose) [288] and polysaccharides such as chitin, the latter being abundant in insects [355]. Additionally, this parasite can use some amino acids such as proline and glutamate as carbon sources, which are also abundant in the intestine of insects [356]. Although this organism is classified as monoxenous, variations in the carbon sources provided by its host are possible.

It is important to highlight that duplication of *pgk* genes in *Leishmania* and *Leptomonas* reported in this work is in accordance with the finding by Durrani *et al.* [321] that these organisms are prone to polyploidy, a feature that they reported in particular for genes of glycosomal matrix proteins.

8.2.3.7. Crithidia

Like in several other kinetoplastids, PGK has been also experimentally characterized in *C. fasciculata* [80,125]. These studies have shown that *Crithidia*, like *T. brucei*, contains three *pgk* genes, A, B and C, in a tandem array located on chromosome 18 [125] (figure 5). Genes B and C encode a cytosolic (PGKB, 45 kDa) and glycosomal (PGKC, 48 kDa) isoenzyme, respectively. The *pgkA* gene has an ORF for a 56 kDa PGK, which has most active site residues conserved and is characterized by having an insert of 80 amino acids. This insert sequence is conserved between *T. brucei* and *C. fasciculata* with 48% positional identity [80], even though the lineages to each of these parasites have separated by almost 300 Myr [357]. The results of our bioinformatics analysis for this review for the three genes and their expressed products, as catalogued in table 3, are in full agreement with those from the previous studies by Swinkels *et al.* [80,125].

No biochemical studies have yet been carried out of the *Crithidia* PGK isoenzymes. However, proteomics analysis has shown that the glycolytic pathway is highly expressed during the early and mid-logarithmic growth phase of *in vitro* cultured choanomastigotes of *C. fasciculata* [358]. During these growth phases, various glycolytic enzymes (ALD, HK, PYK) increase their expression importantly. This could indicate that a high activity of the glycolytic pathway at mid-logarithmic phase is required to provide energy and essential precursors of certain amino acids, ribonucleotides and derived coenzymes. PGK and other enzymes such as TIM, that catalyse reactions that under physiological conditions are easily reversible, are constitutively expressed, unlike the others mentioned above [358]. Furthermore, like other trypanosomatids analysed in this work, genes that code for multidomain PGKs (i.e. a PAS–PGK and a PGK with a CNB domain) were also identified in *Crithidia*. Interestingly, the orthologue of PAS–PGK was shown to be upregulated at the transcript level in amastigotes compared to stationary phase promastigotes [359]. As observed for *T. cruzi*, *Leishmania*, *Leptomonas* and *Crithidia*, the number and nature of PGK isoenzymes can vary within different isolates. However, in the case of these kinetoplastids, the presence of 'dead' PGK enzymes is less common (electronic supplementary material, table SI).

8.2.3.8 Summary of the functions of kinetoplastid PGKs

The differential expression of PGK isoenzymes present in kinetoplastids seems to be related to adaptation of these (free-living and parasitic) organisms to changing environmental conditions. When analysing the genomes of different kinetoplastid species, we found several important data about this glycolytic/gluconeogenic enzyme. (i) The number of genes (potentially) encoding PGK isoenzymes and PGK-related proteins ('dead' PGKs) can vary greatly in these protists, from 1 (in *Perkinsela*) to 6 (in *T. congolense* and *T. cruzi*). (ii) Some of these PGKs have additional domains such as a PAS, CNB and/or some transmembrane regions of which the location within the overall protein structure may vary. The two former modules are known to be responsible for regulation of a wide diversity of cellular processes (from metabolite transport to differentiation) in cells. (iii) (Pseudo)genes for one to several catalytically 'dead' PGK isoenzymes are present in some kinetoplastids. This lack of catalytic activity is most often due to the absence of the binding site for 3PGA/1,3BPGA or ADP/ATP, either resulting from residue substitutions within these sites

or from a major deletion or truncation, or in some other cases only due to the mutation of an essential catalytic residue. PGK isoenzyme orthologues may be present in different kinetoplastid species; however, they may, according to sequence-based predictions, have different subcellular location sites. Additionally, in some species, 'dead' PGK enzymes have been shown to be expressed where they can represent up to 70% of the total PGK-related protein content. (iv) In the case of *Leishmania*, a pseudogene for a short protein was detected that, according to the prediction by the software used, combined stretches of PGK and TIM-related sequences. However, based on levels of similarity, structural organization of the ORF and the exclusive finding of the pseudogene in species from only one genus, one may wonder if it is a vestige of what once has been a gene for a bifunctional PGK–TIM. Further studies may be required to address this issue. (v) Some of these isoenzymes, such as PAS-PGK, are totally absent in African trypanosomes. (vi) The number of genes for PGKs and PGK-related proteins can vary significantly between strains of a species, at least for *T. cruzi*, *Leishmania* and *Leptomonas*. (vii) PGKs and PGK-related proteins can have different (experimentally determined or sequence-based predicted) subcellular locations, such as in glycosomes, mitochondrion, flagellum, lysosomes and/or cytosol. The biological function, if any, of *pgk* gene duplications and the consequent presence of several, sometimes highly diverse PGK isoenzymes in these parasites remains poorly understood. This protein has probably more relevant physiological functions than has been attributed to it so far. As has been documented in other organisms, it is very likely that several of these isoforms have additional important biological functions. However, it should be realized that for many of the PGK isoforms we described in this section about kinetoplastids, the currently available information is based on bioinformatic analysis of their genomes. In those cases, future proteomic analyses will be required to determine if expression of a (stable) protein occurs.

9. Phylogenetic analysis

We have performed a phylogenetic reconstruction of three groups of isoforms that are represented in different species and of which expression has been confirmed in at least some of them: the canonical PGKs, PAS–PGKs and PGK–CNB– HTHs. Figure 6 shows the consensus tree of relationships obtained by the neighbour-joining method for 268 amino acid positions of 57 sequences. The tree topology found for the taxa belonging to the orders Trypanosomatida, Eubodonida and Euglenozoa corresponds to what has been reported previously [293,365]. The phylogenetic analysis reveals that each paralogue of PGK forms a different cluster with the 'canonical' one, including the A, B and C isoenzymes. The PGKA and B sequences from stercorarian *Trypanosomas* are all grouped in a single clade. With respect to the salivarian *Trypanosomas* PGKs (A, B and C) sequences, these are grouped in two clades, one comprising all sequences from *T. b. brucei*, *T. b. gambiense* and *T. evansi*, and the second one grouping the PGK isoenzymes B and C from *T. congolense*. However, separately branching from these two clades is the cytosolic PGK sequence from *T. vivax*, which is within the stercorian clade, close to the PGKB from *T. grayi*. Another clade, subdividing itself in two clusters, comprises in one cluster all PGKB and

C sequences from the different species of *Leishmania* (*L. major*, *L. donovani* and *L. infantum*), and in the second cluster, the PGKB and C sequences of their monoxenic trypanosomatid relatives *C. fasciculata* and *Leptomonas* (*L. pyrrhocoris* and *L. seymouri*). In independent branches are the other protistan sequences, the putative PGK (56 kDa) from *B. saltans*, the unique PGK (55 kDa) of *Perkinsela* sp. and the cytosolic PGK (45 kDa) from *E. gracilis*, these latter ones being most closely related to sequences from *L. pyrrhocoris* (LpyrH10_25_1860) and *C. fasciculata* (CFAC1_180007100), both of them with a molecular weight of 55 kDa like PGKA.

Two other major clusters are formed by the PGK–CNB– HTHs and PAS–PGKs from the kinetoplastids. The cluster of PAS–PGKs is subdivided into two main clades, one of them subdivided further in a group comprising sequences from *Leishmania* spp. and a second group with the PAS– PGKs from *C. fasciculata* and *Leptomonas*. The other cluster comprises all the sequences from stercorarian *Trypanosomas*. Although the PAS–PGK is completely absent in the salivarian *Trypanosomas*, the sequence TvY486_0604100 from *T. vivax* is in this cluster, because this gene product, although lacking a PAS domain, shares a high identity (approx. 60%) with the PGK domain of these sequences. Indeed, this *T. vivax* PGK gene presents synteny with PAS–PGK genes from all trypanosomatids that contain it (electronic supplementary material, table SII). By contrast, branching independently is the PAS–PGK from *B. saltans*, a member of the order Eubodonida. In the PGK–CNB–HTH cluster of *Trypanosomas*, all sequences from the salvarian trypanosomes are grouped in a single clade, which is separated from the clade of sequences from stercorarian trypanosomes. In a similar manner are grouped sequences from *C. fasciculata* and *Leptomonas* (*L. pyrrhocoris* and *L. seymouri*). In a separate branch is the PGK (PCON_0003620) from *P. confusum* which forms the most basal lineage of trypanosomatids, and the PGK of the free-living eubodonid *B. saltans*.

The phylogenetic tree also allows us to distinguish different types of homology, namely 'paralogues' (derived from gene duplication events) and 'orthologues' (from speciation events) [366]. Specifically, 'canonical' PGKs (A, B, C) of *T. brucei, T. b. gambiense* and *T. evansi* each inherited an orthologous gene ancestry, which then underwent one or more rounds of duplication and divergence to produce the different types of canonical PGKs in each descendent lineage. With respect to the duplication event from which arose the paralogous PGKs A and B of *T. cruzi* and *T. rangeli*, these nodes are strongly supported by high maximum-likelihood bootstrap values (86% and 90%, respectively). The same behaviour was found for the rest of sequences from the Trypanosomatida order that are highlighted with a blue oval in figure 6.

10. What significance has *pkg* gene duplication?

Gene duplication is an important mechanism for acquiring new genes, and creating genome rearrangements and genetic variation [367]. It can take place through different processes that include retroposition, duplicative transposition, polyploidization, exchange between homologous chromosomes and exchange events within a single chromosome [367,368]. These mechanisms of gene acquisition have been studied in a wide variety of organisms, including protists, both with a
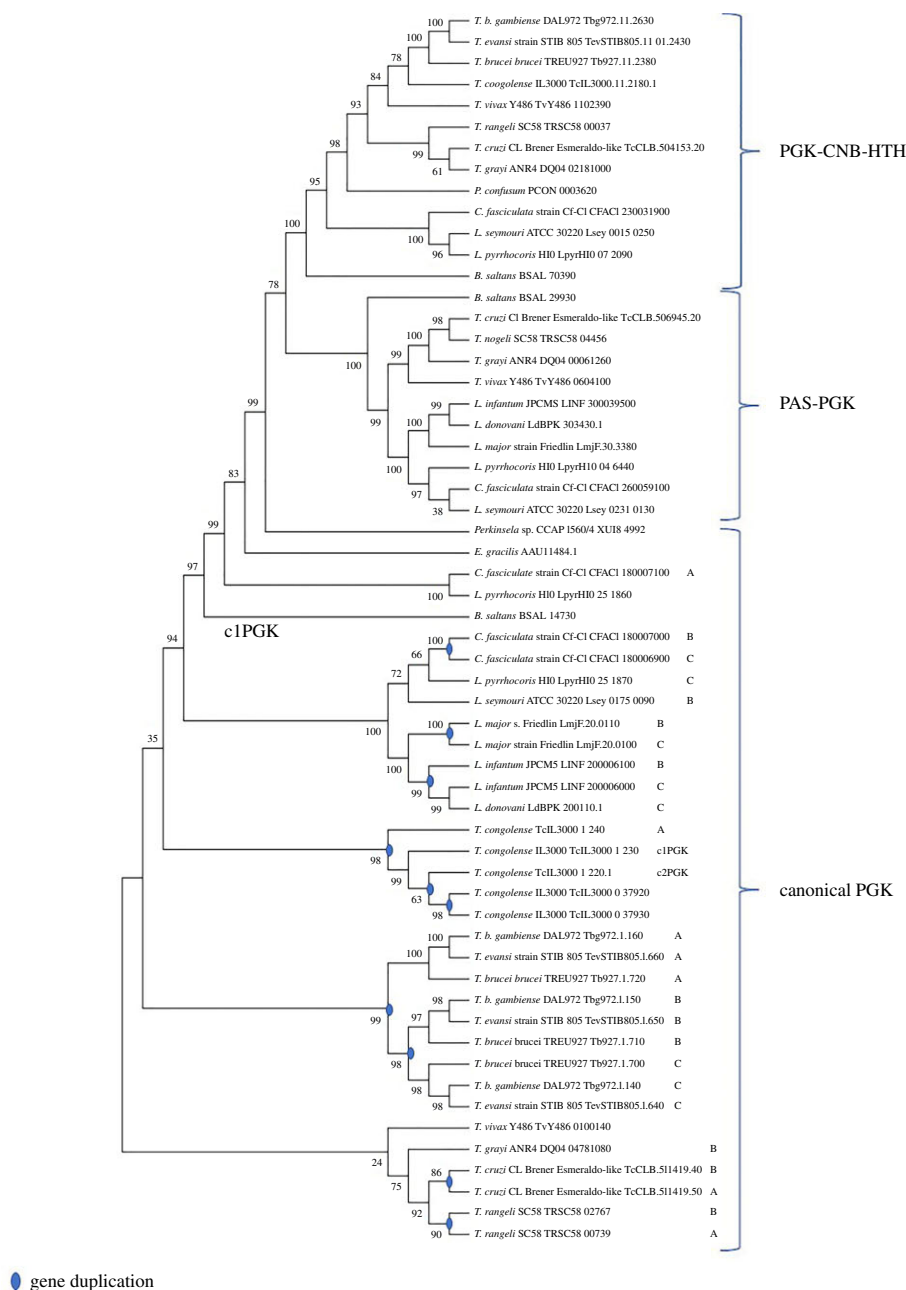
Figure 6. Kinetoplastid phylogenetic tree based on PGK sequences. PGK gene sequences were retrieved from the TriTrypDB, GeneDB and GenBank databases [94,123,360], and the predicted amino acid sequences aligned using Muscle (EMBL-EBI) [95]. Regions of uncertain alignment were omitted from the analysis leaving a total of 57 sequences (19 taxa) and 268 amino acid positions. The evolutionary history was inferred using the neighbour-joining (NJ) and the maximum parsimony (MP) methods [361], with similar results. The branches of the tree represent sequences, orthologous and paralogous, from different members of the Kinetoplastea phylum, with nodes depicted as filled blue ovals representing gene duplication events. Support for the nodes in the NJ and MP analyses was evaluated using the bootstrap procedure, and in each case, the bootstrap consensus tree inferred from 1000 replicates [362] was taken to represent the evolutionary history of the taxa analysed [362]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [362]. For the NJ analysis, the evolutionary distances were computed using the p-distance method [363] and are given in the units of the number of amino acid differences per site. This analysis involved 58 amino acid sequences. All positions containing gaps and missing data were eliminated (complete deletion option). There were a total of 268 positions in the final dataset. For the MP analysis branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [361]. The MP tree was obtained using the subtree-pruning-regrafting (SPR) algorithm ([362], p. 126)] with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates). This analysis involved 57 amino acid sequences. There were a total of 391 positions in the final dataset. Evolutionary analyses by NJ and MP were conducted in MEGA X [364].

free and parasitic lifestyle [369,370]. Gene expansion is considered to have been important for the evolution of adaptability of organisms to environmental variations and of their development to parasitism, for the latter especially involving those genes that contribute to a parasite's pathogenicity and resistance to host immunity [365,369–373]. The origin of the duplication of *pgk* genes in trypanosomatids and selective benefit of it for these organisms have as yet not been addressed.

In some organisms, such as the nematode *Strongyloides papillosus*, the evolution of parasitism is associated with a large expansion and chromosomal clustering of the CAP and Astacin gene families [374]. These two gene families

experienced an extreme extent of duplication that coincided with the appearance of parasitism and were maintained during speciation events that ultimately resulted in different species of parasites that infect a wide variety of vertebrate hosts, including humans [375]. Both families of proteins, which are secreted by the parasites, fulfil important functions during infection, such as immunomodulation, tissue digestion and penetration [375,376]. In addition to being one of the most abundant gene groups in the genome, the *CAP* and *Astacin* genes are differentially expressed during the life cycle of these parasites [371].

On the other hand, studies with *Schistosoma mansoni* revealed that small-scale gene duplications (i.e. tandem, dispersed and proximal (TDP)) may be derived from retrotransposition and unequal crossing over. These genes derived by TDP duplications have undergone small divergence during evolution. An important aspect of this group of small-scale duplicated genes in *S. mansoni* is that they are part of some gene families associated with pathogenesis, involved in parasite survival during infection [372]. In *Fasciola hepatica*, something similar occurs. Although this organism has one of the largest genomes sequenced so far (1.3 Gb) [373], three times larger than that of *S. mansoni* (363 Mb) [377], its ability to infect and survive in different tissue environments (as it migrates from the intestine, through the liver and into the bile ducts) is mainly based on products of duplicated genes [373,378,379]. The genes of the tubulin, proteases and proteins involved in the glycosylation have undergone a marked process of duplication and are those that in turn exhibit an appreciable change in their expression during the life cycle [373,378]. Although gene duplication seems to be a key process for adaptation to the parasitic lifestyle of helminths, we can observe that different evolutionary solutions have been developed between different taxa that involve the duplication of different gene families.

The capacity for gene innovation introduced by duplications seems also to play an important role in apicomplexan protists. In these organisms, certain regions of the genome are enriched in multi-copy genes and some of these duplicated genes are species-specific. This specificity could indicate a role for duplication in adaptation and genetic innovation. Additionally, many of these recently duplicated genes are arranged in tandem. A distinctive feature of these tandemly arranged duplicated genes that are apicomplexan species specific is their distribution along the chromosome; they are located mainly in subtelomeric regions. Until now, the reason for the susceptibility of these regions to gene duplications is unknown. But it is clear that the duplications of these genes (which code for proteases, protein and DNA interacting proteins, etc.) contribute to the pathogenicity and resistance to host immunity of these organisms [380].

In *E. histolytica*, genes involved in lysis of host tissue and nutrient capture from the extracellular environment have undergone extensive gene duplications [381]. Furthermore, in *G. lamblia*, 40% of its genes have been identified as being duplicated and many of these genes are organized in tandem. Most of these duplicated genes encode variant-specific surface proteins (VSPs), ankyrin (Ank), and ADG (ancient duplicated genes) PKs. VSP proteins are essential for the successful parasitic life of *Giardia* in its hosts [382]. The Ank protein is a cytoskeleton component and is involved in cell cycle regulation and signal transduction [383] and ADG PKs are implicated in vesicular trafficking.

In the halophilic stramenopile protist *Halocafeteria seosinensis*, adaptation to halophilic habitats apparently depends on the acquisition of new genes through duplication. A large part of these duplicated genes (43%) are differentially expressed in response to the presence of salt in the environment [384]. In the ciliate *Paramecium tetraurelia*, most of the nearly 40 000 genes arose through at least three successive whole-genome duplications and many of these duplication processes coincided with speciation events. In addition, most of these genes (51%) are present in two copies, both highly expressed. However, duplication has apparently remained limited to gene dosing, rather than contributing to functional innovation [385].

In kinetoplastids, studies conducted with *B. saltans* revealed that its genome has high redundancy, with many tandemly duplicated genes; 34.7% of its genes are present in more than one copy. Some duplicate genes have an extremely high copy number, such as those of the GP46-like surface antigen gene family, of which 391 copies are present in the genome [282]. Some authors proposed that in the case of some asexual organisms (a characteristic attributed to *B. saltans*), the presence of multiple identical gene copies can help to counteract Muller's ratchet phenomenon. This is a process in which genomes of an asexual population accumulate harmful mutations irreversibly. However, the only way these organisms would be able to cope with it is by having multiple copies of the gene in question [386]. In parasitic kinetoplastids, gene duplications could have this and additional functions. Trypanosomatids are characterized by having a marked genomic plasticity that offers great ability to remodel or induce changes in their genome in response to environmental changes. These changes in the genome can occur at the overall chromosomal level (altered number (aneuploidy) and size) or in specific parts of individual chromosomes (gene amplification or deletion) [314,369,387,388]. Analyses of gene amplification/duplication in these organisms indicate that gene duplication is a potential source of genetic diversity and an important evolutionary factor that directly affects fitness [5,282,387,389–391]. Through this mechanism, there is an increase in the variability and alteration of gene expression in response to environmental conditions encountered in the host or by stressor agents like drugs [387]. The massive duplication of some genes would not only promote the creation of paralogous genes, but their formation would also have different consequences such as: (i) the neofunctionalization of duplicated genes; (ii) segregation of the original gene function between duplicates (subfunctionalisation); (iii) a change in the expression of the duplicated genes, in the case of parasites, for example, during the different stages of life cycle development, mediated by possible changes in their regulatory regions; and (iv) loss of function (pseudogenisation, or non-functionalization) [280]. Many of the genes that have undergone significant amplification are related to the parasite–host interaction. So appear the products of these amplified genes to be related to the process of cell adhesion/invasion, immune evasion, cytokinesis and parasite survival. Some of these duplicate gene products in trypanosomatids include surface proteins, peptidases, proteins involved in glycosylation, calpain-like cysteine protease, adenylyl cyclases and Ras-like proteins [369,387,388,392,393]. Another feature of these tandemly arranged duplicate genes is their location. In *T. cruzi* and *Leishmania* spp., these genes are all distributed internally in chromosomes and subtelomeric regions, while in *T. brucei*, several of these multigene families' clusters are enriched in subtelomeric regions [289–291]. In eukaryotes, the subtelomeric regions are preferential sites for the evolution

of multigene families. However, little is known about the mechanisms involved in this process [394]. A distinctive feature of tandem duplicated genes is that many of them are pseudogenes. In the case of *T. brucei*, variant surface glycoproteins (VSGs) are involved in antigenic variation of this parasite. The *T. b. brucei* genome has approximately 800 different VSG genes, of which 18% are gene fragments and 66% full-length pseudogenes [289,291]. Some studies of these VSG pseudogenes have led to the suggestion that they can be used to generate expressed mosaic genes by ectopic recombination, and this recombination in turn allows the generation of new VSG variants [289,394]. The massive duplication into multigenic families in trypanosomatids can constitute evidence of the specialization that these parasites have to undertake to survive, as they must deal with different stressors within their hosts. *T. cruzi* and *Leishmania* spp. use this arsenal of proteins to invade and survive within the host cell, while in *T. brucei*, the multigene families are involved in antigenic variation to evade the humoral immune response and survive in the host bloodstream [387].

In the case of the *pgk* multigenic family, studies have so far mainly been performed from biochemical and structural points of views; the presence of several PGK isoenzymes is attributed to their role in the carbohydrate metabolism of these parasites. Little is known about the significance or reason for the existence of multiple, distinct PGK isoforms and the mechanism and selective force that have led to their preservation. However, the presence of various PGKs in kinetoplastids can have different physiological 'raisons d'être' that are closely related to each other and to the appearance of parasitism. The differentiation and retention of these isoforms may serve different purposes and be involved in different cellular processes.

10.1. Back-up function

Based on studies with *Saccharomyces cerevisiae*, it has been postulated that duplication and conservation of genes for metabolic enzymes are a mechanism for creating genetic robustness by 'genetic buffering' and functional complementation from duplicate genes [395,396]. This 'genetic buffering' through redundant gene networks would include alternative metabolic or regulatory or signal pathways, while complementation would be a way by which deletions in genes will have no significant effect on the phenotype [395,397]. This connection between robustness and redundancy is known as a back-up function. In the case of genetic robustness, functional compensation by duplicates might be a product of functional divergence after gene duplication. Additionally, a duplicate can be recruited into a novel gene network (or other types of functional divergence), setting up a boundary to stop its further divergence. In addition to providing robustness, gene duplication implies a continuous creation of new genes during evolution and increased complexity of molecular pathways [395]. Back-up of important or essential functions with duplicate genes plays an important role in genetic robustness, since individual duplicates would compensate for genetic dysfunctions. However, some members of a family of duplicates can acquire a specialized function, restricted expression or localization pattern that precludes functional complementation [396].

Different PGK isoforms could provide the kinetoplastid parasites with environmental robustness, since a role in the adaptation to environmental variations has been attributed to gene duplications [398]. The presence of multiple isoenzymes could allow organisms to survive through a wide range of environmental conditions without significantly altering their fitness, while they can increase the phenotypic robustness of the metabolic system [399,400]. For example, in *Drosophila melanogaster*, the presence of casein kinase II (CKII) isoforms increases phenotypic robustness and provides a flexible and parsimonious mechanism for adaptation to environmental fluctuations [400].

In *T. gondii*, metabolic plasticity is considered a crucial aspect for its successful parasitism in various host organisms and cell types. *T. gondii* tachyzoites primarily use glucose as energy source and a wide variety of anabolic precursors via glycolysis, the phosphate-pentose pathway and mitochondrial processes. They can also metabolize amino acids such as glutamine, after switching to the use of amino acids as a primary carbon and energy source when the host glucose is limiting [251]. Although *T. gondii* shows this preference for glucose, tachyzoites constitutively express also some enzymes involved in gluconeogenesis, including two isoforms of fructose-1,6-bisphosphatase (FBPase), which catalyse the dephosphorylation of fructose 1,6-bisphosphate (FBP) to fructose 6-phosphate [401]. These TgFBP1 and TgFBP2 isoenzymes are both located in the cytosol. They differ in their Vmax, Km for FBP and sensitivity to AMP. TgFBP2 exhibited a higher specific activity, lower substrate affinity and a markedly lower susceptibility to AMP inhibition. TgFBP1 only has minimal demonstrable FBPase activity *in vivo* in tachyzoites. Knockdown studies have shown that TgFBP2 is essential for growth under glucose-limiting and glucose-replete conditions and for acute virulence in mice. Rescue of the deficiency was possible with an active FBPase, but not by overexpression of TgFBP1, and was associated with global changes in central carbon metabolism and glycosylation pathways. The constitutive expression of TgFBP2, also in the presence of glucose, may have as function to increase the capacity of these parasites to rapidly adapt to changing nutrient conditions in their host cells and thus expand their host cell range [402].

As with the TgFBP isoenzymes, tachyzoites also express two catalytically active PYK enzymes (PYK1 and PYK2) with distinct subcellular location and enzymatic properties [115,403]. PYK1 is in the cytosol, while PYK2 is found mainly within the apicoplast and mitochondrion [403,404]. The cytosolic PYK1 works in canonical glycolysis to convert PEP to pyruvate, which is then transported to the mitochondrion where it enters the Krebs cycle [405]. By contrast, PYK2 catalyses pyruvate formation in the apicoplast which in turn provides substrates for the synthesis of isoprenoids and fatty acids. PYK1 is necessary for parasite growth, while PYK2 is available for the lytic cycle. The absence of both PYK isoenzymes abrogates tachyzoite growth. Co-expression of the enzymes contributes to robust regulation of central carbon metabolism.

Something similar is likely to occur with the PGK isoenzymes of *T. cruzi*. These isoenzymes would allow the use of different carbon sources. Results from biochemical studies on *T. cruzi* suggested that PGKA and B may act in the glycolytic direction, while PGKC does so in the gluconeogenic one [25]. At least in the case of *T. cruzi*, the presence of three active PGK isoenzymes (PGKA, PGKC and PAS-PGK) within glycosomes, but with different affinities for the substrates, could suggest a mechanism for reducing the effects of variations in the environment, thus increasing cellular robustness. The differences in the kinetic characteristics of the cytosolic and glycosomal PGK isoenzymes provide the parasite with greater metabolic buffer capacity. Affinity differences among isoforms will increase the flexibility in compensating larger changes in

the environment [400]. The expression pattern of TgFBP2 in *T. gondii* is reminiscent of the expression pattern of the PGKC in *T. cruzi*. Although this latter isoenzyme is constitutively expressed during all phases of parasite growth, its expression is increased upon entering the stationary phase, when glucose availability becomes low. The pattern of expression and negative regulation by ATP that have been reported for the *T. cruzi* PGKC [124] suggests that this isoenzyme would have a similar function as TgFBP2 (i.e. to sustain an important gluconeogenic flux under glucose-starvation conditions). As for the PAS–PGK of *T. cruzi*, it is, with the present knowledge, difficult to attribute a specific function to its PAS domain. It could serve as a receptor for sensing nutritional changes (such as levels of glucose or other carbon sources in the different cell types where the parasite resides as amastigote, such as heart cells or muscle cells, or extracellularly as trypomastigote) that would modulate the fluxes through the glycolytic and/or gluconeogenic pathways in the parasite. The expression of PGK isoenzymes with different kinetics and mechanism of regulation as the PGKC may then enable rapid fine-tuning of the parasite's carbon metabolism over a wide range of nutrient conditions in the host and possibly contribute to the ability of these parasites to robustly proliferate within many different host cell compartments.

This idea could be supported by findings obtained through theoretical studies that provide evidence that the evolution of organisms in fluctuating environments leads to the development of robustness in metabolic networks, while robustness is lower in organisms that live in stable environments [406]. This could be an explanation for the difference in the number of PGK isoenzymes between the endosymbiotic *Perkinsela* sp. and other kinetoplastids with complex, parasitic life cycles such as *T. brucei*, *T. cruzi* and *Leishmania* spp. (table 3). Additionally, robustness of glucose metabolism may also be necessary for *P. falciparum*; the rapid growth and proliferation of the blood stage of this apicomplexan parasite is supported by robust metabolism of glucose as a source of carbon and energy via glycolysis [407]. Glycolysis would not only be necessary to meet the physiological demand of ATP, but also to support other physiological needs, because glycolytic intermediates are also directed to anabolic reactions. Similarly, in *T. cruzi*, the glycolytic enzymes in glycosomes and cytosol are interwoven in an elaborate metabolic network of catabolic and anabolic processes [340].

In other organisms such as plants of the genus *Alloteropsis*, gene dosage is associated with the emergence and establishment of the C$_4$ carbon fixation pathway, with subsequent neofunctionalization [408]. C$_4$ metabolism is strongly favoured by selection only under specific environmental conditions, such as drought, high temperatures and intense light. Different models suggested that a weak C$_4$ cycle can emerge from enzymes that have not been adapted to the C$_4$ catalytic context; gene duplications increasing the transcript abundance of C$_4$-related genes in plants with a weak C$_4$ cycle would then subsequently increase the strength of the pathway, which is predicted to boost carbon assimilation and fitness [409]. In other plants, such as *Arabidopsis thaliana*, duplicate genes are also associated with an increase in metabolic flux, suggesting that increased gene dosage results in increased metabolic activity [410]. Studies of *Leishmania* have shown duplication and differential expression of a large variety of genes, such as those for several transporters (including an inosine/guanosine transporter, two putative nucleoside transporters, two amino acid transporters, a glucose transporter and a pteridine transporter), proteins associated with drug resistance such as proteins involved in redox pathways, trypanothione synthetase and some hypothetical proteins [411,412]. Gene dosing is an adaptive trait that confers phenotypic plasticity in natural populations of *Leishmania* by rapidly regulating up or down the expression of transporter proteins to cope with the effects of environmental stress [411].

The duplication of *pkg* genes in *T. cruzi* may be related to gene dosage, since the creation of the glycosomal PGKA, although having relatively low catalytic activity [25], could have had the function of contributing to ATP generation inside the glycosomes and at the same time providing a mechanism for adaptation to environments with abundant glucose such as the bloodstream. The presence of a PGKA may even have become an adaptive advantage for the epimastigote form of *T. cruzi* (experiencing low glucose availability in the insect), by providing the parasite with a tool to be prepared for entering the mammalian bloodstream. PGKA could act within an adaptation mechanism to changes in glucose availability, somewhat similar to the role exerted by some glycolytic isoenzymes in yeasts [413]. The difference in location of the trypanosomatid PGK isoenzymes could have been the result of subcellular relocation upon the duplication process. In *S. cerevisiae*, too, proteins involved in catabolic processes (such as glycolysis) frequently developed new compartments ('neolocalization') or sublocalization patterns after duplication [414]. In the case of the PGK of *T. cruzi*, after duplication, a process of simultaneous neo- and sublocalization of the product of the *pgkA* gene has probably occurred. Additionally, the PGKA evolved new functional roles through changes of its kinetic properties, expression levels and probably, its interaction partners. The unique insert of 80 residues found in the N-terminal domain of PGKA of different kinetoplastids (salivarian and stercorian trypanosomes, *Crithidia* and *Leptomonas*) could be relevant for this latter property. This insert must have been acquired in a duplicated gene in a common ancestor. An analysis of these PGKAs, using different server predictors for segments of structural disorder, revealed several of such segments within the insert (W. Q. *et al.* 2020, unpublished results). Elements with a disordered structure within a protein are often involved in interactions with other proteins. In the case of *T. cruzi* PGKA, the insert may be responsible for the observed association of this isoenzyme with the glycosomal membrane [124], through electrostatic interactions of amino acids in the disordered segments with the membrane phospholipid head groups or with membrane proteins. Alternatively, PGKA may interact with other glycosomal proteins, for example, to channel its product (3PGA or 1,3BPGA), separate from the metabolite flux within the main glycolytic/gluconeogenic pathway complexes, to an enzyme that depends on it. In this respect, the finding of a putative glycerate kinase within the *T. cruzi* glycosomal proteome is suggestive [104]. However, this protein has not been found in the *T. brucei* glycosomal proteome, nor its gene identified in the genome databases of other trypanosomatids containing a PGKA. The lower level of conservation of the primary structure of the PGKA insert than the remainder of the enzyme (see §8.2.3.3.1) and the variation in the positions of the elements of disordered structure in the insert could indicate that the enzyme may be involved in different interactions in the other trypanosomatids. Future research is necessary to reveal the function(s) of this peculiar enzyme.

Finally, the 'PGKL' is another trypanosomatid protein that resulted from a duplication of a *pgk* gene, followed by recombination with other genes to produce a gene encoding a multidomain protein, PGK-CNB-HTH, and with between 1 and 4 predicted transmembrane segments at its C-terminal side. However, our analysis, and that by [60], predicted that the PGK is not active; it is considered a 'dead' enzyme. Also the HTH domain has evolved such that it is doubtful whether it retained DNA-binding capacity. However, and intriguingly, the protein is present in many trypanosomatids, thus must already have been present in the common ancestor, and has been relatively well conserved with an identity over 50%. Moreover, it has been shown to be expressed in several of these organisms. In *T. brucei*, the protein has been located in the flagellum, but from the sequences is predicted that it may be located elsewhere in other organisms. Together, these data suggest that the protein plays a role in the biology of these organisms, but its function remains enigmatic. In addition, genes encoding related proteins, but in which a potential HTH domain sequence could not be recognized, are also found in many kinetoplastids (table 3; electronic supplementary material, table SI). Whether they have been derived from *pgkL* genes or evolved independently is currently unknown.

In humans, the presence of duplicate genes contributes to attenuate the phenotypic consequences of deleterious human mutations, representing functional support [415]. Most duplicated genes are associated with metabolic and cell physiological processes, such as catabolism, anabolism and responses to stimuli [415]. Alterations in the gene copy number of metabolic enzymes are an omnipresent characteristic in all human cancers, and which also appear to have an impact on clinical aggressiveness and responsiveness to drugs [416].

10.2.

The neo- and/or subfunctionalization of PGKs could allow the adaptation of these protists to the different environmental changes to which they are exposed during their life cycle. The infection of alternative hosts during the life cycle could probably have been the selective force for the evolution of PGK isoforms, while gene duplication could have provided the gross genetic material. This proposition provides support for an alternative to Ohno's hypothesis. A *pgk* gene encodes a protein product that fulfilled a catalytic function; however, after the duplication process, the two child genes were able to specialize for additional functions to the specific one. This may have happened with the products of the *pgkA/B/C* gene array. These three isoenzymes have been shown to possess the ability to catalyse the reaction within the glycolytic and/or gluconeogenic pathway in *T. cruzi* [25,124]. In turn, another possibility is that the division of the functions of the ancestral *pgk* gene can have led to a neofunctionalization process. In this process, a series of changes in the coding region of a child gene (substitutions of non-synonymous nucleotides, addition of coding sequences for a domain and/or sorting signal for a subcellular compartment) that adapted its product to a new specialized function could have been selectively advantageous and this would then have allowed the fixation of the gene [417,418]. It is further possible that the appearance of one or more additional domains in a PGK isoenzyme has provided an ancestral trypanosomatid with an enzyme, the PAS–PGK, with new and specialized functions within metabolism. Functions such as adaptation to environmental changes (such as internal oxygen sensor and carbon sources) and phosphorylation of glycosomal proteins have been suggested as possibilities for the PAS-PGK of *T. cruzi* [4]. It is also likely to have functions similar to those reported for *Leishmania*, where it has been shown that deletion of the PAS–PGK gene results in a 15% decreased glycolytic flux in promastigotes, as measured by lactate production, as well as an increase in the number of autophagosomes and an increased cell death [28]. Gene innovation introduced by duplication seems thus to play an important role in these parasites, similar to what has been proposed for other parasitic protists [380].

11. Conclusion and possible trends of future research

PGK is an enzyme that is distributed and highly conserved in Archaea, Bacteria and eukaryotes. Its main function is related to carbohydrate metabolism, specifically in glycolysis and gluconeogenesis, but the enzyme can also participate in a wide variety of other biological processes such as cell invasion, angiogenesis and tumour growth, flagellar movement, viral replication, induction of autophagy and cell death, thus fulfilling additional functions not related to its metabolic one. In many eukaryotic organisms, different PGK isoenzymes have been identified. The physiological significance of *pgk* gene duplications and the presence of different PGK isoenzymes are poorly understood. Our analyses of the genomes of most kinetoplastid parasites showed that *pgk* genes are often organized as tandem arrays and concentrated in specific chromosomes. Other thus far unreported *pgk* genes and *pseudogenes* were identified in some kinetoplastids. Several of these additionally identified *pgk* genes code for multidomain PGK enzymes and others for 'dead' PGK enzymes. So far, most studies of kinetoplastid PGK isoenzymes have focused almost exclusively on their biochemical and structural characterization. For that reason, we set out to address a little more the possible physiological significance of the presence of multiple PGKs in kinetoplastids and other protists. Due to the diversity of functions that PGK can fulfil, different evolutionary scenarios and functions are proposed to understand the physiological significance of PGK isoforms in kinetoplastids. Specifically, in the case of *T. cruzi*, a function beyond glycolysis/gluconeogenesis is proposed for some of the non-canonical PGKs. Providing metabolic plasticity is considered one of the main functions of the presence of isoenzymes. This plasticity would be crucial for adaptation to different environmental conditions and successful parasitism of this kinetoplastid in the extracellular and intracellular niches of its different hosts and the various host cell types that it parasitizes. Future molecular and biochemical studies of PGK isoforms of which the genes were identified in this study could give a clearer view of their physiological importance in *T. cruzi*, and by inference maybe also in related human parasites. Additionally, this analysis could also shed light on the possibility that some isoforms could serve for explorative studies for their use as drug targets in these parasites.

References

1. Cicenas J, Zalyte E, Bairoch A, Gaudet P. 2018 Kinases and cancer. *Cancers* **10**, 63. (doi:10.3390/cancers10030063)

2. Rauch J, Volinsky N, Romano D, Kolch W. 2011 The secret life of kinases: functions beyond catalysis. *Cell Commun. Signal.* **9**, 23. (doi:10.1186/1478-811X-9-23)

3. Parsons M, Worthey EA, Ward PN, Mottram JC. 2005 Comparative analysis of the kinomes of three pathogenic trypanosomatids: *Leishmania major*, *Trypanosoma brucei* and *Trypanosoma cruzi*. *BMC Genomics* **6**, 127. (doi:10.1186/1471-2164-6-127)

4. Rojas-Pirela M, Rigden DJ, Michels PA, Cáceres AJ, Concepción JL, Quiñones W. 2018 Structure and function of Per-ARNT-Sim domains and their possible role in the life-cycle biology of *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* **219**, 52–66. (doi:10.1016/j.molbiopara.2017.11.002)

5. Jackson AP *et al.* 2016 Kinetoplastid phylogenomics reveals the evolutionary innovations associated with the origins of parasitism. *Curr. Biol.* **26**, 161–172. (doi:10.1016/j.cub.2015.11.055)

6. Adl SM *et al.* 2019 Revisions to the classification, nomenclature, and diversity of eukaryotes. *J. Eukaryot. Microbiol.* **66**, 4–119. (doi:10.1111/jeu.12691)

7. Saldivia M *et al.* 2020 Targeting the trypanosome kinetochore with CLK1 protein kinase inhibitors. *Nat. Microbiol.* **5**, 1207–1216. (doi:10.1038/s41564-020-0745-6)

8. Stortz JA *et al.* 2017 Genome-wide and protein kinase-focused RNAi screens reveal conserved and novel damage response pathways in *Trypanosoma brucei*. *PLoS Pathog.* **13**, e1006477. (doi:10.1371/journal.ppat.1006477)

9. Ellis J, Sarkar M, Hendriks E, Matthews K. 2004 A novel ERK-like, CRK-like protein kinase that modulates growth in *Trypanosoma brucei* via an autoregulatory C-terminal extension. *Mol. Microbiol.* **53**, 1487–1499. (doi:10.1111/j.1365-2958.2004.04218.x)

10. Portal D *et al.* 2003 Trypanosoma cruzi TcSRPK, the first protozoan member of the SRPK family, is biochemically and functionally conserved with metazoan SR protein-specific kinases. *Mol. Biochem. Parasitol.* **127**, 9–21. (doi:10.1016/s0166-6851(02)00299-2)

11. Parsons M, Valentine M, Carter V. 1993 Protein kinases in divergent eukaryotes: identification of protein kinase activities regulated during trypanosome development. *Proc. Natl Acad. Sci. USA* **90**, 2656–2660. (doi:10.1073/pnas.90.7.2656)

12. Deshmukh K, Anamika K, Srinivasan N. 2010 Evolution of domain combinations in protein kinases and its implications for functional diversity. *Prog. Biophys. Mol. Biol.* **102**, 1–15. (doi:10.1016/j.pbiomolbio.2009.12.009)

13. Lu Z, Hunter T. 2018 Metabolic kinases moonlighting as protein kinases. *Trends Biochem. Sci.* **43**, 301–310. (doi:10.1016/j.tibs.2018.01.006)

14. Li X, Zheng Y, Lu Z. 2016 PGK1 is a new member of the protein kinome. *Cell Cycle Georget. Tex* **15**, 1803–1804. (doi:10.1080/15384101.2016.1179037)

15. Yang W, Lu Z. 2015 Pyruvate kinase M2 at a glance. *J. Cell Sci.* **128**, 1655–1660. (doi:10.1242/jcs.166629)

16. Lukeš J, Skalický T, Týč J, Votýpka J, Yurchenko V. 2014 Evolution of parasitism in kinetoplastid flagellates. *Mol. Biochem. Parasitol.* **195**, 115–122. (doi:10.1016/j.molbiopara.2014.05.007)

17. Palmai Z, Seifert C, Gräter F, Balog E. 2014 An allosteric signaling pathway of human 3-phosphoglycerate kinase from force distribution analysis. *PLoS Comput. Biol.* **10**, e1003444. (doi:10.1371/journal.pcbi.1003444)

18. Bücher T. 1947 Über ein phosphatübertragendes gärungsferment. *Biochim. Biophys. Acta* **1**, 292–314. (doi:10.1016/0006-3002(47)90143-1)

19. Banks RD, Blake CC, Evans PR, Haser R, Rice DW, Hardy GW, Merrett M, Phillips AW. 1979 Sequence, structure and activity of phosphoglycerate kinase: a possible hinge-bending enzyme. *Nature* **279**, 773–777. (doi:10.1038/279773a0)

20. Grall M, Srivastava IK, Schmidt M, Garcia AM, Mauël J, Perrin LH. 1992 *Plasmodium falciparum*: identification and purification of the phosphoglycerate kinase of the malaria parasite. *Exp. Parasitol.* **75**, 10–18. (doi:10.1016/0014-4894(92)90117-s)

21. Encalada R, Rojo-Domínguez A, Rodríguez-Zavala JS, Pardo JP, Quezada H, Moreno-Sánchez R, Saavedra E. 2009 Molecular basis of the unusual catalytic preference for GDP/GTP in Entamoeba histolytica 3-phosphoglycerate kinase. *FEBS J.* **276**, 2037–2047. (doi:10.1111/j.1742-4658.2009.06939.x)

22. Misset O, Opperdoes FR. 1987 The phosphoglycerate kinases from *Trypanosoma brucei*. A comparison of the glycosomal and the cytosolic isoenzymes and their sensitivity towards suramin. *Eur. J. Biochem.* **162**, 493–500. (doi:10.1111/j.1432-1033.1987.tb10667.x)

23. Gibson WC, Swinkels BW, Borst P. 1988 Post-transcriptional control of the differential expression of phosphoglycerate kinase genes in *Trypanosoma brucei*. *J. Mol. Biol.* **201**, 315–325. (doi:10.1016/0022-2836(88)90140-4)

24. Concepción JL, Adjé CA, Quiñones W, Chevalier N, Dubourdieu M, Michels PA. 2001 The expression and intracellular distribution of phosphoglycerate kinase isoenzymes in *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* **118**, 111–121. (doi:10.1016/s0166-6851(01)00381-4)

25. Barros-Álvarez X, Cáceres AJ, Michels PAM, Concepción JL, Quiñones W. 2014 The phosphoglycerate kinase isoenzymes have distinct roles in the regulation of carbohydrate metabolism in *Trypanosoma cruzi*. *Exp. Parasitol.* **143**, 39–47. (doi:10.1016/j.exppara.2014.05.010)

26. Rivero LA, Concepción JL, Quintero-Troconis E, Quiñones W, Michels PAM, Acosta H. 2016 *Trypanosoma evansi* contains two auxiliary enzymes of glycolytic metabolism: phosphoenolpyruvate carboxykinase and pyruvate phosphate dikinase. *Exp. Parasitol.* **165**, 7–15. (doi:10.1016/j.exppara.2016.03.003)

27. Villafraz O, Rondón-Mercado R, Cáceres AJ, Concepción JL, Quiñones W. 2018 Molecular and biochemical characterization of natural and recombinant phosphoglycerate kinase B from *Trypanosoma rangeli*. *Exp. Parasitol.* **187**, 42–48. (doi:10.1016/j.exppara.2018.03.009)

28. Adhikari A, Biswas S, Mukherjee A, Das S, Adak S. 2019 PAS domain-containing phosphoglycerate kinase deficiency in *Leishmania major* results in increased autophagosome formation and cell death. *Biochem. J.* **476**, 1303–1321. (doi:10.1042/BCJ20190041)

29. Bücher T. 1942 Isolierung und Kristallisation eines phosphatübertragenden Gärungsferments. *Naturwissenschaften* **30**, 756–757. (doi:10.1007/BF01488658)

30. Blake CC, Evans PR, Scopes RK. 1972 Structure of horse-muscle phosphoglycerate kinase at 6 angstrom resolution. *Nat. New Biol.* **235**, 195–198. (doi:10.1038/newbio235195a0)

31. Wendell PL, Bryant TN, Watson HC. 1972 Low resolution structure of yeast phosphoglycerate kinase. *Nat. New Biol.* **240**, 134–136. (doi:10.1038/newbio240134a0)

32. Bücher T. 1989 Commentary on 'Uber ein phosphatübertragendes Gärungsferment'. *Biochim. Biophys. Acta* **1000**, 223–227. (doi:10.1016/S0006-3002(89)80020-4)

33. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000 The protein data bank. *Nucleic Acids Res.* **28**, 235–242. (doi:10.1093/nar/28.1.235)

34. Auerbach G, Huber R, Grättinger M, Zaiss K, Schurig H, Jaenicke R, Jacob U. 1997 Closed structure of phosphoglycerate kinase from *Thermotoga maritima* reveals the catalytic mechanism and determinants of thermal stability. *Struct. Lond. Engl. 1993* **5**, 1475–1483. (doi:10.1016/s0969-2126(97)00297-9)

35. Fulde M *et al*. 2014 Pneumococcal phosphoglycerate kinase interacts with plasminogen and its tissue activator. *Thromb. Haemost.* **111**, 401–416. (doi:10.1160/TH13-05-0421)

36. Zheng H *et al*. 2012 Crystal structures of putative phosphoglycerate kinases from *B. anthracis* and *C. jejuni. J. Struct. Funct. Genomics* **13**, 15–26. (doi:10.1007/s10969-012-9131-9)

37. Lee JH, Im YJ, Bae J, Kim D, Kim M-K, Kang GB, Lee D-S, Eom SH. 2006 Crystal structure of *Thermus caldophilus* phosphoglycerate kinase in the open conformation. *Biochem. Biophys. Res. Commun.* **350**, 1044–1049. (doi:10.1016/j.bbrc.2006.09.151)

38. Young TA, Skordalakes E, Marqusee S. 2007 Comparison of proteolytic susceptibility in phosphoglycerate kinases from yeast and *E. coli*: modulation of conformational ensembles without altering structure or stability. *J. Mol. Biol.* **368**, 1438–1447. (doi:10.1016/j.jmb.2007.02.077)

39. Mandelman D, Ballut L, Wolff DA, Feller G, Gerday C, Haser R, Aghajari N. 2019 Structural determinants increasing flexibility confer cold adaptation in psychrophilic phosphoglycerate kinase. *Extrem. Life Extreme Cond.* **23**, 495–506. (doi:10.1007/s00792-019-01102-x)

40. Franklin MC *et al*. 2015 Structural genomics for drug design against the pathogen *Coxiella burnetii*. *Proteins* **83**, 2124–2136. (doi:10.1002/prot.24841)

41. Davies GJ, Gamblin SJ, Littlechild JA, Dauter Z, Wilson KS, Watson HC. 1994 Structure of the ADP complex of the 3-phosphoglycerate kinase from *Bacillus stearothermophilus* at 1.65 A. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 202–209. (doi:10.1107/S0907444993011138)

42. Zerrad L, Merli A, Schröder GF, Varga A, Gráczer É, Pernot P, Round A, Vas M, Bowler MW. 2011 A spring-loaded release mechanism regulates domain movement and catalysis in phosphoglycerate kinase. *J. Biol. Chem.* **286**, 14 040–14 048. (doi:10.1074/jbc.M110.206813)

43. Watson HC *et al*. 1982 Sequence and structure of yeast phosphoglycerate kinase. *EMBO J.* **1**, 1635–1640. (doi:10.1002/j.1460-2075.1982.tb01366.x)

44. Chen X *et al*. 2015 Terazosin activates Pgk1 and Hsp90 to promote stress resistance. *Nat. Chem. Biol.* **11**, 19–25. (doi:10.1038/nchembio.1657)

45. Sawyer GM, Monzingo AF, Poteet EC, O'Brien DA, Robertus JD. 2008 X-ray analysis of phosphoglycerate kinase 2, a sperm-specific isoform from *Mus musculus. Proteins* **71**, 1134–1144. (doi:10.1002/prot.21801)

46. Szilágyi AN, Ghosh M, Garman E, Vas M. 2001 A 1.8 A resolution structure of pig muscle 3-phosphoglycerate kinase with bound MgADP and 3-phosphoglycerate in open conformation: new insight into the role of the nucleotide in domain closure. *J. Mol. Biol.* **306**, 499–511. (doi:10.1006/jmbi.2000.4294)

47. Smith CD, Chattopadhyay D, Pal B. 2011 Crystal structure of *Plasmodium falciparum* phosphoglycerate kinase: evidence for anion binding in the basic patch. *Biochem. Biophys. Res. Commun.* **412**, 203–206. (doi:10.1016/j.bbrc.2011.07.045)

48. Bernstein BE, Michels PA, Hol WG. 1997 Synergistic effects of substrate-induced conformational changes in phosphoglycerate kinase activation. *Nature* **385**, 275–278. (doi:10.1038/385275a0)

49. Rice DW. 1981 The use of phase combination in the refinement of phosphoglycerate kinase at 2.5Å resolution. *Acta Crystallogr. A* **37**, 491–500. (doi:10.1107/S0567739481001186)

50. Watson HC, Littlechild JA. 1990 Isoenzymes of phosphoglycerate kinase: evolutionary conservation of the structure of this glycolytic enzyme. *Biochem. Soc. Trans.* **18**, 187–190. (doi:10.1042/bst0180187)

51. Hess D, Krüger K, Knappik A, Palm P, Hensel R. 1995 Dimeric 3-phosphoglycerate kinases from hyperthermophilic Archaea. *Eur. J. Biochem.* **233**, 227–237. (doi:10.1111/j.1432-1033.1995.227_1.x)

52. Jones CE, Fleming TM, Cowan DA, Littlechild JA, Piper PW. 1995 The phosphoglycerate kinase and glyceraldehyde-3-phosphate dehydrogenase genes from the thermophilic archaeon *Sulfolobus solfataricus* overlap by 8-bp. Isolation, sequencing of the genes and expression in *Escherichia coli. Eur. J. Biochem.* **233**, 800–808. (doi:10.1111/j.1432-1033.1995.800_3.x)

53. He Y, Luo Y, Zhang D, Wang X, Zhang P, Li H, Ejaz S, Liang S. 2019 PGK1-mediated cancer progression and drug resistance. *Am. J. Cancer Res.* **9**, 2280–2302.

54. Bosco MB, Aleanzi MC, Iglesias AÁ. 2012 Plastidic phosphoglycerate kinase from *Phaeodactylum tricornutum*: on the critical role of cysteine residues for the enzyme function. *Protist* **163**, 188–203. (doi:10.1016/j.protis.2011.07.001)

55. Tsukamoto Y, Fukushima Y, Hara S, Hisabori T. 2013 Redox control of the activity of phosphoglycerate kinase in *Synechocystis* sp. PCC6803. *Plant Cell Physiol.* **54**, 484–491. (doi:10.1093/pcp/pct002)

56. Jindal HK, Vishwanatha JK. 1990 Functional identity of a primer recognition protein as phosphoglycerate kinase. *J. Biol. Chem.* **265**, 6540–6543.

57. Lay AJ, Jiang XM, Kisker O, Flynn E, Underwood A, Condron R, Hogg PJ. 2000 Phosphoglycerate kinase acts in tumour angiogenesis as a disulphide reductase. *Nature* **408**, 869–873. (doi:10.1038/35048596)

58. Kinnby B, Booth NA, Svensäter G. 2008 Plasminogen binding by oral streptococci from dental plaque and inflammatory lesions. *Microbiol. Read. Engl.* **154**, 924–931. (doi:10.1099/mic.0.2007/013235-0)

59. Burnham C-AD, Shokoples SE, Tyrrell GJ. 2005 Phosphoglycerate kinase inhibits epithelial cell invasion by group B streptococci. *Microb. Pathog.* **38**, 189–200. (doi:10.1016/j.micpath.2005.02.002)

60. Brown RWB, Collingridge PW, Gull K, Rigden DJ, Ginger ML. 2014 Evidence for loss of a partial flagellar glycolytic pathway during trypanosomatid evolution. *PLoS ONE* **9**, e103026. (doi:10.1371/journal.pone.0103026)

61. Prasanth KR, Chuang C, Nagy PD. 2017 Co-opting ATP-generating glycolytic enzyme PGK1 phosphoglycerate kinase facilitates the assembly of viral replicase complexes. *PLoS Pathog.* **13**, e1006689. (doi:10.1371/journal.ppat.1006689)

62. Vas M, Varga A, Gráczer E. 2010 Insight into the mechanism of domain movements and their role in enzyme function: example of 3-phosphoglycerate kinase. *Curr. Protein Pept. Sci.* **11**, 118–147. (doi:10.2174/138920310790848403)

63. Suzuki K, Imahori K. 1974 Phosphoglycerate kinase of *Bacillus stearothermophilus. J. Biochem. (Tokyo)* **76**, 771–782.

64. Reddy GK, Wendisch VF. 2014 Characterization of 3-phosphoglycerate kinase from *Corynebacterium glutamicum* and its impact on amino acid production. *BMC Microbiol.* **14**, 54. (doi:10.1186/1471-2180-14-54)

65. Bentahir M, Feller G, Aittaleb M, Lamotte-Brasseur J, Himri T, Chessa JP, Gerday C. 2000 Structural, kinetic, and calorimetric characterization of the cold-active phosphoglycerate kinase from the antarctic *Pseudomonas* sp. TACII18. *J. Biol. Chem.* **275**, 11 147–11 153. (doi:10.1074/jbc.275.15.11147)

66. Nojima H, Oshima T, Noda H. 1979 Purification and properties of phosphoglycerate kinase from *Thermus thermophilus* strain HB8. *J. Biochem. (Tokyo)* **85**, 1509–1517. (doi:10.1093/oxfordjournals.jbchem.a132480)

67. Pawluk A, Scopes RK, Griffiths-Smith K. 1986 Isolation and properties of the glycolytic enzymes from *Zymomonas mobilis*: the five enzymes from glyceraldehyde-3-phosphate dehydrogenase through to pyruvate kinase. *Biochem. J.* **238**, 275–281. (doi:10.1042/bj2380275)

68. Kaushik S, Krishnarjuna B, Raghothama S, Aggarwal S, Raghunathan V, Ganjiwale A. 2012 Theoretical and in vitro studies of a C-terminal peptide from PGKC of *Leishmania mexicana mexicana. Mol. Biochem. Parasitol.* **185**, 27–35. (doi:10.1016/j.molbiopara.2012.06.004)

69. Zomer AW, Allert S, Chevalier N, Callens M, Opperdoes FR, Michels PA. 1998 Purification and characterisation of the phosphoglycerate kinase isoenzymes of *Trypanosoma brucei* expressed in *Escherichia coli. Biochim. Biophys. Acta* **1386**, 179–188. (doi:10.1016/s0167-4838(98)00095-8)

70. Kuntz GW, Krietsch WK. 1982 Phosphoglycerate kinase from spinach, blue-green algae, and yeast. *Methods Enzymol.* **90**(Pt E), 110–114. (doi:10.1016/s0076-6879(82)90115-x)

71. Krietsch WK, Bücher T. 1970 3-phosphoglycerate kinase from rabbit sceletal muscle and yeast.

Eur. J. Biochem. **17**, 568–580. (doi:10.1111/j.1432-1033.1970.tb01202.x)

72. Chen SH, Donahue RP, Scott CR. 1976 Characterization of phosphoglycerate kinase from human spermatozoa. *Fertil. Steril.* **27**, 699–701.

73. Blake CC, Rice DW. 1981 Phosphoglycerate kinase. *Phil. Trans. R. Soc. Lond. B* **293**, 93–104. (doi:10.1098/rstb.1981.0063)

74. Blake CC, Evans PR. 1974 Structure of horse muscle phosphoglycerate kinase: some results on the chain conformation, substrate binding and evolution of the molecule from a 3 angstrom Fourier map. *J. Mol. Biol.* **84**, 585–601. (doi:10.1016/0022-2836(74)90118-1)

75. May A, Vas M, Harlos K, Blake C. 1996 2.0 A resolution structure of a ternary complex of pig muscle phosphoglycerate kinase containing 3-phospho-D-glycerate and the nucleotide Mn adenylylimidodiphosphate. *Proteins* **24**, 292–303. (doi:10.1002/(SICI)1097-0134(199603)24:3<292::AID-PROT2>3.0.CO;2-J)

76. Fields PA, Somero GN. 1998 Hot spots in cold adaptation: localized increases in conformational flexibility in lactate dehydrogenase A4 orthologs of Antarctic notothenioid fishes. *Proc. Natl Acad. Sci. USA* **95**, 11 476–11 481. (doi:10.1073/pnas.95.19.11476)

77. Harlos K, Vas M, Blake CF. 1992 Crystal structure of the binary complex of pig muscle phosphoglycerate kinase and its substrate 3-phospho-D-glycerate. *Proteins* **12**, 133–144. (doi:10.1002/prot.340120207)

78. Scopes RK. 1978 Binding of substrates and other anions to yeast phosphoglycerate kinase. *Eur. J. Biochem.* **91**, 119–129. (doi:10.1111/j.1432-1033.1978.tb20944.x)

79. Cliff MJ *et al*. 2010 Transition state analogue structures of human phosphoglycerate kinase establish the importance of charge balance in catalysis. *J. Am. Chem. Soc.* **132**, 6507–6516. (doi:10.1021/ja100974t)

80. Swinkels BW, Loiseau A, Opperdoes FR, Borst P. 1992 A phosphoglycerate kinase-related gene conserved between *Trypanosoma brucei* and *Crithidia fasciculata*. *Mol. Biochem. Parasitol.* **50**, 69–78. (doi:10.1016/0166-6851(92)90245-f)

81. McKoy G, Badal M, Prescott Q, Lux H, Hart DT. 1997 Characterisation of phosphoglycerate kinase genes in *Leishmania major* and evidence for the absence of a third closely related gene or isoenzyme. *Mol. Biochem. Parasitol.* **90**, 169–181. (doi:10.1016/s0166-6851(97)00172-2)

82. Dhar A, Samiotakis A, Ebbinghaus S, Nienhaus L, Homouz D, Gruebele M, Cheung MS. 2010 Structure, function, and folding of phosphoglycerate kinase are strongly perturbed by macromolecular crowding. *Proc. Natl Acad. Sci. USA* **107**, 17 586–17 591. (doi:10.1073/pnas.1006760107)

83. Bowler MW. 2013 Conformational dynamics in phosphoglycerate kinase, an open and shut case? *FEBS Lett.* **587**, 1878–1883. (doi:10.1016/j.febslet.2013.05.012)

84. Schurig H, Beaucamp N, Ostendorp R, Jaenicke R, Adler E, Knowles JR. 1995 Phosphoglycerate kinase and triosephosphate isomerase from the hyperthermophilic bacterium *Thermotoga maritima* form a covalent bifunctional enzyme complex. *EMBO J.* **14**, 442–451. (doi:10.1002/j.1460-2075.1995.tb07020.x)

85. Stover NA, Punia RS, Bowen MS, Dolins SB, Clark TG. 2012 Tetrahymena Genome Database Wiki: a community-maintained model organism database. *Database J. Biol. Databases Curation* **2012**, bas007. (doi:10.1093/database/bas007)

86. Pruitt KD, Tatusova T, Maglott DR. 2005 NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33**, D501–D504. (doi:10.1093/nar/gki025)

87. UniProt Consortium T. 2018 UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **46**, 2699. (doi:10.1093/nar/gky092)

88. Kreppel L, Fey P, Gaudet P, Just E, Kibbe WA, Chisholm RL, Kimmel AR. 2004 dictyBase: a new Dictyostelium discoideum genome database. *Nucleic Acids Res.* **32**, D332–D333. (doi:10.1093/nar/gkh138)

89. Aurrecoechea C *et al*. 2011 AmoebaDB and MicrosporidiaDB: functional genomic resources for Amoebozoa and Microsporidia species. *Nucleic Acids Res.* **39**, D612–D619. (doi:10.1093/nar/gkq1006)

90. Aurrecoechea C *et al*. 2009 GiardiaDB and TrichDB: integrated genomic resources for the eukaryotic protist pathogens *Giardia lamblia* and *Trichomonas vaginalis*. *Nucleic Acids Res.* **37**, D526–D530. (doi:10.1093/nar/gkn631)

91. Gajria B *et al*. 2008 ToxoDB: an integrated *Toxoplasma gondii* database resource. *Nucleic Acids Res.* **36**, D553–D556. (doi:10.1093/nar/gkm981)

92. Aurrecoechea C *et al*. 2009 PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res.* **37**, D539–D543. (doi:10.1093/nar/gkn814)

93. Heiges M *et al*. 2006 CryptoDB: a Cryptosporidium bioinformatics resource update. *Nucleic Acids Res.* **34**, D419–D422. (doi:10.1093/nar/gkj078)

94. Aslett M *et al*. 2010 TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res.* **38**, D457–D462. (doi:10.1093/nar/gkp851)

95. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R. 2013 Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res.* **41**, W597–W600. (doi:10.1093/nar/gkt376)

96. Marchler-Bauer A *et al*. 2015 CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43**, D222–D226. (doi:10.1093/nar/gku1221)

97. Finn RD *et al*. 2017 InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* **45**, D190–D199. (doi:10.1093/nar/gkw1107)

98. Letunic I, Bork P. 2018 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* **46**, D493–D496. (doi:10.1093/nar/gkx922)

99. Artimo P *et al*. 2012 ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* **40**, W597–W603. (doi:10.1093/nar/gks400)

100. Narasimhan G, Bu C, Gao Y, Wang X, Xu N, Mathee K. 2002 Mining protein sequences for motifs. *J. Comput. Biol.* **9**, 707–720. (doi:10.1089/106652702761034145)

101. Lin W-Z, Fang J-A, Xiao X, Chou K-C. 2011 iDNA-Prot: identification of DNA binding proteins using random forest with grey model. *PLoS ONE* **6**, e24756. (doi:10.1371/journal.pone.0024756)

102. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001 Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580. (doi:10.1006/jmbi.2000.4315)

103. Käll L, Krogh A, Sonnhammer ELL. 2007 Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* **35**, W429–W432. (doi:10.1093/nar/gkm256)

104. Acosta H, Burchmore R, Naula C, Gualdrón-López M, Quintero-Troconis E, Cáceres AJ, Michels PAM, Concepción JL, Quiñones W. 2019 Proteomic analysis of glycosomes from *Trypanosoma cruzi* epimastigotes. *Mol. Biochem. Parasitol.* **229**, 62–74. (doi:10.1016/j.molbiopara.2019.02.008)

105. Almagro Armenteros JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O. 2017 DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinforma. Oxf. Engl.* **33**, 3387–3395. (doi:10.1093/bioinformatics/btx431)

106. Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. 2007 WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* **35**, W585–W587. (doi:10.1093/nar/gkm259)

107. Chou K-C, Shen H-B. 2008 Cell-PLoc: a package of Web servers for predicting subcellular localization of proteins in various organisms. *Nat. Protoc.* **3**, 153–162. (doi:10.1038/nprot.2007.494)

108. Xiong J *et al*. 2013 Tetrahymena functional genomics database (TetraFGD): an integrated resource for Tetrahymena functional genomics. *Database J. Biol. Databases Curation* **2013**, bat008. (doi:10.1093/database/bat008)

109. Pruitt KD, Tatusova T, Maglott DR. 2007 NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D61–D65. (doi:10.1093/nar/gkl842)

110. Nowitzki U, Gelius-Dietrich G, Schwieger M, Henze K, Martin W. 2004 Chloroplast phosphoglycerate kinase from *Euglena gracilis*: endosymbiotic gene replacement going against the tide. *Eur. J. Biochem.* **271**, 4123–4131. (doi:10.1111/j.1432-1033.2004.04350.x)

111. Mesquita A, Calvo-Garrido J, Carilla-Latorre S, Escalante R. 2013 Monitoring autophagy in Dictyostelium. *Methods Mol. Biol.* **983**, 461–470. (doi:10.1007/978-1-62703-302-2_26)

112. Howe KL *et al*. 2020 Ensembl Genomes 2020-enabling non-vertebrate genomic research. *Nucleic Acids Res.* **48**, D689–D695. (doi:10.1093/nar/gkz890)

113. Fritz-Laylin LK, Cande WZ. 2010 Ancestral centriole and flagella proteins identified by analysis of

Naegleria differentiation. *J. Cell Sci.* **123**, 4024–4031. (doi:10.1242/jcs.077453)

114. Huang K-Y *et al.* 2014 Adaptive responses to glucose restriction enhance cell survival, antioxidant capability, and autophagy of the protozoan parasite *Trichomonas vaginalis. Biochim. Biophys. Acta* **1840**, 53–64. (doi:10.1016/j.bbagen.2013.08.008)

115. Fleige T, Fischer K, Ferguson DJP, Gross U, Bohne W. 2007 Carbohydrate metabolism in the *Toxoplasma gondii* apicoplast: localization of three glycolytic isoenzymes, the single pyruvate dehydrogenase complex, and a plastid phosphate translocator. *Eukaryot. Cell* **6**, 984–996. (doi:10.1128/EC. 00061-07)

116. Costa RM, Nogueira F, de Sousa KP, Vitorino R, Silva MS. 2013 Immunoproteomic analysis of *Plasmodium falciparum* antigens using sera from patients with clinical history of imported malaria. *Malar. J.* **12**, 100. (doi:10.1186/1475-2875-12-100)

117. Cornillot E *et al.* 2012 Sequencing of the smallest Apicomplexan genome from the human pathogen *Babesia microti. Nucleic Acids Res.* **40**, 9102–9114. (doi:10.1093/nar/gks700)

118. Morales J *et al.* 2016 Differential remodelling of peroxisome function underpins the environmental and metabolic adaptability of diplonemids and kinetoplastids. *Proc. R. Soc. B* **283**, 20160520. (doi:10.1098/rspb.2016.0520)

119. Osinga KA, Swinkels BW, Gibson WC, Borst P, Veeneman GH, Van Boom JH, Michels PA, Opperdoes FR. 1985 Topogenesis of microbody enzymes: a sequence comparison of the genes for the glycosomal (microbody) and cytosolic phosphoglycerate kinases of *Trypanosoma brucei. EMBO J.* **4**, 3811–3817. (doi:10. 1002/j.1460-2075.1985.tb04152.x)

120. Alexander K, Parsons M. 1991 A phosphoglycerate kinase-like molecule localized to glycosomal microbodies: evidence that the topogenic signal is not at the C-terminus. *Mol. Biochem. Parasitol.* **46**, 1–10. (doi:10.1016/0166-6851(91)90193-a)

121. Alexander K, Parsons M. 1993 Characterization of a divergent glycosomal microbody phosphoglycerate kinase from *Trypanosoma brucei. Mol. Biochem. Parasitol.* **60**, 265–272. (doi:10.1016/0166-6851(93)90137-m)

122. Parker HL, Hill T, Alexander K, Murphy NB, Fish WR, Parsons M. 1995 Three genes and two isozymes: gene conversion and the compartmentalization and expression of the phosphoglycerate kinases of *Trypanosoma (Nannomonas) congolense. Mol. Biochem. Parasitol.* **69**, 269–279. (doi:10.1016/ 0166-6851(94)00208-5)

123. Logan-Klumpler FJ *et al.* 2012 GeneDB—an annotation database for pathogens. *Nucleic Acids Res.* **40**, D98–108. (doi:10.1093/nar/ gkr1032)

124. Barros-Álvarez X, Cáceres AJ, Ruiz MT, Michels PAM, Concepción JL, Quiñones W. 2015 The glycosomal-membrane associated phosphoglycerate kinase isoenzyme A plays a role in sustaining the glucose flux in *Trypanosoma cruzi* epimastigotes. *Mol. Biochem. Parasitol.* **200**, 5–8. (doi:10.1016/j. molbiopara.2015.04.003)

125. Swinkels BW, Evers R, Borst P. 1988 The topogenic signal of the glycosomal (microbody) phosphoglycerate kinase of *Crithidia fasciculata* resides in a carboxy-terminal extension. *EMBO J.* **7**, 1159–1165. (doi:10. 1002/j.1460-2075.1988.tb02926.x)

126. Crowhurst GS, Isupov MN, Fleming T, Littlechild JA. 1998 Two glycolytic enzymes from *Sulfolobus solfataricus. Biochem. Soc. Trans.* **26**, S275. (doi:10. 1042/bst026s275)

127. Kirino H, Aoki M, Aoshima M, Hayashi Y, Ohba M, Yamagishi A, Wakagi T, Oshima T. 1994 Hydrophobic interaction at the subunit interface contributes to the thermostability of 3-isopropylmalate dehydrogenase from an extreme thermophile, *Thermus thermophilus. Eur. J. Biochem.* **220**, 275–281. (doi:10.1111/j.1432-1033.1994. tb18623.x)

128. Myre MA, O'Day DH. 2004 Calmodulin binds to and inhibits the activity of phosphoglycerate kinase. *Biochim. Biophys. Acta* **1693**, 177–183. (doi:10. 1016/j.bbamcr.2004.08.003)

129. Pan J-Y, Wu H, Liu X, Li P-P, Li H, Wang S-Y, Peng X-X. 2011 Complexome of *Escherichia coli* cytosolic proteins under normal native conditions. *Mol. Biosyst.* **7**, 2651–2663. (doi:10.1039/c1mb05103b)

130. Gao X-H *et al.* 2015 Quantitative H2S-mediated protein sulfhydration reveals metabolic reprogramming during the integrated stress response. *eLife* **4**, e10067. (doi:10.7554/eLife.10067)

131. Bräsen C, Esser D, Rauch B, Siebers B. 2014 Carbohydrate metabolism in Archaea: current insights into unusual enzymes and pathways and their regulation. *Microbiol. Mol. Biol. Rev.* **78**, 89–175. (doi:10.1128/MMBR.00041-13)

132. Fokina KV, Dainyak MB, Nagradova NK, Muronetz VI. 1997 A study on the complexes between human erythrocyte enzymes participating in the conversions of 1,3-diphosphoglycerate. *Arch. Biochem. Biophys.* **345**, 185–192. (doi:10.1006/abbi. 1997.0222)

133. Axelrod B, Bandurski RS. 1953 Phosphoglyceryl kinase in higher plants. *J. Biol. Chem.* **204**, 939–948.

134. Rao DR, Oesper P. 1961 Purification and properties of muscle phosphoglycerate kinase. *Biochem. J.* **81**, 405–411. (doi:10.1042/bj0810405)

135. Nowak T, Suelter C. 1981 Pyruvate kinase: activation by and catalytic role of the monovalent and divalent cations. *Mol. Cell. Biochem.* **35**, 65–75. (doi:10.1007/BF02354821)

136. Pacold I, Anderson LE. 1973 Energy charge control of the Calvin cycle enzyme 3-phosphoglyceric acid kinase. *Biochem. Biophys. Res. Commun.* **51**, 139–143. (doi:10.1016/0006-291x(73)90519-6)

137. Larsson-Raźnikiewicz M, Arvidsson L. 1971 Inhibition of phosphoglycerate kinase by products and product homologues. *Eur. J. Biochem.* **22**, 506–512. (doi:10.1111/j.1432-1033.1971.tb01570.x)

138. Morisse S, Zaffagnini M, Gao X-H, Lemaire SD, Marchand CH. 2014 Insight into protein S-nitrosylation in *Chlamydomonas reinhardtii. Antioxid. Redox Signal.* **21**, 1271–1284. (doi:10. 1089/ars.2013.5632)

139. Morisse S, Michelet L, Bedhomme M, Marchand CH, Calvaresi M, Trost P, Fermani S, Zaffagnini M, Lemaire SD. 2014 Thioredoxin-dependent redox regulation of chloroplastic phosphoglycerate kinase from *Chlamydomonas reinhardtii. J. Biol. Chem.* **289**, 30 012–30 024. (doi:10.1074/jbc.M114.597997)

140. Michelet L, Zaffagnini M, Vanacker H, Le Maréchal P, Marchand C, Schroda M, Lemaire SD, Decottignies P. 2008 In vivo targets of S-thiolation in *Chlamydomonas reinhardtii. J. Biol. Chem.* **283**, 21 571–21 578. (doi:10.1074/jbc.M802331200)

141. Zaffagnini M, Bedhomme M, Groni H, Marchand CH, Puppo C, Gontero B, Cassier-Chauvat C, Decottignies P, Lemaire SD. 2012 Glutathionylation in the photosynthetic model organism *Chlamydomonas reinhardtii*: a proteomic survey. *Mol. Cell. Proteomics MCP* **11**, M111.014142. (doi:10.1074/mcp.M111. 014142)

142. Chen Z, Zhuang W, Wang Z, Xiao W, Don W, Li X, Chen X. 2019 MicroRNA-450b-3p inhibits cell growth by targeting phosphoglycerate kinase 1 in hepatocellular carcinoma. *J. Cell. Biochem.* **120**, 18 805–18 815. (doi:10.1002/jcb.29196)

143. Ge J, Li J, Na S, Wang P, Zhao G, Zhang X. 2019 miR-548c-5p inhibits colorectal cancer cell proliferation by targeting PGK1. *J. Cell. Physiol.* **234**, 18 872–18 878. (doi:10.1002/jcp.28525)

144. Yu T *et al.* 2017 MetaLnc9 facilitates lung cancer metastasis via a PGK1-activated AKT/mTOR pathway. *Cancer Res.* **77**, 5782–5794. (doi:10.1158/ 0008-5472.CAN-17-0671)

145. Nitulescu GM *et al.* 2018 The Akt pathway in oncology therapy and beyond (Review). *Int. J. Oncol.* **53**, 2319–2331. (doi:10.3892/ijo. 2018.4597)

146. Cai Q, Wang S, Jin L, Weng M, Zhou D, Wang J, Tang Z, Quan Z. 2019 Long non-coding RNA GBCDRlnc1 induces chemoresistance of gallbladder cancer cells by activating autophagy. *Mol. Cancer* **18**, 82. (doi:10.1186/s12943-019-1016-0)

147. Jeffery CJ. 2014 An introduction to protein moonlighting. *Biochem. Soc. Trans.* **42**, 1679–1683. (doi:10.1042/BST20140226)

148. Jeffery CJ. 2017 Moonlighting proteins—nature's Swiss army knives. *Sci. Prog.* **100**, 363–373. (doi:10. 3184/003685017X15063357842574)

149. Jeffery CJ. 2018 Protein moonlighting: what is it, and why is it important? *Phil. Trans. R. Soc. B* **373**, 20160523. (doi:10.1098/rstb.2016.0523)

150. Sriram G, Martinez JA, McCabe ERB, Liao JC, Dipple KM. 2005 Single-gene disorders: what role could moonlighting enzymes play? *Am. J. Hum. Genet.* **76**, 911–924. (doi:10.1086/430799)

151. Gómez-Arreaza A, Acosta H, Quiñones W, Concepción JL, Michels PAM, Avilán L. 2014 Extracellular functions of glycolytic enzymes of parasites: unpredicted use of ancient proteins. *Mol. Biochem. Parasitol.* **193**, 75–81. (doi:10.1016/j. molbiopara.2014.02.005)

152. Huangyang P, Simon MC. 2018 Hidden features: exploring the non-canonical functions of metabolic enzymes. *Dis. Model. Mech.* **11**, dmm033365. (doi:10.1242/dmm.033365)

153. Hughes MJG et al. 2002 Identification of major outer surface proteins of Streptococcus agalactiae. Infect. Immun. **70**, 1254–1259. (doi:10.1128/iai.70.3.1254-1259.2002)

154. Alloush HM, López-Ribot JL, Masten BJ, Chaffin WL. 1997 3-phosphoglycerate kinase: a glycolytic enzyme protein present in the cell wall of Candida albicans. Microbiol. Read. Engl. **143**(Pt 2), 321–330. (doi:10.1099/00221287-143-2-321)

155. Labroussaa F, Arricau-Bouvery N, Dubrana M-P, Saillard C. 2010 Entry of Spiroplasma citri into Circulifer haematoceps cells involves interaction between spiroplasma phosphoglycerate kinase and leafhopper actin. Appl. Environ. Microbiol. **76**, 1879–1886. (doi:10.1128/AEM.02384-09)

156. Boone TJ, Burnham C-AD, Tyrrell GJ. 2011 Binding of group B streptococcal phosphoglycerate kinase to plasminogen and actin. Microb. Pathog. **51**, 255–261. (doi:10.1016/j.micpath.2011.06.005)

157. Miles LA, Plow EF. 1985 Binding and activation of plasminogen on the platelet surface. J. Biol. Chem. **260**, 4303–4311.

158. Plow EF, Freaney DE, Plescia J, Miles LA. 1986 The plasminogen system and cell surfaces: evidence for plasminogen and urokinase receptors on the same cell type. J. Cell Biol. **103**, 2411–2420. (doi:10.1083/jcb.103.6.2411)

159. Boone TJ, Tyrrell GJ. 2012 Identification of the actin and plasminogen binding regions of group B streptococcal phosphoglycerate kinase. J. Biol. Chem. **287**, 29 035–29 044. (doi:10.1074/jbc.M112.361261)

160. Boukouris AE, Zervopoulos SD, Michelakis ED. 2016 Metabolic enzymes moonlighting in the nucleus: metabolic regulation of gene transcription. Trends Biochem. Sci. **41**, 712–730. (doi:10.1016/j.tibs.2016.05.013)

161. Rutter J, Michnoff CH, Harper SM, Gardner KH, McKnight SL. 2001 PAS kinase: an evolutionarily conserved PAS domain-regulated serine/threonine kinase. Proc. Natl. Acad. Sci. USA **98**, 8991–8996. (doi:10.1073/pnas.161284798)

162. Fjeld CC, Birdsong WT, Goodman RH. 2003 Differential binding of NAD+ and NADH allows the transcriptional corepressor carboxyl-terminal binding protein to serve as a metabolic sensor. Proc. Natl Acad. Sci. USA **100**, 9202–9207. (doi:10.1073/pnas.1633591100)

163. Wright RHG et al. 2016 ADP-ribose-derived nuclear ATP synthesis by NUDIX5 is required for chromatin remodeling. Science **352**, 1221–1225. (doi:10.1126/science.aad9335)

164. Li X, Egervari G, Wang Y, Berger SL, Lu Z. 2018 Regulation of chromatin and gene expression by metabolic enzymes and metabolites. Nat. Rev. Mol. Cell Biol. **19**, 563–578. (doi:10.1038/s41580-018-0029-7)

165. Enzo E et al. 2015 Aerobic glycolysis tunes YAP/TAZ transcriptional activity. EMBO J. **34**, 1349–1370. (doi:10.15252/embj.201490379)

166. Yalcin A et al. 2014 6-Phosphofructo-2-kinase (PFKFB3) promotes cell cycle progression and suppresses apoptosis via Cdk1-mediated phosphorylation of p27. Cell Death Dis. **5**, e1337. (doi:10.1038/cddis.2014.292)

167. Li B et al. 2014 Fructose-1,6-bisphosphatase opposes renal carcinoma progression. Nature **513**, 251–255. (doi:10.1038/nature13557)

168. Zhou Y, Yi X, Stoffer JB, Bonafe N, Gilmore-Hebert M, McAlpine J, Chambers SK. 2008 The multifunctional protein glyceraldehyde-3-phosphate dehydrogenase is both regulated and controls colony-stimulating factor-1 messenger RNA stability in ovarian cancer. Mol. Cancer Res. **6**, 1375–1384. (doi:10.1158/1541-7786.MCR-07-2170)

169. Popanda O, Fox G, Thielmann HW. 1998 Modulation of DNA polymerases alpha, delta and epsilon by lactate dehydrogenase and 3-phosphoglycerate kinase. Biochim. Biophys. Acta **1397**, 102–117. (doi:10.1016/s0167-4781(97)00229-7)

170. Wang W, Wang L, Endoh A, Hummelke G, Hawks CL, Hornsby PJ. 2005 Identification of alpha-enolase as a nuclear DNA-binding protein in the zona fasciculata but not the zona reticularis of the human adrenal cortex. J. Endocrinol. **184**, 85–94. (doi:10.1677/joe.1.05909)

171. Wang H-J et al. 2014 JMJD5 regulates PKM2 nuclear translocation and reprograms HIF-1α-mediated glucose metabolism. Proc. Natl Acad. Sci. USA **111**, 279–284. (doi:10.1073/pnas.1311249111)

172. Anderson LE, Bryant JA, Carol AA. 2004 Both chloroplastic and cytosolic phosphoglycerate kinase isozymes are present in the pea leaf nucleus. Protoplasma **223**, 103–110. (doi:10.1007/s00709-004-0041-y)

173. Brice DC, Bryant JA, Dambrauskas G, Drury SC, Littlechild JA. 2004 Cloning and expression of cytosolic phosphoglycerate kinase from pea (Pisum sativum L.). J. Exp. Bot. **55**, 955–956. (doi:10.1093/jxb/erh096)

174. Vishwanatha JK, Jindal HK, Davis RG. 1992 The role of primer recognition proteins in DNA replication: association with nuclear matrix in HeLa cells. J. Cell Sci. **101**, 25–34.

175. Foiani M, Liberi G, Lucchini G, Plevani P. 1995 Cell cycle-dependent phosphorylation and dephosphorylation of the yeast DNA polymerase alpha-primase B subunit. Mol. Cell. Biol. **15**, 883–891. (doi:10.1128/mcb.15.2.883)

176. Lay AJ, Jiang X-M, Daly E, Sun L, Hogg PJ. 2002 Plasmin reduction by phosphoglycerate kinase is a thiol-independent process. J. Biol. Chem. **277**, 9062–9068. (doi:10.1074/jbc.M111531200)

177. Yu L, Chen X, Wang L, Chen S. 2016 The sweet trap in tumors: aerobic glycolysis and potential targets for therapy. Oncotarget **7**, 38 908–38 926. (doi:10.18632/oncotarget.7676)

178. Fiorillo A, Petrosino M, Ilari A, Pasquo A, Cipollone A, Maggi M, Chiaraluce R, Consalvi V. 2018 The phosphoglycerate kinase 1 variants found in carcinoma cells display different catalytic activity and conformational stability compared to the native enzyme. PLoS ONE **13**, e0199191. (doi:10.1371/journal.pone.0199191)

179. Hogg PJ. 2002 Biological regulation through protein disulfide bond cleavage. Redox Rep. Commun. Free Radic. Res. **7**, 71–77. (doi:10.1179/135100002125000299)

180. Xie H, Tong G, Zhang Y, Liang S, Tang K, Yang Q. 2017 PGK1 drives hepatocellular carcinoma metastasis by enhancing metabolic process. Int. J. Mol. Sci. **18**, 1630. (doi:10.3390/ijms18081630)

181. Irshad K et al. 2015 A combined gene signature of hypoxia and notch pathway in human glioblastoma and its prognostic relevance. PLoS ONE **10**, e0118201. (doi:10.1371/journal.pone.0118201)

182. Zieker D et al. 2010 Phosphoglycerate kinase 1 promoting tumor progression and metastasis in gastric cancer detected in a tumor mouse model using positron emission tomography/magnetic resonance imaging. Cell. Physiol. Biochem. **26**, 147–154. (doi:10.1159/000320545)

183. Li X et al. 2016 Mitochondria-translocated PGK1 functions as a protein kinase to coordinate glycolysis and the TCA cycle in tumorigenesis. Mol. Cell **61**, 705–719. (doi:10.1016/j.molcel.2016.02.009)

184. Hu H et al. 2017 Acetylation of PGK1 promotes liver cancer cell proliferation and tumorigenesis. Hepatology (Baltimore Md) **65**, 515–528. (doi:10.1002/hep.28887)

185. Qian X et al. 2017 Phosphoglycerate kinase 1 phosphorylates beclin1 to induce autophagy. Mol. Cell **65**, 917–931.e6. (doi:10.1016/j.molcel.2017.01.027)

186. Cheng S-F, Huang Y-P, Chen L-H, Hsu Y-H, Tsai C-H. 2013 Chloroplast phosphoglycerate kinase is involved in the targeting of bamboo mosaic virus to chloroplasts in Nicotiana benthamiana plants. Plant Physiol. **163**, 1598–1608. (doi:10.1104/pp.113.229666)

187. Budziszewska M, Obrępalska-Stęplowska A. 2018 The role of the chloroplast in the replication of positive-sense single-stranded plant RNA viruses. Front. Plant Sci. **9**, 1776. (doi:10.3389/fpls.2018.01776)

188. Ogino T, Iwama M, Kinouchi J, Shibagaki Y, Tsukamoto T, Mizumoto K. 1999 Involvement of a cellular glycolytic enzyme, phosphoglycerate kinase, in Sendai virus transcription. J. Biol. Chem. **274**, 35 999–36 008. (doi:10.1074/jbc.274.50.35999)

189. Ogino T, Yamadera T, Nonaka T, Imajoh-Ohmi S, Mizumoto K. 2001 Enolase, a cellular glycolytic enzyme, is required for efficient transcription of Sendai virus genome. Biochem. Biophys. Res. Commun. **285**, 447–455. (doi:10.1006/bbrc.2001.5160)

190. Orias E, Cervantes MD, Hamilton EP. 2011 Tetrahymena thermophila, a unicellular eukaryote with separate germline and somatic genomes. Res. Microbiol. **162**, 578–586. (doi:10.1016/j.resmic.2011.05.001)

191. Hamilton EP et al. 2016 Structure of the germline genome of Tetrahymena thermophila and relationship to the massively rearranged somatic genome. eLife **5**, e19090. (doi:10.7554/eLife.19090)

192. Saettone A, Nabeel-Shah S, Garg J, Lambert J-P, Pearlman RE, Fillingham J. 2019 Functional

proteomics of nuclear proteins in *Tetrahymena thermophila*: a review. *Genes* **10**, 333. (doi:10.3390/genes10050333)

193. Noto T, Mochizuki K. 2017 Whats, hows and whys of programmed DNA elimination in *Tetrahymena*. *Open Biol.* **7**, 170172. (doi:10.1098/rsob.170172)

194. Ruehle MD, Orias E, Pearson CG. 2016 *Tetrahymena* as a unicellular model eukaryote: genetic and genomic tools. *Genetics* **203**, 649–665. (doi:10.1534/genetics.114.169748)

195. Eisen JA et al. 2006 Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol.* **4**, e286. (doi:10.1371/journal.pbio.0040286)

196. Endo M, Sugai T. 2011 Amitotic division of the macronucleus in *Tetrahymena thermophila*: DNA distribution by genomic unit. *Zool. Sci.* **28**, 482–490. (doi:10.2108/zsj.28.482)

197. Lin R, Cook RG, Allis CD. 1991 Proteolytic removal of core histone amino termini and dephosphorylation of histone H1 correlate with the formation of condensed chromatin and transcriptional silencing during *Tetrahymena* macronuclear development. *Genes Dev.* **5**, 1601–1610. (doi:10.1101/gad.5.9.1601)

198. Vohra GB, Golding GB, Tsao N, Pearlman RE. 1992 A phylogenetic analysis based on the gene encoding phosphoglycerate kinase. *J. Mol. Evol.* **35**, 466.

199. Yukitake H, Furusawa M, Taira T, Iguchi-Ariga SMM, Ariga H. 2002 AMAP-1, a novel testis-specific AMY-1-binding protein, is differentially expressed during the course of spermatogenesis. *Biochim. Biophys. Acta* **1577**, 126–132. (doi:10.1016/s0167-4781(02)00411-6)

200. Kushida Y, Takaine M, Nakano K, Sugai T, Vasudevan KK, Guha M, Jiang Y-Y, Gaertig J, Numata O. 2017 Kinesin-14 is important for chromosome segregation during mitosis and meiosis in the ciliate *Tetrahymena thermophila*. *J. Eukaryot. Microbiol.* **64**, 293–307. (doi:10.1111/jeu.12366)

201. Ashraf K, Nabeel-Shah S, Garg J, Saettone A, Derynck J, Gingras A-C, Lambert J-P, Pearlman RE, Fillingham J. 2019 Proteomic analysis of histones H2A/H2B and variant Hv1 in *Tetrahymena thermophila* reveals an ancient network of chaperones. *Mol. Biol. Evol.* **36**, 1037–1055. (doi:10.1093/molbev/msz039)

202. Kataoka K, Noto T, Mochizuki K. 2016 Phosphorylation of an HP1-like protein is a prerequisite for heterochromatin body formation in *Tetrahymena* DNA elimination. *Proc. Natl Acad. Sci. USA* **113**, 9027–9032. (doi:10.1073/pnas.1606012113)

203. Stargell LA, Gorovsky MA. 1994 TATA-binding protein and nuclear differentiation in *Tetrahymena thermophila*. *Mol. Cell. Biol.* **14**, 723–734. (doi:10.1128/mcb.14.1.723)

204. Miao W et al. 2009 Microarray analyses of gene expression during the *Tetrahymena thermophila* life cycle. *PLoS ONE* **4**, e4429. (doi:10.1371/journal.pone.0004429)

205. Vanclová AMGN et al. 2019 Proteome of the secondary plastid of *Euglena gracilis* reveals metabolic quirks and colourful history. *bioRxiv*, 573709. (doi:10.1101/573709)

206. Sulli C, Schwartzbach SD. 1995 The polyprotein precursor to the *Euglena* light-harvesting chlorophyll a/b-binding protein is transported to the Golgi apparatus prior to chloroplast import and polyprotein processing. *J. Biol. Chem.* **270**, 13 084–13 090. (doi:10.1074/jbc.270.22.13084)

207. Sulli C, Schwartzbach SD. 1996 A soluble protein is imported into *Euglena* chloroplasts as a membrane-bound precursor. *Plant Cell* **8**, 43–53. (doi:10.1105/tpc.8.1.43)

208. Enomoto T, Sulli C, Schwartzbach SD. 1997 A soluble chloroplast protease processes the *Euglena* polyprotein precursor to the light harvesting chlorophyll a/b binding protein of photosystem II. *Plant Cell Physiol.* **38**, 743–746. (doi:10.1093/oxfordjournals.pcp.a029229)

209. Koziol AG, Durnford DG. 2008 *Euglena* light-harvesting complexes are encoded by multifarious polyprotein mRNAs that evolve in concert. *Mol. Biol. Evol.* **25**, 92–100. (doi:10.1093/molbev/msm232)

210. Houlné G, Schantz R. 1993 Expression of polyproteins in *Euglena*. *Crit. Rev. Plant Sci.* **12**, 1–17. (doi:10.1080/07352689309382354)

211. Durnford DG, Gray MW. 2006 Analysis of *Euglena gracilis* plastid-targeted proteins reveals different classes of transit sequences. *Eukaryot. Cell* **5**, 2079–2091. (doi:10.1128/EC.00222-06)

212. Künne A, Pistorius E, Kloppstech K, de Groot E. 1998 Circadian synthesis of light-harvesting-chlorophyll-proteins in *Euglena gracilis* is under translational control. *Z. Naturforschung C J. Biosci.* **53**, 1017–1026. (doi:10.1515/znc-1998-11-1213)

213. Russell GK, Draffan AG, Schmidt GW, Lyman H. 1978 Light-induced enzyme formation in a chlorophyll-less mutant of *Euglena gracilis*. *Plant Physiol.* **62**, 678–682. (doi:10.1104/pp.62.5.678)

214. Walla OJ, de Groot EJ, Schweiger M. 1994 On the molecular mechanism of the circadian clock. The 41 000 M(r) clock protein of Chlorella was identified as 3-phosphoglycerate kinase. *J. Cell Sci.* **107**(Pt 2), 719–726.

215. Rijo-Ferreira F, Pinto-Neves D, Barbosa-Morais NL, Takahashi JS, Figueiredo LM. 2017 *Trypanosoma brucei* metabolism is under circadian control. *Nat. Microbiol.* **2**, 17032. (doi:10.1038/nmicrobiol.2017.32)

216. Catalano A, O'Day DH. 2008 Calmodulin-binding proteins in the model organism *Dictyostelium*: a complete & critical review. *Cell. Signal.* **20**, 277–291. (doi:10.1016/j.cellsig.2007.08.017)

217. Beitner R. 1993 Control of glycolytic enzymes through binding to cell structures and by glucose-1,6-bisphosphate under different conditions. The role of Ca^{2+} and calmodulin. *Int. J. Biochem.* **25**, 297–305. (doi:10.1016/0020-711x(93)90616-m)

218. Beitner R. 1998 Calmodulin antagonists and cell energy metabolism in health and disease. *Mol. Genet. Metab.* **64**, 161–168. (doi:10.1006/mgme.1998.2691)

219. Buschmeier B, Meyer HE, Mayr GW. 1987 Characterization of the calmodulin-binding sites of muscle phosphofructokinase and comparison with known calmodulin-binding domains. *J. Biol. Chem.* **262**, 9454–9462.

220. Mayr GW, Heilmeyer LM. 1983 Phosphofructokinase is a calmodulin binding protein. *FEBS Lett.* **159**, 51–57. (doi:10.1016/0014-5793(83)80415-3)

221. Singh P, Salih M, Leddy JJ, Tuana BS. 2004 The muscle-specific calmodulin-dependent protein kinase assembles with the glycolytic enzyme complex at the sarcoplasmic reticulum and modulates the activity of glyceraldehyde-3-phosphate dehydrogenase in a Ca2+/calmodulin-dependent manner. *J. Biol. Chem.* **279**, 35 176–35 182. (doi:10.1074/jbc.M402282200)

222. Scherer A, Kuhl S, Wessels D, Lusche DF, Raisley B, Soll DR. 2010 Ca^{2+} chemotaxis in *Dictyostelium discoideum*. *J. Cell Sci.* **123**, 3756–3767. (doi:10.1242/jcs.068619)

223. O'Day DH, Mathavarajah S, Myre MA, Huber RJ. 2020 Calmodulin-mediated events during the life cycle of the amoebozoan *Dictyostelium discoideum*. *Biol. Rev. Camb. Philos. Soc.* **95**, 472–490. (doi:10.1111/brv.12573)

224. Fritz-Laylin LK et al. 2010 The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. *Cell* **140**, 631–642. (doi:10.1016/j.cell.2010.01.032)

225. Opperdoes FR, De Jonckheere JF, Tielens AGM. 2011 *Naegleria gruberi* metabolism. *Int. J. Parasitol.* **41**, 915–924. (doi:10.1016/j.ijpara.2011.04.004)

226. Nerad TA, Visvesvara G, Daggett PM. 1983 Chemically defined media for the cultivation of Naegleria: pathogenic and high temperature tolerant species. *J. Protozool.* **30**, 383–387. (doi:10.1111/j.1550-7408.1983.tb02935.x)

227. Henze K, Horner DS, Suguri S, Moore DV, Sánchez LB, Müller M, Embley TM. 2001 Unique phylogenetic relationships of glucokinase and glucosephosphate isomerase of the amitochondriate eukaryotes *Giardia intestinalis*, *Spironucleus barkhanus* and *Trichomonas vaginalis*. *Gene* **281**, 123–131. (doi:10.1016/s0378-1119(01)00773-9)

228. Wu G, Henze K, Müller M. 2001 Evolutionary relationships of the glucokinase from the amitochondriate protist, *Trichomonas vaginalis*. *Gene* **264**, 265–271. (doi:10.1016/s0378-1119(01)00329-8)

229. Cáceres AJ, Quiñones W, Gualdrón M, Cordeiro A, Avilán L, Michels PAM, Concepción JL. 2007 Molecular and biochemical characterization of novel glucokinases from *Trypanosoma cruzi* and *Leishmania* spp. *Mol. Biochem. Parasitol.* **156**, 235–245. (doi:10.1016/j.molbiopara.2007.08.007)

230. Mertens E, De Jonckheere J, Van Schaftingen E. 1993 Pyrophosphate-dependent phosphofructokinase from the amoeba Naegleria fowleri, an AMP-sensitive enzyme. *Biochem. J.* **292**(Pt 3), 797–803. (doi:10.1042/bj2920797)

231. Krisfalusi M, Miki K, Magyar PL, O'Brien DA. 2006 Multiple glycolytic enzymes are tightly bound to the fibrous sheath of mouse spermatozoa. *Biol. Reprod.* **75**, 270–278. (doi:10.1095/biolreprod.105.049684)

232. Bexkens ML et al. 2018 Lipids are the preferred substrate of the protist *Naegleria gruberi*, relative of

a human brain pathogen. *Cell Rep.* **25**, 537–543.e3. (doi:10.1016/j.celrep.2018.09.055)

233. Capewell LG, Harris AM, Yoder JS, Cope JR, Eddy BA, Roy SL, Visvesvara GS, Fox LM, Beach MJ. 2015 Diagnosis, clinical course, and treatment of primary amoebic meningoencephalitis in the United States, 1937–2013. *J. Pediatr. Infect. Dis. Soc.* **4**, e68–e75. (doi:10.1093/jpids/piu103)

234. Milanes JE *et al*. 2019 Enzymatic and structural characterization of the *Naegleria fowleri* glucokinase. *Antimicrob. Agents Chemother.* **63**, e02410–18. (doi:10.1128/AAC.02410-18)

235. Barnard JP, Reynafarje B, Pedersen PL. 1993 Glucose catabolism in African trypanosomes. Evidence that the terminal step is catalyzed by a pyruvate transporter capable of facilitating uptake of toxic analogs. *J. Biol. Chem.* **268**, 3654–3661.

236. Reeves RE. 1984 Metabolism of *Entamoeba histolytica* Schaudinn, 1903. *Adv. Parasitol.* **23**, 105–142. (doi:10.1016/s0065-308x(08)60286-9)

237. McLaughlin J, Aley S. 1985 The biochemistry and functional morphology of the Entamoeba. *J. Protozool.* **32**, 221–240. (doi:10.1111/j.1550-7408.1985.tb03043.x)

238. Loftus B *et al*. 2005 The genome of the protist parasite *Entamoeba histolytica*. *Nature* **433**, 865–868. (doi:10.1038/nature03291)

239. Saavedra E, Encalada R, Vázquez C, Olivos-García A, Michels PAM, Moreno-Sánchez R. 2019 Control and regulation of the pyrophosphate-dependent glucose metabolism in *Entamoeba histolytica*. *Mol. Biochem. Parasitol.* **229**, 75–87. (doi:10.1016/j.molbiopara.2019.02.002)

240. Reeves RE, South DJ. 1974 Phosphoglycerate kinase (GTP). An enzyme from *Entamoeba histolytica* selective for guanine nucleotides. *Biochem. Biophys. Res. Commun.* **58**, 1053–1057. (doi:10.1016/s0006-291x(74)80250-0)

241. Clark CG *et al*. 2007 Structure and content of the *Entamoeba histolytica* genome. *Adv. Parasitol.* **65**, 51–190. (doi:10.1016/S0065-308X(07)65002-7)

242. Albach RA. 1989 Nucleic acids of *Entamoeba histolytica*. *J. Protozool.* **36**, 197–205. (doi:10.1111/j.1550-7408.1989.tb01074.x)

243. Han J, Collins LJ. 2012 Reconstruction of sugar metabolic pathways of *Giardia lamblia*. *Int. J. Proteomics* **2012**, 980829. (doi:10.1155/2012/980829)

244. Horváthová L *et al*. 2012 Transcriptomic identification of iron-regulated and iron-independent gene copies within the heavily duplicated *Trichomonas vaginalis* genome. *Genome Biol. Evol.* **4**, 1017–1029. (doi:10.1093/gbe/evs078)

245. Fang Y-K, Huang K-Y, Huang P-J, Lin R, Chao M, Tang P. 2015 Gene-expression analysis of cold-stress response in the sexually transmitted protist *Trichomonas vaginalis*. *J. Microbiol. Immunol. Infect. Wei Mian Yu Gan Ran Za Zhi* **48**, 662–675. (doi:10.1016/j.jmii.2014.07.013)

246. Arese P, Cappuccinelli P. 1974 Glycolysis and pentose phosphate cycle in *Trichomonas vaginalis*: I. Enzyme activity pattern and the constant proportion quintet. *Int. J. Biochem.* **5**, 859–865. (doi:10.1016/0020-711X(74)90121-9)

247. Mundodi V, Kucknoor AS, Chang T-H, Alderete JF. 2006 A novel surface protein of *Trichomonas vaginalis* is regulated independently by low iron and contact with vaginal epithelial cells. *BMC Microbiol.* **6**, 6. (doi:10.1186/1471-2180-6-6)

248. Kummer S, Hayes GR, Gilbert RO, Beach DH, Lucas JJ, Singh BN. 2008 Induction of human host cell apoptosis by *Trichomonas vaginalis* cysteine proteases is modulated by parasite exposure to iron. *Microb. Pathog.* **44**, 197–203. (doi:10.1016/j.micpath.2007.09.004)

249. Alderete JF, Provenzano D, Lehker MW. 1995 Iron mediates *Trichomonas vaginalis* resistance to complement lysis. *Microb. Pathog.* **19**, 93–103. (doi:10.1006/mpat.1995.0049)

250. Flügge U-I, Häusler RE, Ludewig F, Fischer K. 2003 Functional genomics of phosphate antiport systems of plastids. *Physiol. Plant.* **118**, 475–482. (doi:10.1034/j.1399-3054.2003.00137.x)

251. MacRae JI, Sheiner L, Nahid A, Tonkin C, Striepen B, McConville MJ. 2012 Mitochondrial metabolism of glucose and glutamine is required for intracellular growth of *Toxoplasma gondii*. *Cell Host Microbe* **12**, 682–692. (doi:10.1016/j.chom.2012.09.013)

252. Jacot D, Waller RF, Soldati-Favre D, MacPherson DA, MacRae JI. 2016 Apicomplexan energy metabolism: carbon source promiscuity and the quiescence hyperbole. *Trends Parasitol.* **32**, 56–70. (doi:10.1016/j.pt.2015.09.001)

253. Hicks KE, Read M, Holloway SP, Sims PF, Hyde JE. 1991 Glycolytic pathway of the human malaria parasite *Plasmodium falciparum*: primary sequence analysis of the gene encoding 3-phosphoglycerate kinase and chromosomal mapping studies. *Gene* **100**, 123–129. (doi:10.1016/0378-1119(91)90357-h)

254. Pal B, Pybus B, Muccio DD, Chattopadhyay D. 2004 Biochemical characterization and crystallization of recombinant 3-phosphoglycerate kinase of *Plasmodium falciparum*. *Biochim. Biophys. Acta* **1699**, 277–280. (doi:10.1016/j.bbapap.2004.01.003)

255. Liu J, Bolick DT, Kolling GL, Fu Z, Guerrant RL. 2016 Protein malnutrition impairs intestinal epithelial cell turnover, a potential mechanism of increased cryptosporidiosis in a murine model. *Infect. Immun.* **84**, 3542–3549. (doi:10.1128/IAI.00705-16)

256. Abrahamsen MS *et al*. 2004 Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* **304**, 441–445. (doi:10.1126/science.1094786)

257. Entrala E, Mascaró C. 1997 Glycolytic enzyme activities in *Cryptosporidium parvum* oocysts. *FEMS Microbiol. Lett.* **151**, 51–57. (doi:10.1016/s0378-1097(97)00136-5)

258. Mogi T, Kita K. 2010 Diversity in mitochondrial metabolic pathways in parasitic protists *Plasmodium* and *Cryptosporidium*. *Parasitol. Int.* **59**, 305–312. (doi:10.1016/j.parint.2010.04.005)

259. Kersey PJ *et al*. 2018 Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. *Nucleic Acids Res.* **46**, D802–D808. (doi:10.1093/nar/gkx1011)

260. Zhang H, Guo F, Zhou H, Zhu G. 2012 Transcriptome analysis reveals unique metabolic features in the *Cryptosporidium parvum* oocysts associated with environmental survival and stresses. *BMC Genomics* **13**, 647. (doi:10.1186/1471-2164-13-647)

261. Vannier E, Gewurz BE, Krause PJ. 2008 Human babesiosis. *Infect. Dis. Clin. North Am.* **22**, 469–488, viii–ix. (doi:10.1016/j.idc.2008.03.010)

262. Jones SH, Lew AE, Jorgensen WK, Barker SC. 1997 *Babesia bovis*: genome size, number of chromosomes and telomeric probe hybridisation. *Int. J. Parasitol.* **27**, 1569–1573. (doi:10.1016/s0020-7519(97)00099-4)

263. Mosqueda J, Hernández-Silva DJ, Hidalgo-Ruiz M. 2017 Genome-based vaccinology applied to bovine Babesiosis. In *Farm animals diseases, recent omic trends and new strategies of treatment* (ed. RE Quiroz-Castañeda). Rijeka, Croatia: Intech.

264. Brayton KA *et al*. 2007 Genome sequence of *Babesia bovis* and comparative analysis of apicomplexan hemoprotozoa. *PLoS Pathog.* **3**, 1401–1413. (doi:10.1371/journal.ppat.0030148)

265. Magni R, Luchini A, Liotta L, Molestina RE. 2019 Analysis of the Babesia microti proteome in infected red blood cells by a combination of nanotechnology and mass spectrometry. *Int. J. Parasitol.* **49**, 139–144. (doi:10.1016/j.ijpara.2018.08.004)

266. Barry DN. 1984 Metabolism of Babesia parasites in vitro. Glucose and energy metabolism of *B. bovis*. *Aust. J. Exp. Biol. Med. Sci.* **62**(Pt 1), 53–61. (doi:10.1038/icb.1984.5)

267. Upston JM, Gero AM. 1990 Increased glucose permeability in *Babesia bovis*-infected erythrocytes. *Int. J. Parasitol.* **20**, 69–76. (doi:10.1016/0020-7519(90)90175-m)

268. Shikano S, Shimada T, Nakada K, Hashiguchi R, Ono K. 1995 Enzyme activities related to glucose metabolism in *Babesia microti* and *Babesia rodhaini*. *J. Vet. Med. Sci.* **57**, 93–97. (doi:10.1292/jvms.57.93)

269. Rajapakshage BKW *et al*. 2012 Analysis of energy generation and glycolysis pathway in diminazene aceturate-resistant *Babesia gibsoni* isolate in vitro. *Jpn. J. Vet. Res.* **60**, 51–61.

270. Gentekaki E *et al*. 2017 Extreme genome diversity in the hyper-prevalent parasitic eukaryote Blastocystis. *PLoS Biol.* **15**, e2003769. (doi:10.1371/journal.pbio.2003769)

271. Nakayama T, Ishida K, Archibald JM. 2012 Broad distribution of TPI-GAPDH fusion proteins among eukaryotes: evidence for glycolytic reactions in the mitochondrion? *PLoS ONE* **7**, e52340. (doi:10.1371/journal.pone.0052340)

272. Abrahamian M, Kagda M, Ah-Fong AMV, Judelson HS. 2017 Rethinking the evolution of eukaryotic metabolism: novel cellular partitioning of enzymes in stramenopiles links serine biosynthesis to glycolysis in mitochondria. *BMC Evol. Biol.* **17**, 241. (doi:10.1186/s12862-017-1087-8)

273. Río Bártulos C *et al*. 2018 Mitochondrial glycolysis in a major lineage of eukaryotes. *Genome Biol. Evol.* **10**, 2310–2325. (doi:10.1093/gbe/evy164)

274. Denoeud F *et al*. 2011 Genome sequence of the stramenopile Blastocystis, a human anaerobic parasite. *Genome Biol.* **12**, R29. (doi:10.1186/gb-2011-12-3-r29)

275. Lukeš J, Flegontova O, Horák A. 2015 Diplonemids. *Curr. Biol.* **25**, R702–R704. (doi:10.1016/j.cub.2015.04.052)

276. Butenko A *et al*. 2020 Evolution of metabolic capabilities and molecular features of diplonemids, kinetoplastids, and euglenids. *BMC Biol.* **18**, 23. (doi:10.1186/s12915-020-0754-1)

277. Makiuchi T, Annoura T, Hashimoto M, Hashimoto T, Aoki T, Nara T. 2011 Compartmentalization of a glycolytic enzyme in Diplonema, a non-kinetoplastid euglenozoan. *Protist* **162**, 482–489. (doi:10.1016/j.protis.2010.11.003)

278. Adjé CA, Opperdoes FR, Michels PA. 1998 Molecular analysis of phosphoglycerate kinase in *Trypanoplasma borreli* and the evolution of this enzyme in kinetoplastida. *Gene* **217**, 91–99. (doi:10.1016/s0378-1119(98)00356-4)

279. Azevedo A, Toledo JS, Defina T, Pedrosa AL, Cruz AK. 2015 Leishmania major phosphoglycerate kinase transcript and protein stability contributes to differences in isoform expression levels. *Exp. Parasitol.* **159**, 222–226. (doi:10.1016/j.exppara.2015.09.008)

280. Jackson AP. 2007 Tandem gene arrays in *Trypanosoma brucei*: comparative phylogenomic analysis of duplicate sequence variation. *BMC Evol. Biol.* **7**, 54. (doi:10.1186/1471-2148-7-54)

281. Adjé CA, Opperdoes FR, Michels PA. 1997 Organization, sequence and stage-specific expression of the phosphoglycerate kinase genes of *Leishmania mexicana* mexicana. *Mol. Biochem. Parasitol.* **90**, 155–168. (doi:10.1016/s0166-6851(97)00152-7)

282. Opperdoes FR, Butenko A, Flegontov P, Yurchenko V, Lukeš J. 2016 Comparative metabolism of free-living *Bodo saltans* and parasitic trypanosomatids. *J. Eukaryot. Microbiol.* **63**, 657–678. (doi:10.1111/jeu.12315)

283. Leslie M. 2013 'Dead' enzymes show signs of life. *Science* **340**, 25–27. (doi:10.1126/science.340.6128.25)

284. Vasiliou V *et al*. 2013 ALDH16A1 is a novel non-catalytic enzyme that may be involved in the etiology of gout via protein protein interactions with HPRT1. *Chem. Biol. Interact.* **202**, 22–31. (doi:10.1016/j.cbi.2012.12.018)

285. Eyers PA, Murphy JM. 2016 The evolving world of pseudoenzymes: proteins, prejudice and zombies. *BMC Biol.* **14**, 98. (doi:10.1186/s12915-016-0322-x)

286. Tanifuji G *et al*. 2017 Genome sequencing reveals metabolic and cellular interdependence in an amoeba-kinetoplastid symbiosis. *Sci. Rep.* **7**, 11688. (doi:10.1038/s41598-017-11866-x)

287. Porcel BM *et al*. 2014 The streamlined genome of Phytomonas spp. relative to human pathogenic kinetoplastids reveals a parasite tailored for plants. *PLoS Genet.* **10**, e1004007. (doi:10.1371/journal.pgen.1004007)

288. Flegontov P *et al*. 2016 Genome of *Leptomonas pyrrhocoris*: a high-quality reference for monoxenous trypanosomatids and new insights into evolution of *Leishmania*. *Sci. Rep.* **6**, 23704. (doi:10.1038/srep23704)

289. Berriman M *et al*. 2005 The genome of the African trypanosome *Trypanosoma brucei*. *Science* **309**, 416–422. (doi:10.1126/science.1112642)

290. Ivens AC *et al*. 2005 The genome of the kinetoplastid parasite, *Leishmania major*. *Science* **309**, 436–442. (doi:10.1126/science.1112680)

291. El-Sayed NM *et al*. 2005 The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* **309**, 409–415. (doi:10.1126/science.1112631)

292. Lukeš J, Butenko A, Hashimi H, Maslov DA, Votýpka J, Yurchenko V. 2018 trypanosomatids are much more than just Trypanosomes: clues from the expanded family tree. *Trends Parasitol.* **34**, 466–480. (doi:10.1016/j.pt.2018.03.002)

293. Flegontov P *et al*. 2013 Paratrypanosoma is a novel early-branching trypanosomatid. *Curr. Biol.* **23**, 1787–1793. (doi:10.1016/j.cub.2013.07.045)

294. Skalický T *et al*. 2017 Extensive flagellar remodeling during the complex life cycle of Paratrypanosoma, an early-branching trypanosomatid. *Proc. Natl Acad. Sci. USA* **114**, 11 757–11 762. (doi:10.1073/pnas.1712311114)

295. Cho J, King JS, Qian X, Harwood AJ, Shears SB. 2008 Dephosphorylation of 2,3-bisphosphoglycerate by MIPP expands the regulatory capacity of the Rapoport-Luebering glycolytic shunt. *Proc. Natl Acad. Sci. USA* **105**, 5998–6003. (doi:10.1073/pnas.0710980105)

296. Fischbach A, Adelt S, Müller A, Vogel G. 2006 Disruption of inositol biosynthesis through targeted mutagenesis in *Dictyostelium discoideum*: generation and characterization of inositol-auxotrophic mutants. *Biochem. J.* **397**, 509–518. (doi:10.1042/BJ20060277)

297. Peterson GC, Sommer JM, Klosterman S, Wang CC, Parsons M. 1997 *Trypanosoma brucei*: identification of an internal region of phosphoglycerate kinase required for targeting to glycosomal microbodies. *Exp. Parasitol.* **85**, 16–23. (doi:10.1006/expr.1996.4114)

298. Blattner J, Helfert S, Michels P, Clayton C. 1998 Compartment of phosphoglycerate kinase in *Trypanosoma brucei* plays a critical role in parasite energy metabolism. *Proc. Natl Acad. Sci. USA* **95**, 11 596–11 600. (doi:10.1073/pnas.95.20.11596)

299. Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM. 2005 The many faces of the helix–turn–helix domain: transcription regulation and beyond. *FEMS Microbiol. Rev.* **29**, 231–262. (doi:10.1016/j.femsre.2004.12.008)

300. Wilson KP, Shewchuk LM, Brennan RG, Otsuka AJ, Matthews BW. 1992 *Escherichia coli* biotin holoenzyme synthetase/bio repressor crystal structure delineates the biotin- and DNA-binding domains. *Proc. Natl Acad. Sci. USA* **89**, 9257–9261. (doi:10.1073/pnas.89.19.9257)

301. Nguyen HT. 2014 Understanding mechanisms of motility in *Trypanosoma brucei*. UCLA. See https://escholarship.org/uc/item/7md8b9j5.

302. Oberholzer M, Saada EA, Hill KL. 2015 Cyclic AMP regulates social behavior in African trypanosomes. *mBio* **6**, e01954-14. (doi:10.1128/mBio.01954-14)

303. Lopez MA, Nguyen HT, Oberholzer M, Hill KL. 2011 Social parasites. *Curr. Opin. Microbiol.* **14**, 642–648. (doi:10.1016/j.mib.2011.09.012)

304. Shaw S, DeMarco SF, Rehmann R, Wenzler T, Florini F, Roditi I, Hill KL. 2019 Flagellar cAMP signaling controls trypanosome progression through host tissues. *Nat. Commun.* **10**, 803. (doi:10.1038/s41467-019-08696-y)

305. Imhof S, Vu XL, Bütikofer P, Roditi I. 2015 A glycosylation mutant of *Trypanosoma brucei* links social motility defects in vitro to impaired colonization of tsetse flies in vivo. *Eukaryot. Cell* **14**, 588–592. (doi:10.1128/EC.00023-15)

306. Oberholzer M, Marti G, Baresic M, Kunz S, Hemphill A, Seebeck T. 2007 The *Trypanosoma brucei* cAMP phosphodiesterases TbrPDEB1 and TbrPDEB2: flagellar enzymes that are essential for parasite virulence. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **21**, 720–731. (doi:10.1096/fj.06-6818com)

307. Rojas F, Matthews KR. 2019 Quorum sensing in African trypanosomes. *Curr. Opin. Microbiol.* **52**, 124–129. (doi:10.1016/j.mib.2019.07.001)

308. DeMarco SF, Saada EA, Lopez MA, Hill KL. 2020 Identification of positive chemotaxis in the protozoan pathogen *Trypanosoma brucei*. *mSphere* **5**, e00685–20. (doi:10.1128/mSphere.00685-20)

309. Silvester E, Ivens A, Matthews KR. 2018 A gene expression comparison of *Trypanosoma brucei* and *Trypanosoma congolense* in the bloodstream of the mammalian host reveals species-specific adaptations to density-dependent development. *PLoS Negl. Trop. Dis.* **12**, e0006863. (doi:10.1371/journal.pntd.0006863)

310. Awuoche EO, Weiss BL, Mireji PO, Vigneron A, Nyambega B, Murilla G, Aksoy S. 2018 Expression profiling of *Trypanosoma congolense* genes during development in the tsetse fly vector *Glossina morsitans morsitans*. *Parasit. Vectors* **11**, 380. (doi:10.1186/s13071-018-2964-8)

311. Stoco PH *et al*. 2014 Genome of the avirulent human-infective trypanosome—*Trypanosoma rangeli*. *PLoS Negl. Trop. Dis.* **8**, e3176. (doi:10.1371/journal.pntd.0003176)

312. Pils B, Schultz J. 2004 Inactive enzyme-homologues find new function in regulatory processes. *J. Mol. Biol.* **340**, 399–404. (doi:10.1016/j.jmb.2004.04.063)

313. Adrain C, Freeman M. 2012 New lives for old: evolution of pseudoenzyme function illustrated by iRhoms. *Nat. Rev. Mol. Cell Biol.* **13**, 489–498. (doi:10.1038/nrm3392)

314. Jackson BC, Thompson DC, Charkoftaki G, Vasiliou V. 2015 Dead enzymes in the aldehyde dehydrogenase gene family: role in drug metabolism and toxicology. *Expert Opin. Drug Metab. Toxicol.* **11**, 1839–1847. (doi:10.1517/17425255.2016.1108406)

315. Kung JE, Jura N. 2016 Structural basis for the non-catalytic functions of protein kinases. *Struct. Lond. Engl.* **1993** 24, 7–24. (doi:10.1016/j.str.2015.10.020)

316. Anantharaman V, Zhang D, Aravind L. 2010 OST-HTH: a novel predicted RNA-binding domain. *Biol. Direct* **5**, 13. (doi:10.1186/1745-6150-5-13)

317. Bekaert M, Conant GC. 2014 Gene duplication and phenotypic changes in the evolution of mammalian metabolic networks. *PLoS ONE* **9**, e87115. (doi:10.1371/journal.pone.0087115)

318. Carnes J et al. 2015 Genome and phylogenetic analyses of *Trypanosoma evansi* reveal extensive similarity to *T. brucei* and multiple independent origins for dyskinetoplasty. *PLoS Negl. Trop. Dis.* **9**, e3404. (doi:10.1371/journal.pntd.0003404)

319. Roy N et al. 2010 Proteomics of *Trypanosoma evansi* infection in rodents. *PLoS ONE* **5**, e9796. (doi:10.1371/journal.pone.0009796)

320. Jackson AP, Goyard S, Xia D, Foth BJ, Sanders M, Wastling JM, Minoprio P, Berriman M. 2015 Global gene expression profiling through the complete life cycle of *Trypanosoma vivax*. *PLoS Negl. Trop. Dis.* **9**, e0003975. (doi:10.1371/journal.pntd.0003975)

321. Durrani H, Hampton M, Rumbley JN, Zimmer SL. 2020 A global analysis of enzyme compartmentalization to glycosomes. *Pathog. Basel Switz.* **9**, 281. (doi:10.3390/pathogens9040281)

322. Misset O, Bos OJ, Opperdoes FR. 1986 Glycolytic enzymes of *Trypanosoma brucei*: simultaneous purification, intraglycosomal concentrations and physical properties. *Eur. J. Biochem.* **157**, 441–453. (doi:10.1111/j.1432-1033.1986.tb09687.x)

323. Rondón-Mercado R, Acosta H, Cáceres AJ, Quiñones W, Concepción JL. 2017 Subcellular localization of glycolytic enzymes and characterization of intermediary metabolism of *Trypanosoma rangeli*. *Mol. Biochem. Parasitol.* **216**, 21–29. (doi:10.1016/j.molbiopara.2017.06.007)

324. Kelly S, Ivens A, Manna PT, Gibson W, Field MC. 2014 A draft genome for the African crocodilian trypanosome *Trypanosoma grayi*. *Sci. Data* **1**, 140024. (doi:10.1038/sdata.2014.24)

325. Martins NO et al. 2015 Molecular characterization of a novel family of *Trypanosoma cruzi* surface membrane proteins (TcSMP) involved in mammalian host cell invasion. *PLoS Negl. Trop. Dis.* **9**, e0004216. (doi:10.1371/journal.pntd.0004216)

326. Vickerman K. 1976 The diversity of the kinetoplastid flagellates. In *Biology of the Kinetoplastida* (eds WHR Lumsden, DA Evans), pp. 1–34. London, UK: Academic Press.

327. Podlipaev S. 2001 The more insect trypanosomatids under study-the more diverse Trypanosomatidae appears. *Int. J. Parasitol.* **31**, 648–652. (doi:10.1016/s0020-7519(01)00139-4)

328. Votýpka J, Kostygov AY, Kraeva N, Grybchuk-Ieremenko A, Tesařová M, Grybchuk D, Lukeš J, Yurchenko V. 2014 *Kentomonas* gen. n., a new genus of endosymbiont-containing trypanosomatids of Strigomonadinae subfam. n. *Protist* **165**, 825–838. (doi:10.1016/j.protis.2014.09.002)

329. de Souza W, Motta MC. 1999 Endosymbiosis in protozoa of the Trypanosomatidae family. *FEMS Microbiol. Lett.* **173**, 1–8. (doi:10.1111/j.1574-6968.1999.tb13477.x)

330. Alves JMP, Voegtly L, Matveyev AV, Lara AM, da Silva FM, Serrano MG, Buck GA, Teixeira MMG, Camargo EP. 2011 Identification and phylogenetic analysis of heme synthesis genes in trypanosomatids and their bacterial endosymbionts. *PLoS ONE* **6**, e23518. (doi:10.1371/journal.pone.0023518)

331. Alves JMP, Serrano MG, Maia da Silva F, Voegtly LJ, Matveyev AV, Teixeira MMG, Camargo EP, Buck GA. 2013 Genome evolution and phylogenomic analysis of *Candidatus* kinetoplastibacterium, the betaproteobacterial endosymbionts of *Strigomonas* and *Angomonas*. *Genome Biol. Evol.* **5**, 338–350. (doi:10.1093/gbe/evt012)

332. Freymuller E, Camargo EP. 1981 Ultrastructural differences between species of trypanosomatids with and without endosymbionts. *J. Protozool.* **28**, 175–182. (doi:10.1111/j.1550-7408.1981.tb02829.x)

333. Motta MC, Soares MJ, Attias M, Morgado J, Lemos AP, Saad-Nehme J, Meyer-Fernandes JR, De Souza W. 1997 Ultrastructural and biochemical analysis of the relationship of *Crithidia deanei* with its endosymbiont. *Eur. J. Cell Biol.* **72**, 370–377.

334. Loyola-Machado AC, Azevedo-Martins AC, Catta-Preta CMC, de Souza W, Galina A, Motta MCM. 2017 The symbiotic bacterium fuels the energy metabolism of the host trypanosomatid *Strigomonas culicis*. *Protist* **168**, 253–269. (doi:10.1016/j.protis.2017.02.001)

335. Azevedo-Martins AC et al. 2015 Mitochondrial respiration and genomic analysis provide insight into the influence of the symbiotic bacterium on host trypanosomatid oxygen consumption. *Parasitology* **142**, 352–362. (doi:10.1017/S0031182014001139)

336. Brunoro GVF et al. 2019 Quantitative proteomic map of the trypanosomatid *Strigomonas culicis*: the biological contribution of its endosymbiotic bacterium. *Protist* **170**, 125698. (doi:10.1016/j.protis.2019.125698)

337. Alcolea PJ, Alonso A, García-Tabares F, Mena MC, Ciordia S, Larraga V. 2016 Increased abundance of proteins involved in resistance to oxidative and nitrosative stress at the last stages of growth and development of *Leishmania amazonensis* promastigotes revealed by proteome analysis. *PLoS ONE* **11**, e0164344. (doi:10.1371/journal.pone.0164344)

338. Srinivasan S, Aggarwal S, Raghunathan V. 2018 Theoretical structural model of the glycolytic enzyme phosphoglycerate kinase C from *Leishmania mexicana mexicana*: transition from a continuous to discontinuous helix at the C-terminus and the role of GXXXG motif. *Allergy Drugs Clin. Immunol.* **2**, 109.

339. Quiñones W, Cáceres AJ, Ruiz MT, Concepción JL. 2015 Glycosomal membrane proteins and lipids from *Leishmania mexicana*. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **182**, 27–36. (doi:10.1016/j.cbpb.2014.11.012)

340. Quiñones W, Acosta H, Gonçalves CS, Motta MCM, Gualdrón-López M, Michels PAM. 2020 Structure, properties, and function of glycosomes in *Trypanosoma cruzi*. *Front. Cell. Infect. Microbiol.* **10**, 25. (doi:10.3389/fcimb.2020.00025)

341. Blum JJ. 1993 Intermediary metabolism of *Leishmania*. *Parasitol. Today* **9**, 118–122. (doi:10.1016/0169-4758(93)90168-F)

342. Hart DT, Coombs GH. 1982 Leishmania mexicana: energy metabolism of amastigotes and promastigotes. *Exp. Parasitol.* **54**, 397–409. (doi:10.1016/0014-4894(82)90049-2)

343. Saunders EC, Naderer T, Chambers J, Landfear SM, McConville MJ. 2018 *Leishmania mexicana* can utilize amino acids as major carbon sources in macrophages but not in animal models. *Mol. Microbiol.* **108**, 143–158. (doi:10.1111/mmi.13923)

344. Acosta H, Yuan M, Cáceres AJ, Quiñones WA, Concepción JL, Michels PAM. 2017 Chapter 15: Carbon metabolism as a drug target in *Leishmania*. In *Drug discovery for Leishmaniasis* (eds L Rivas, C Gil), pp. 297–315. London, UK: Royal Society of Chemistry. (doi:10.1039/9781788010177-00297)

345. Staudt C, Puissant E, Boonen M. 2016 Subcellular trafficking of mammalian lysosomal proteins: an extended view. *Int. J. Mol. Sci.* **18**, 47. (doi:10.3390/ijms18010047)

346. Bonifacino JS, Traub LM. 2003 Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annu. Rev. Biochem.* **72**, 395–447. (doi:10.1146/annurev.biochem.72.121801.161800)

347. Rowan BG, Weigel NL, O'Malley BW. 2000 Phosphorylation of steroid receptor coactivator-1. Identification of the phosphorylation sites and phosphorylation through the mitogen-activated protein kinase pathway. *J. Biol. Chem.* **275**, 4475–4483. (doi:10.1074/jbc.275.6.4475)

348. Kollara A, Brown TJ. 2010 Variable expression of nuclear receptor coactivator 4 (NcoA4) during mouse embryonic development. *J. Histochem. Cytochem. Off. J. Histochem. Soc.* **58**, 595–609. (doi:10.1369/jhc.2010.955294)

349. Mancias JD, Wang X, Gygi SP, Harper JW, Kimmelman AC. 2014 Quantitative proteomics identifies NCOA4 as the cargo receptor mediating ferritinophagy. *Nature* **509**, 105–109. (doi:10.1038/nature13148)

350. Yu JS, Noll KM. 1995 The hyperthermophilic bacterium *Thermotoga neapolitana* possesses two isozymes of the 3-phosphoglycerate kinase/triosephosphate isomerase fusion protein. *FEMS Microbiol. Lett.* **131**, 307–312. (doi:10.1111/j.1574-6968.1995.tb07792.x)

351. Sienkiewicz N, Jarosławski S, Wyllie S, Fairlamb AH. 2008 Chemical and genetic validation of dihydrofolate reductase-thymidylate synthase as a drug target in African trypanosomes. *Mol. Microbiol.* **69**, 520–533. (doi:10.1111/j.1365-2958.2008.06305.x)

352. Landau MJ, Sharma H, Anderson KS. 2013 Selective peptide inhibitors of bifunctional thymidylate synthase-dihydrofolate reductase from *Toxoplasma gondii* provide insights into domain-domain communication and allosteric regulation. *Protein Sci. Publ. Protein Soc.* **22**, 1161–1173. (doi:10.1002/pro.2300)

353. Anderson KS. 2017 Understanding the molecular mechanism of substrate channeling and domain

communication in protozoal bifunctional TS-DHFR. *Protein Eng. Des. Sel.* **30**, 253–261. (doi:10.1093/protein/gzx004)

354. Chevalier N, Bertrand L, Rider MH, Opperdoes FR, Rigden DJ, Michels PAM. 2005 6-Phosphofructo-2-kinase and fructose-2,6-bisphosphatase in Trypanosomatidae: molecular characterization, database searches, modelling studies and evolutionary analysis. *FEBS J.* **272**, 3542–3560. (doi:10.1111/j.1742-4658.2005.04774.x)

355. Merzendorfer H. 2006 Insect chitin synthases: a review. *J. Comp. Physiol. [B]* **176**, 1–15. (doi:10.1007/s00360-005-0005-3)

356. Bringaud F, Barrett MP, Zilberstein D. 2012 Multiple roles of proline transport and metabolism in trypanosomatids. *Front. Biosci. Landmark Ed.* **17**, 349–374. (doi:10.2741/3931)

357. Lake JA, de la Cruz VF, Ferreira PC, Morel C, Simpson L. 1988 Evolution of parasitism: kinetoplastid protozoan history reconstructed from mitochondrial rRNA gene sequences. *Proc. Natl Acad. Sci. USA* **85**, 4779–4783. (doi:10.1073/pnas.85.13.4779)

358. Alcolea PJ, Alonso A, García-Tabares F, Toraño A, Larraga V. 2014 An insight into the proteome of Crithidia fasciculata choanomastigotes as a comparative approach to axenic growth, peanut lectin agglutination and differentiation of Leishmania spp. promastigotes. *PLoS ONE* **9**, e113837. (doi:10.1371/journal.pone.0113837)

359. Alcolea PJ, Alonso A, Gómez MJ, Moreno I, Domínguez M, Parro V, Larraga V. 2010 Transcriptomics throughout the life cycle of Leishmania infantum: high down-regulation rate in the amastigote stage. *Int. J. Parasitol.* **40**, 1497–1516. (doi:10.1016/j.ijpara.2010.05.013)

360. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2013 GenBank. *Nucleic Acids Res.* **41**, D36–D42. (doi:10.1093/nar/gks1195)

361. Saitou N, Nei M. 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425. (doi:10.1093/oxfordjournals.molbev.a040454)

362. Felsenstein J. 1985 Confidence limits on phylogenies: an approach using the bootstrap. *Evol. Int. J. Org. Evol.* **39**, 783–791. (doi:10.1111/j.1558-5646.1985.tb00420.x)

363. Nei M, Kumar S, Nei EPP of BM. 2000 *Molecular evolution and phylogenetics*. Oxford, UK: Oxford University Press.

364. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018 MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549. (doi:10.1093/molbev/msy096)

365. Jackson AP. 2015 Preface. The evolution of parasite genomes and the origins of parasitism. *Parasitology* **142**(Suppl. 1), S1–S5. (doi:10.1017/S0031182014001516)

366. Storz JF. 2016 Gene duplication and evolutionary innovations in hemoglobin-oxygen transport. *Physiology (Bethesda, MD)* **31**, 223–232. (doi:10.1152/physiol.00060.2015)

367. Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R. 2013 Gene duplication as a major force in evolution. *J. Genet.* **92**, 155–161. (doi:10.1007/s12041-013-0212-8)

368. Reams AB, Roth JR. 2015 Mechanisms of gene duplication and amplification. *Cold Spring Harb. Perspect. Biol.* **7**, a016592. (doi:10.1101/cshperspect.a016592)

369. Reis-Cunha JL et al. 2015 Chromosomal copy number variation reveals differential levels of genomic plasticity in distinct Trypanosoma cruzi strains. *BMC Genomics* **16**, 499. (doi:10.1186/s12864-015-1680-4)

370. Harding T, Simpson AGB. 2018 Recent advances in halophilic protozoa research. *J. Eukaryot. Microbiol.* **65**, 556–570. (doi:10.1111/jeu.12495)

371. Baskaran P, Jaleta TG, Streit A, Rödelsperger C. 2017 Duplications and positive selection drive the evolution of parasitism-associated gene families in the nematode Strongyloides papillosus. *Genome Biol. Evol.* **9**, 790–801. (doi:10.1093/gbe/evx040)

372. Wang S, Zhu X-Q, Cai X. 2017 Gene duplication analysis reveals no ancient whole genome duplication but extensive small-scale duplications during genome evolution and adaptation of Schistosoma mansoni. *Front. Cell. Infect. Microbiol.* **7**, 412. (doi:10.3389/fcimb.2017.00412)

373. Cwiklinski K, Dalton JP, Dufresne PJ, La Course J, Williams DJ, Hodgkinson J, Paterson S. 2015 The Fasciola hepatica genome: gene duplication and polymorphism reveals adaptation to the host environment and the capacity for rapid evolution. *Genome Biol.* **16**, 71. (doi:10.1186/s13059-015-0632-2)

374. Hunt VL et al. 2016 The genomic basis of parasitism in the Strongyloides clade of nematodes. *Nat. Genet.* **48**, 299–307. (doi:10.1038/ng.3495)

375. Moyle M et al. 1994 A hookworm glycoprotein that inhibits neutrophil function is a ligand of the integrin CD11b/CD18. *J. Biol. Chem.* **269**, 10 008–10 015.

376. Williamson AL et al. 2006 Ancylostoma caninum MTP-1, an astacin-like metalloprotease secreted by infective hookworm larvae, is involved in tissue migration. *Infect. Immun.* **74**, 961–967. (doi:10.1128/IAI.74.2.961-967.2006)

377. Berriman M et al. 2009 The genome of the blood fluke Schistosoma mansoni. *Nature* **460**, 352–358. (doi:10.1038/nature08160)

378. McVeigh P, Cwiklinski K, Garcia-Campos A, Mulcahy G, O'Neill SM, Maule AG, Dalton JP. 2018 In silico analyses of protein glycosylating genes in the helminth Fasciola hepatica (liver fluke) predict protein-linked glycan simplicity and reveal temporally-dynamic expression profiles. *Sci. Rep.* **8**, 11700. (doi:10.1038/s41598-018-29673-3)

379. Pandey T et al. 2020 Draft genome of the liver fluke Fasciola gigantica. *ACS Omega* **5**, 11 084–11 091. (doi:10.1021/acsomega.0c00980)

380. DeBarry JD, Kissinger JC. 2014 A survey of innovation through duplication in the reduced genomes of twelve parasites. *PLoS ONE* **9**, e99213. (doi:10.1371/journal.pone.0099213)

381. Shabardina V, Kischka T, Kmita H, Suzuki Y, Makałowski W. 2018 Environmental adaptation of Acanthamoeba castellanii and Entamoeba histolytica at genome level as seen by comparative genomic analysis. *Int. J. Biol. Sci.* **14**, 306–320. (doi:10.7150/ijbs.23869)

382. Sun J, Jiang H, Flores R, Wen J. 2010 Gene duplication in the genome of parasitic Giardia lamblia. *BMC Evol. Biol.* **10**, 49. (doi:10.1186/1471-2148-10-49)

383. Elmendorf HG, Rohrer SC, Khoury RS, Bouttenot RE, Nash TE. 2005 Examination of a novel head-stalk protein family in Giardia lamblia characterised by the pairing of ankyrin repeats and coiled-coil domains. *Int. J. Parasitol.* **35**, 1001–1011. (doi:10.1016/j.ijpara.2005.03.009)

384. Harding T, Roger AJ, Simpson AGB. 2017 Adaptations to high salt in a halophilic protist: differential expression and gene acquisitions through duplications and gene transfers. *Front. Microbiol.* **8**, 944. (doi:10.3389/fmicb.2017.00944)

385. Aury J-M et al. 2006 Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia. *Nature* **444**, 171–178. (doi:10.1038/nature05230)

386. Muller HJ. 1964 The relation of recombination to mutational advance. *Mutat. Res.* **106**, 2–9. (doi:10.1016/0027-5107(64)90047-8)

387. Reis-Cunha JL, Valdivia HO, Bartholomeu DC. 2018 Gene and chromosomal copy number variations as an adaptive mechanism towards a parasitic lifestyle in trypanosomatids. *Curr. Genomics* **19**, 87–97. (doi:10.2174/1389202918666170911161311)

388. Sinha R, Das S, Das S, Shadab M, Chowdhury R, Tripathy S, Ali N. 2018 Genome plasticity in cultured Leishmania donovani: comparison of early and late passages. *Front. Microbiol.* **9**, 1279. (doi:10.3389/fmicb.2018.01279)

389. Rogers MB et al. 2011 Chromosome and gene copy number variation allow major structural change between species and strains of Leishmania. *Genome Res.* **21**, 2129–2142. (doi:10.1101/gr.122945.111)

390. Laffitte M-CN, Leprohon P, Papadopoulou B, Ouellette M. 2016 Plasticity of the Leishmania genome leading to gene copy number variations and drug resistance. *F1000Research* **5**, 2350. (doi:10.12688/f1000research.9218.1)

391. de Souza DAS, Pavoni DP, Krieger MA, Ludwig A. 2018 Evolutionary analyses of myosin genes in trypanosomatids show a history of expansion, secondary losses and neofunctionalization. *Sci. Rep.* **8**, 1376. (doi:10.1038/s41598-017-18865-y)

392. da Silva MS, Hovel-Miner GA, Briggs EM, Elias MC, McCulloch R. 2018 Evaluation of mechanisms that may generate DNA lesions triggering antigenic variation in African trypanosomes. *PLoS Pathog.* **14**, e1007321. (doi:10.1371/journal.ppat.1007321)

393. Salmon D, Bachmaier S, Krumbholz C, Kador M, Gossmann JA, Uzureau P, Pays E, Boshart M. 2012 Cytokinesis of Trypanosoma brucei bloodstream forms depends on expression of adenylyl cyclases of the ESAG4 or ESAG4-like subfamily. *Mol. Microbiol.* **84**, 225–242. (doi:10.1111/j.1365-2958.2012.08013.x)

394. Marcello L, Barry JD. 2007 Analysis of the VSG gene silent archive in *Trypanosoma brucei* reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure. *Genome Res.* **17**, 1344–1352. (doi:10.1101/gr.6421207)

395. Gu X. 2003 Evolution of duplicate genes versus genetic robustness against null mutations. *Trends Genet. TIG* **19**, 354–356. (doi:10.1016/S0168-9525(03)00139-2)

396. Kuepfer L, Sauer U, Blank LM. 2005 Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Res.* **15**, 1421–1430. (doi:10.1101/gr.3992505)

397. Gu Z, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li W-H. 2003 Role of duplicate genes in genetic robustness against null mutations. *Nature* **421**, 63–66. (doi:10.1038/nature01198)

398. Keane OM, Toft C, Carretero-Paulet L, Jones GW, Fares MA. 2014 Preservation of genetic and regulatory robustness in ancient gene duplicates of *Saccharomyces cerevisiae*. *Genome Res.* **24**, 1830–1841. (doi:10.1101/gr.176792.114)

399. Lempe J, Lachowiec J, Sullivan AM, Queitsch C. 2013 Molecular mechanisms of robustness in plants. *Curr. Opin. Plant Biol.* **16**, 62–69. (doi:10.1016/j.pbi.2012.12.002)

400. Tomaiuolo M, Bertram R, Houle D. 2008 Enzyme isoforms may increase phenotypic robustness. *Evol. Int. J. Org. Evol.* **62**, 2884–2893. (doi:10.1111/j.1558-5646.2008.00483.x)

401. Xia D *et al*. 2008 The proteome of Toxoplasma gondii: integration with the genome provides novel insights into gene expression and annotation. *Genome Biol.* **9**, R116. (doi:10.1186/gb-2008-9-7-r116)

402. Blume M, Nitzsche R, Sternberg U, Gerlic M, Masters SL, Gupta N, McConville MJ. 2015 A *Toxoplasma gondii* gluconeogenic enzyme contributes to robust central carbon metabolism and is essential for replication and virulence. *Cell Host Microbe* **18**, 210–220. (doi:10.1016/j.chom.2015.07.008)

403. Saito T, Nishi M, Lim MI, Wu B, Maeda T, Hashimoto H, Takeuchi T, Roos DS, Asai T. 2008 A novel GDP-dependent pyruvate kinase isozyme from *Toxoplasma gondii* localizes to both the apicoplast and the mitochondrion. *J. Biol. Chem.* **283**, 14 041–14 052. (doi:10.1074/jbc.M709015200)

404. Maeda T, Saito T, Harb OS, Roos DS, Takeo S, Suzuki H, Tsuboi T, Takeuchi T, Asai T. 2009 Pyruvate kinase type-II isozyme in *Plasmodium falciparum* localizes to the apicoplast. *Parasitol. Int.* **58**, 101–105. (doi:10.1016/j.parint.2008.10.005)

405. Xia N *et al*. 2019 Pyruvate homeostasis as a determinant of parasite growth and metabolic plasticity in *Toxoplasma gondii*. *mBio* **10**, e00898–19. (doi:10.1128/mBio.00898-19)

406. Millard P, Smallbone K, Mendes P. 2017 Metabolic regulation is sufficient for global and robust coordination of glucose uptake, catabolism, energy production and growth in *Escherichia coli*. *PLoS Comput. Biol.* **13**, e1005396. (doi:10.1371/journal.pcbi.1005396)

407. Salcedo-Sora JE, Caamano-Gutierrez E, Ward SA, Biagini GA. 2014 The proliferating cell hypothesis: a metabolic framework for *Plasmodium* growth and development. *Trends Parasitol.* **30**, 170–175. (doi:10.1016/j.pt.2014.02.001)

408. Bianconi ME, Dunning LT, Moreno-Villena JJ, Osborne CP, Christin P-A. 2018 Gene duplication and dosage effects during the early emergence of C4 photosynthesis in the grass genus *Alloteropsis*. *J. Exp. Bot.* **69**, 1967–1980. (doi:10.1093/jxb/ery029)

409. Heckmann D, Schulze S, Denton A, Gowik U, Westhoff P, Weber APM, Lercher MJ. 2013 Predicting C4 photosynthesis evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape. *Cell* **153**, 1579–1588. (doi:10.1016/j.cell.2013.04.058)

410. Bekaert M, Edger PP, Pires JC, Conant GC. 2011 Two-phase resolution of polyploidy in the *Arabidopsis* metabolic network gives rise to relative and absolute dosage constraints. *Plant Cell* **23**, 1719–1728. (doi:10.1105/tpc.110.081281)

411. Iantorno SA *et al*. 2017 Gene expression in Leishmania is regulated predominantly by gene dosage. *mBio* **8**, e01393–17. (doi:10.1128/mBio.01393-17)

412. Dumetz F *et al*. 2018 Molecular preadaptation to antimony resistance in *Leishmania donovani* on the Indian Subcontinent. *mSphere* **3**, e00548–17. (doi:10.1128/mSphere.00548-17)

413. Solis-Escalante D, Kuijpers NGA, Barrajon-Simancas N, van den Broek M, Pronk JT, Daran J-M, Daran-Lapujade P. 2015 A minimal set of glycolytic genes reveals strong redundancies in *Saccharomyces cerevisiae* central metabolism. *Eukaryot. Cell* **14**, 804–816. (doi:10.1128/EC.00064-15)

414. Marques AC, Vinckenbosch N, Brawand D, Kaessmann H. 2008 Functional diversification of duplicate genes through subcellular adaptation of encoded proteins. *Genome Biol.* **9**, R54. (doi:10.1186/gb-2008-9-3-r54)

415. Hsiao T-L, Vitkup D. 2008 Role of duplicate genes in robustness against deleterious human mutations. *PLoS Genet.* **4**, e1000014. (doi:10.1371/journal.pgen.1000014)

416. Sinkala M, Mulder N, Patrick Martin D. 2019 Metabolic gene alterations impact the clinical aggressiveness and drug responses of 32 human cancers. *Commun. Biol.* **2**, 414. (doi:10.1038/s42003-019-0666-1)

417. Hughes AL. 2002 Adaptive evolution after gene duplication. *Trends Genet.* **18**, 433–434. (doi:10.1016/s0168-9525(02)02755-5)

418. Zhang J, Zhang Y, Rosenberg HF. 2002 Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat. Genet.* **30**, 411–415. (doi:10.1038/ng852)