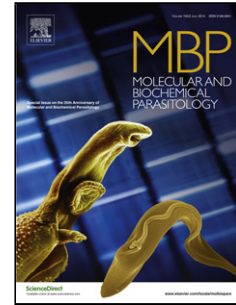


## Accepted Manuscript

Title: Structure and function of Per-ARNT-Sim domains and their possible role in the life-cycle biology of *Trypanosoma cruzi*

Authors: Maura Rojas-Pirela, Daniel J. Rigden, Paul A. Michels, Ana J. Cáceres, Juan Luis Concepción, Wilfredo Quiñones



PII: S0166-6851(17)30135-4  
DOI: <https://doi.org/10.1016/j.molbiopara.2017.11.002>  
Reference: MOLBIO 11096

To appear in: *Molecular & Biochemical Parasitology*

Received date: 30-4-2017  
Revised date: 12-10-2017  
Accepted date: 2-11-2017

Please cite this article as: Rojas-Pirela Maura, Rigden Daniel J, Michels Paul A, Cáceres Ana J, Concepción Juan Luis, Quiñones Wilfredo. Structure and function of Per-ARNT-Sim domains and their possible role in the life-cycle biology of *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology* <https://doi.org/10.1016/j.molbiopara.2017.11.002>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Review****Structure and function of Per-ARNT-Sim domains and their possible role in the life-cycle biology of *Trypanosoma cruzi***

Maura Rojas-Pirela<sup>a</sup>, Daniel J. Rigden<sup>b</sup>, Paul A. Michels<sup>c</sup>, Ana J. Cáceres<sup>a</sup>, Juan Luis Concepción<sup>a</sup>, Wilfredo Quiñones<sup>a,\*</sup>

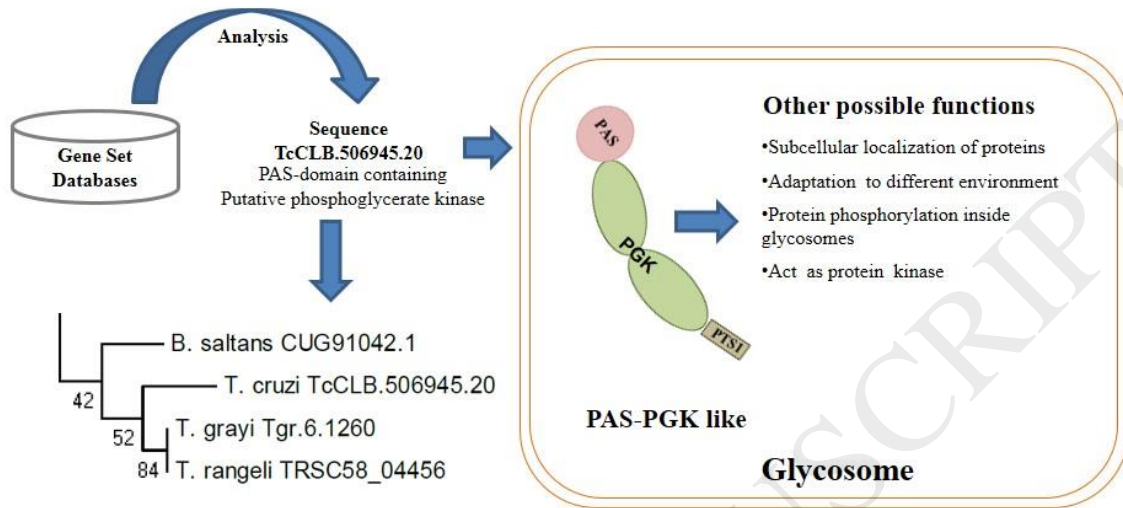
<sup>a</sup> Laboratorio de Enzimología de Parásitos, Departamento de Biología, Facultad de Ciencias, Universidad de Los Andes, Mérida 5101, Venezuela

<sup>b</sup> Institute of Integrative Biology, University of Liverpool, Liverpool, L69 7ZB, United Kingdom.

<sup>c</sup> Centre for Immunity, Infection and Evolution and Centre for Translational and Chemical Biology, School of Biological Sciences, The University of Edinburgh, The King's Buildings, Edinburgh EH9 3FL, Scotland, United Kingdom

\* Corresponding author. Laboratorio de Enzimología de Parásitos. Facultad de Ciencias. Universidad de Los Andes. La Hechicera, 5101-Mérida, Venezuela. Tel.: +58 274 2401302. Fax: +58 274 2401291. *E-mail address*: wilqui@ula.ve

**Graphicalabstract**



## HIGHLIGHTS

- PAS domains are ubiquitous in nature and endow multiple functions to proteins
- Analysis of kinetoplastid genomes revealed PAS domains in various protein kinases
- A PAS-phosphoglycerate kinase (PGK)-like protein was identified in *T. cruzi*
- The PAS-PGK-like protein contains a PTS1 sequence for import into glycosomes

## Abstract

Per-ARNT-Sim (PAS) domains of proteins play important roles as modules for signalling and cellular regulation processes in widely diverse organisms such as Archaea, Bacteria, protists, plants, yeasts, insects and vertebrates. These domains are present in many proteins where they are used as sensors of stimuli and modules for protein interactions. Characteristically, they can bind a broad spectrum of molecules. Such binding causes the domain to trigger a specific cellular response or to make the protein containing the domain susceptible to responding to additional physical or chemical signals. Different PAS proteins have the ability to sense redox potential, light, oxygen, energy levels, carboxylic acids, fatty acids and several other stimuli. Such proteins have been found to be involved in cellular processes such as development, virulence, sporulation, adaptation to hypoxia, circadian cycle, metabolism and gene regulation and expression. Our analysis of the genome of different kinetoplastid species revealed the presence of PAS domains also in different predicted kinases from these protists. Open-reading frames coding for these PAS-kinases are unusually large. In addition, the products of these genes appear to contain in their structure combinations of domains uncommon in other eukaryotes. The physiological significance of PAS domains in these parasites, specifically in *Trypanosoma cruzi*, is discussed.

## 1. Introduction

Protein domains are parts of the polypeptide sequence that can evolve, fold into a three-dimensional structure, and function independently from the rest of the protein. Domain sizes can vary widely; however, the majority of domains comprise between 100 and 200 residues, with the most common size being about 100 residues [1]. Protein domains are present in the three domains of life: archaea, bacteria and eukaryotes. In the case of eukaryotes, many proteins possess one or multiple domains, with different architectures [2]. It has been estimated that there are at least 1200 families of protein domains [3]. Many of these families have specific roles in diverse cellular processes, such as apoptosis, modulation of the cytoskeleton, vesicle trafficking, DNA binding, protein-protein interactions and regulation of intercellular or intracellular signalling. However, some families are considered to contain “promiscuous domains” or “versatile domains”. Moreover, the possibility of combining different domains in a single polypeptide provides proteins with the ability to be involved in broad spectra of processes that are key in interaction networks in the cell, especially those that contribute to signal translation [4,5].

In eukaryotes about 215 promiscuous protein domains have been identified [4], which include the Per-ARNT-Sim (PAS) domains. The versatile PAS domain is present in all kingdoms of life [6]. PAS is an acronym derived from the three eukaryotic proteins in which the domain was first recognized by sequence homology: the period circadian protein (**P**er), the vertebrate aryl hydrocarbon receptor nuclear translocator (**A**RNT) and single-minded (**S**im), a *Drosophila* protein involved in embryonic development [7]. One of the first structural studies of a PAS domain involved the photoactive yellow protein (PYP), a mediator of the phototactic negative response by the phototrophic bacterium *Ectothiohalobium halophilum* [8]. This globular protein is considered to contain the prototype structure of a PAS domain [9,10] (Fig. 1).

During the last few years, studies of proteins from all domains of life have resulted in 21,000 entries annotated as PAS domain in the Pfam database [6]. The more than 200 proteins having a PAS domain comprise receptors, signal transducers, kinases, transcription factors, ion channels, chemotaxis proteins, cyclic nucleotide phosphodiesterases and proteins involved in embryological development of the central nervous system [11]. In this paper, we review the known functionality of PAS domains as a prelude to considering the

enigmatic role of this domain in the biology of trypanosomatid parasites, with particular emphasis on *Trypanosoma cruzi*, since the presence of these regulatory domains in parasitic protists was so far unknown.

## 2. Localization and structure of PAS domains

PAS domain proteins have different subcellular locations. The domain may be contained in proteins present in the plasma membrane, exposed to either the intra- or the extracellular environment [11]. Membrane proteins with PAS domains (Aer, FixL, ArcB and DcuS) have been previously studied [12-16]. In all such membrane proteins analysed until now, the PAS domain is located adjacent to a transmembrane region. Possibly, such a location allows the PAS domain to interact with other modules present in the same or other, adjacent membrane proteins [17]. PAS domains can also be part of soluble cytoplasmic proteins (such as NifL, guanylyl cyclase and histidine kinase (HK), photoactive yellow protein (PYP) and transcription factors). These proteins may have a single or multiple PAS domains in their structure [18-21].

In all these proteins, the PAS domain functions as a sensor of changes or various stimuli in its environment (oxygen, pH, light, ions, glucose, voltage, oligomerisation state, carboxylic acids, energy level or fatty acids) and as modulator of protein-protein interactions, allowing organisms to “sense” and respond appropriately to the changes or stimuli [22-24]. These domains are often associated with other regulatory modules in multi-domain proteins. Moreover, a PAS protein can contain a single or multiple, tandemly-organized PAS domains [11]. Some studies suggest that PAS domains in tandem can regulate recruitment and oligomerisation state of proteins [25,26]. The presence of a PAS domain in a protein can confer association specificity of its effector domain. In prokaryotes, PAS domains have been shown to induce the formation of homodimers [25], and in eukaryotes the formation of heterodimers [24]. Oligomerisation is a necessary requirement for the function of many proteins. This process is crucial to trigger and regulate different physiological processes, such as gene expression, activity of enzymes and cell-cell adhesion [27]. In the case of some PAS proteins with enzymatic activity, such as the histidine kinases, dimerisation is a prerequisite to achieve phosphorylation in *trans* [11]. In addition, the PAS domain can

mediate the oligomerisation of transcription factors such as bHLH/PAS, and it can confer specificity and distinct recognition of target genes [21].

Initially, PAS domains were considered homologous regions of about 50 amino acids in the Per, ARNT and Sim proteins [28]. However, subsequent studies revealed that PAS domains can also be formed by regions comprising approximately 100 to 150 amino acids [10]. PAS domains adopt a globular structure formed by several  $\alpha$ -helices and antiparallel  $\beta$ -sheets. The role of each segment in this structure is discussed in [9]. Both the N- and C-terminus have different helices, called flanking helices, which can vary in length and sequence. These helices are involved in protein-protein interactions and their residues can contribute to the formation of bonds with ligands. Taylor *et al.* [28], when analysing a sequence alignment of 300 PAS domains, found that each of the elements or segments of secondary structure are retained along the alignment. They concluded that it is very likely that all PAS domains have the PAS core, a helical connector, and a  $\beta$ -scaffold as structural elements (Fig. 1). The  $\beta$ -scaffold is the most conserved part, both with regard to its amino-acid sequence and the number of sheets, making this a defining feature of a PAS domain [11]. In addition, the  $\alpha$ -helix, which is attached to the C-terminal end of the PAS domain, can link the domain to another protein module.

Although PAS domains from different proteins vary in sequence, orientation, length, and number of  $\alpha$ -helices, they maintain a fairly conserved three-dimensional (3D) structure [29].

### 3. Ligand binding in the PAS domain family

Detection of a specific stimulus by a PAS domain is determined, in most cases, by the presence of a specific structure in the ligand. In addition to the chromophore of the PYP protein, the *p*-coumaric acid, other molecules have been identified as ligands of PAS domains. These include cofactors such as flavin adenine dinucleotide (FAD), for example in *Escherichia coli*, where the protein Aer is a membrane flavoprotein having FAD bound to its N-terminal PAS domain. The FAD acts as a redox sensor and mediates aerotaxis, a positive aerotactic response [30]. *Klebsiella pneumoniae* and *Azotobacter vinelandii* possess a histidine kinase with a FAD-binding PAS domain that regulates the expression of

genes involved in nitrogen fixation (*nif* genes) through the interaction with the DNA-binding protein NifA [31]. Many proteins contain PAS domains belonging to a sub-class that use flavin mononucleotide (FMN) as sensing intermediate. These domains are known as LOV domains, because they have the ability to respond to stimuli such as light, oxygen and voltage [22]. LOV proteins non-covalently bind the FMN, and then, in response to a stimulus (such as blue light), a covalent bond is formed between a conserved cysteine residue and the FMN. This covalent bond induces a structural change in the PAS domain that is propagated to the rest of the protein [32].

The gas-sensing PAS domains (Heme-PAS) can join heme groups of both type b and c [33,34]. Heme-PAS are commonly found in three types of proteins, the histidine kinase protein, phosphodiesterase guanylyl cyclase, and transcription factors of type bHLH [35-38].

In bacteria, several chemoreceptors of carboxylic acids have been identified. *Pseudomonas putida* KT2440 possess a chemoreceptor with a PAS domain that recognizes C<sub>2</sub> and C<sub>3</sub> carboxylic acids and so mediates taxis for these compounds [39]. *E. coli* also have an integral membrane sensor for carboxylic acids, the histidine kinase DcuS [16,40], that senses C<sub>4</sub> and C<sub>6</sub> carboxylates through a periplasmically exposed PAS domain. DcuS is closely related to sensor kinase CitA of *K. pneumoniae*, that is part of a two-component system regulating the transport and metabolism of citrate [40,41]. In *Rhizobium meliloti*, DctB is a transmembrane sensor histidine kinase that phosphorylates DctD in response to binding carboxylic acid, and controls the expression of another integral membrane protein [42]. Unlike DcuS and CitA, the DctB structure has two tandemly-arranged periplasmic PAS domains, and uses different backbone and side-chain interactions [42].

In prokaryotes, fatty acid-binding PAS domains are commonly associated to proteins having a key role in the virulence of pathogenic bacteria. In *Mycobacterium tuberculosis*, oleic and palmitic acid can function as ligands of PAS domain protein RV1364c, a multidomain protein which regulates the stress-dependent regulatory factor  $\delta$  ( $\delta$ F) [43]. Furthermore, in *Xanthomonas campestris* the perception of *cis*-2-unsaturated fatty acids by a PAS domain regulates expression of a subset of genes that contribute to virulence (for example, genes involved in the type IV secretion system) and motility. It is the domain of the RpfC/RpfG two-component system that is implicated in sensing these fatty acids [44].



Fatty acid-binding PAS domains have also been reported for human cells; unsaturated C18 fatty acids such as oleic and linoleic acid are high-affinity ligands of the PAS domain in the hypoxia-inducible factor 3 $\alpha$  (HIF-3 $\alpha$ ) [44]. HIF transcription factors are  $\alpha$ - $\beta$  dimers that mediate the expression of genes involved in the cellular response to hypoxia. In these factors, the PAS domain has a crucial function in forming active HIF heterodimers and recruiting co-regulators. For HIF-3 $\alpha$ , the fatty acids can have a role as structural cofactors; their binding by the PAS domain can be necessary to acquire a stable structural conformation which allows its translocation to the nucleus and perform its function (as with other proteins of the same family) [44].

It has been found that PAS domains can also bind divalent metals, as occurs in some pathogenic bacteria such as *Salmonella typhimurium* that sense these metals in the host environment. Certain concentrations of divalent cations (such as Mg<sup>2+</sup> and Ca<sup>2+</sup>) promote remodelling of the bacterial envelope and activation of genes associated with virulence, including intracellular survival, invasion, phagosome alteration, acid stress and cationic antimicrobial peptides resistance [45]. All these virulence properties are regulated by the PhoQ/PhoP signal transduction system. The membrane protein PhoQ is an Mg<sup>2+</sup> and Ca<sup>2+</sup> sensing histidine kinase. PhoQ has a periplasmic PAS domain rich in acidic residues, which is involved in the binding of, among others, these divalent ions [46,47].

In *Saccharomyces cerevisiae*, the two PAS kinases, Psk1 and Psk2, have been widely studied. They are activated for two different pathways, the cell integrity stress pathway and the glucose repression pathway [48]. For Psk1, the activation also occurs in response to non-fermentative carbon sources, through direct phosphorylation of Snf1, the master of the fermentation/respiration switch [23]. Snf1 is activated by phosphorylation, through the formation of a complex with three other kinases; SaK1, Tos3 and Elm1. Once phosphorylated, Snf1 activates Psk1/Psk2 through phosphorylation in its kinase domain [49-51].

The mammalian PAS kinases are enzymes involved in the expression of genes for insulin in pancreatic beta cells and for regulating glucose homeostasis in peripheral tissues in mice [45]. Four mammalian PAS kinase substrates have been reported, namely pancreatic duodenal homeobox-1 (Pdx-1) [52], glycogen synthase (Gsy) [53], eukaryotic translation elongation factor 1A1 (eEF1A1) [54], and ribosomal protein S6 (S6) [55]. In yeast, five

substrates of Psk have been identified: UDP-glucose pyrophosphorylase (Ugp1), Cap Associated Factor (Caf20), Translation Initiation Factor (Tif11 (eIF1A)), Suppressor Sro9 of Rho3, and glycogen synthase (Gsy2) [56]. The activated Psk1 induces the activation of the poly(A)-binding protein 1 (Pbp1) by phosphorylation, which then inhibits the target of rapamycin complex 1 (TORC1) through sequestration at stress granules [57]. In both yeast and mammals, the association of small metabolites to the PAS domain could induce a disruption of the interaction between the PAS domain and the kinase domain, subsequently causing an activation of the kinase [20].

In *Bacillus subtilis* a PAS histidine kinase is required for the initiation of sporulation in response to nutrient depletion. The kinase KinA has a N-terminal half composed of three PAS domains in tandem (named PAS<sub>A</sub>, PAS<sub>B</sub> and PAS<sub>C</sub>), acting together as a sensor module that is critical for triggering kinase activity [58]. Each PAS domain has a specific function in the protein. It has been shown that domain PAS<sub>A</sub> binds ATP and catalyses the exchange of a phospho group between ATP and nucleoside diphosphates [59]. This hydrolysis reaction drives the conformational changes that activate or deactivate a KinA. Apparently, the PAS<sub>A</sub> domain possesses a nucleotide-diphosphate kinase (NDPK)-like activity. NDPK enzymes are phosphotransferases, which catalyse the transfer of the  $\gamma$ -phospho group from a (deoxy)nucleoside triphosphate (as ATP or GTP) to a (deoxy)nucleoside diphosphate (as ADP or GDP) [60]. For pathogenic protists such as trypanosomatids, it has been postulated that phosphotransferases control communication between spatially separated pathways of consumption and production of ATP [61]. In KinA, the ATP bound to the PAS<sub>A</sub> domain could probably serve as a mediator of the state of binding of ligands to other signal-sensing domains by driving the conformational change of the kinase. Thus, different ligand molecules could promote or prevent ATP binding or ATP hydrolysis, and therefore the kinase sensor can be activated or inhibited by different signals [59]. Other studies have also revealed the importance of the PAS<sub>A</sub> domain in the kinase activity of KinA [25,62].

In all the sensor kinases documented here, the PAS domain undergoes a conformational change in response to a signal (binding of a small ligand or another protein). This change serves as a signal that is transmitted to the rest of PAS-containing protein, causing its structural reorganisation, which leads to modification of its activity. This conformational

mechanism is what enables the PAS domain to be a versatile signal transducer and orchestrate many cellular processes in diverse organisms. The regulation of proteins by PAS domains can occur through different mechanisms: (a) by regulation of intermolecular interactions. Ligand binding can lead to deployment and rotation of the protein. In addition, once unfolded, the protein can acquire a structure that will allow it to form homodimers or heterodimers [22]; (b) by steric effects/hindrance. The binding of a ligand causes a change in the spatial arrangement of the PAS domain. In the absence of the stimulus, the orientation of the PAS domain protein prevents it from acquiring a conformational state that would allow it to fulfil its function or to interact with other proteins [63]; and (c) through its cellular localization. In this latter case, the PAS domain acts as a cellular localisation module. Some sequences for determining subcellular localisation (such as NLS and NES) may be located within the domain. The exposure of the subcellular localisation sequence located in the PAS domain is dependent on the presence of ligand. The binding of ligand to the PAS domain induces a conformational change in the domain that allows the topogenic sequence to be revealed, so that the protein can be routed to its destination compartment and perform its role [64].

#### **4. PAS domains in diverse organisms**

The PAS domains are widely distributed in proteins from eukaryotes, Bacteria and Archaea. In prokaryotes, they are specifically associated with histidine sensor kinase proteins of two-component regulatory systems (both in simple systems and in phospho-relay systems) [17]. These histidine kinases are involved in the regulation of various processes such as sporulation [58], nitrogen fixation [65], aerotaxis [13], stress response [66], nodulation [67], degradation of hydrocarbons [68], polar organelle development [69] and virulence of pathogenic bacteria [70,71].

In eukaryotic organisms, PAS domains are present in proteins involved in many cellular processes, from the regulation of the biological clock to glucose homeostasis. In plants, they are present in photoreceptors (phytochrome and phototropins [72,73], circadian clock proteins [74] and several transcription factors [75]. These proteins function in different pathways that control the development and response of plants to stress adaptation. In

addition, their presence in circadian clock proteins suggests they might serve as a link between environmental conditions and the biological clock [76].

As in plants, mammalian PAS domains function as regulatory modules in various transcription factors belonging to the bHLH-PAS family, involved in processes such as the circadian rhythm response to xenobiotics and adaptation response to hypoxia [77]. In mammals, the regulation of glucose homeostasis also involves the participation of PAS domains through a mechanism associated with the synthesis and secretion of insulin [20]. This domain is associated with serine/threonine protein kinases, in which the domain acts as a regulation module. Malfunctioning of these PAS kinases (named **PASK**, **PASKIN**, and **PSK**) is associated with various metabolic syndromes, such as type 2 diabetes. In yeasts, these PAS kinases operate very similarly. Moreover, they stimulate in these organisms the partition of glucose towards the biosynthesis of structural carbohydrates in response to certain stimuli such as damage of cell integrity [20].

In other organisms such as nematodes, insects and fish, PAS domains are linked to type bHLH-PAS transcription factors, similar to those found in mammals, and are involved in the same kind of cellular processes [78].

In protist organisms, little is known as yet about the presence of PAS domains. Studies focused on *Dictyostelium discoideum* revealed the presence the two phosphorelay sensor kinases, DhkA and DhkB, which regulate the expression of various genes involved in the differentiation of this organism [79,80]. These proteins are members of the family of two-component signalling systems. Structurally, they consist of a highly conserved kinase domain and a PAS domain that functions as a regulatory response element. Also other genes coding for kinases with a PAS domain have been described in *Dictyostelium* which are involved in osmotic stress response [81]. Upon sequencing the genome of *Paramecium tetraurelia* the presence of a gene coding for a hypothetical membrane protein with a PAS domain was found, however the function of this protein remains unknown [82]. One of the significant findings from annotation of the nuclear genome of *Naegleria* was the identification of more than 50 proteins with PAS domains hinting at significant capacity for environmental perception by this protist [83].

There exists little knowledge as to whether PAS-proteins function in other protists, including relatively well-studied parasitic organisms such as *Plasmodium falciparum*,

*Toxoplasma gondii*, *Cryptosporidium parvum* and kinetoplastids (*Trypanosoma cruzi*, *Trypanosoma brucei*, *Trypanosoma rangeli* and *Leishmania* spp.). This may be due to the near absence of these regulatory modules in their proteomes, different from other organisms like plants. Some authors suggested that the near absence of these domains in obligate parasites is due to their life in a stable environment, where is no necessity for a constant redox sensing [28].

A similar suggestion was made by Galparin *et al.* [84] upon analysing the complete genomes of free-living prokaryotes (or not-obligate parasitic ones) and obligate parasitic prokaryotes. These authors had found that many regulatory domains (such as PAS, GGDEF, EAL and HD-GYP) are very abundant in all free-living bacteria but less so or even almost absent in obligate parasitic bacteria. In *Aquifex aeolicus* and *Helicobacter pylori*, two bacteria with almost the same number of genes but with a free-living and parasitic life style, respectively, they discovered a marked difference in the presence of these domains in their genomes. This could suggest that these domains could be particularly important for detecting or sensing the more diverse environmental stimuli encountered by free-living or non-obligate parasitic bacteria. Extreme cases of the near absence of these regulatory domains are the obligate-parasitic bacteria *Mycoplasma* and *Buchnera*. The minimal genomes of these organisms do not encode any such signalling proteins at all [84]. It seems as if these regulatory domains have no use whatsoever in these organisms, despite having a versatile life style. Another important aspect of this study by Galperin *et al.* [84] is that it revealed that signalling domains are generally less abundant and less evenly distributed in Archaea than they are in Bacteria. This skewed phylogenetic distribution suggests that signal transduction could have emerged in the early evolution of bacteria, with subsequently its mass loss in species developing a parasitic life style and be spread by horizontal transfer between Archaea [84]. A similar scenario could have occurred in unicellular eukaryotes, specifically in protists. However, to search support for this notion, it will be necessary to compare the presence of these regulatory modules in many species of free-living and obligate-parasitic species. In the case of *Dictyostelium* other PAS kinases have been studied, and their importance in the response of this organism to different environmental conditions [81].

The presence of these regulatory domains in parasitic protists was so far unknown. However, our analysis of the genome sequences of various protists that cause important diseases in humans and animals revealed a variety of diverse proteins present in the trypanosomatids with regulatory domains (such as PAS domain) [85]. This will be discussed in the next section.

## 5. PAS domains in proteins of the Kinetoplastea

Kinetoplastea are flagellated protists including free-living organisms as well as parasites of diverse invertebrates, vertebrates and plants. Some of these kinetoplastids cause severe diseases in humans (such as Chagas disease, African trypanosomiasis (or sleeping sickness) and various forms of visceral and cutaneous leishmaniasis) and are transmitted by different insect vectors [86]. These parasitic kinetoplastids belonging to the Trypanosomatidae family (including *Trypanosoma* and *Leishmania* species) have very complex life cycles, involving alternating hosts, during which they undergo drastic morphological and metabolic changes. However, there exists little knowledge of how these organisms sense the environmental changes encountered during the life cycle and respond by simultaneously regulating and coordinating intracellular processes required to adapt to these different environments [87]. There is evidence demonstrating that the drastic morphological and metabolic changes of the parasites manifested in each environment are accompanied by changes of the protein profile [88-91]. However, the signal transduction pathways that mediate these changes remain largely unknown, although progress is being made in the unravelling of pathways such as those involved in differentiation steps of some parasites [92,93]. Some posttranslational changes in proteins (such as phosphorylation) during the parasite developmental cycles have been observed [94-96]. In *Trypanosoma cruzi*, proteins of glycosomes (the kinetoplastids' peroxisome-like organelles containing glycolytic enzymes) such as pyruvate phosphate dikinase (PPDK), undergo phosphorylation and proteolytic cleavage in response to nutritional changes. This post-translational modification of PPDK results in inactivation of the enzyme [97]. Probably, in these parasites, as in other parasitic organisms such as the malaria causing *Plasmodium* spp., phosphorylation plays a crucial role in modulating protein functions and thereby controls various aspects of the

biology of these organisms [98]. This would explain the presence of a large number of eukaryotic protein kinases (ePKs), typical and atypical, in the genome of some trypanosomatids [87]. In the case of *T. cruzi* 190 protein kinase genes have been found in the genome [87]. *Trypanosoma brucei* and *Leishmania major* have 176 and 199 PKs, respectively [87]. This high number of PKs represents approximately 2% of each genome; this suggests a key role for phosphorylation in the biology of these parasites. Many of these kinases belong to the families STE, CMGC and NEK [74]. Several of these kinetoplastid protein kinases have been characterised, notably proteins involved in transducing signals from the surface of the cell to the nucleus [99-104].

Importantly, approximately 8% of PKs in kinetoplastid species analysed are predicted to be catalytically inactive, based on the presence of mutations in residues essential for catalytic activity [87]. In the case of an apparent absence of catalytic activity, these kinases might work through a different mechanism, such as via regulation of protein-protein interactions. There is evidence for “dead” enzymes with unusual functions in metazoa accounting for a rich source of biological regulators [105,106]. *Trypanosoma rangeli*, an avirulent human parasite, has very similar genetic characteristics to human pathogenic trypanosomes; its genome encodes 151 ePKs, which corresponds to 1.94% of its total number of identified coding sequences. Indeed, several of the kinases are predicted to be catalytically inactive [107]. Another remarkable feature of the genomes of the parasites is the scarcity or even complete lack of some accessory domains (such as SH3, SH2, FN-III and immunoglobulin-like domain) present in proteins of other eukaryotic organisms [87,108]. In contrast, the presence of other accessory domains and a different domain architecture of some proteins, were observed [87,107]. The genome of plant pathogens of the genus *Phytomonas* encodes 89 ePKs. This corresponds to 1.39% of its total number of identified coding sequences, similar to the percentage found for other parasitic members of the Kinetoplastea analysed so far [109]. In contrast, a considerably higher number was identified in the free-living kinetoplastid *Bodo saltans* in which the genome encodes 562 PKs, belonging to different kinase families, corresponding to 2.96 % of its total number [110].

PAS domains are also present in kinetoplastids, both in free-living and, in lower numbers, in parasitic ones, where they are primarily linked to protein kinases (Table 1 and Supplementary Information online, Table S1). On the contrary, no PAS domain was

detectable in the genome of *Perkinsela* sp., the organism of kinetoplastid ancestry present as an endosymbiont in *Paramoeba* spp. This difference with the free-living and parasitic kinetoplastids could be interpreted as an indication for the loss of the capacity to sense the environment [111]. Most of the genes encoding the kinetoplastid protein kinases with PAS domains are located in a preserved manner in certain chromosomes of these protists. In addition, several of these genes have unusually large open-reading frames. These protein kinases are characterized by slightly acidic isoelectric points, except BSAL\_46115 in *Bodo saltans*. Additionally, there is a PAS-containing phosphoglycerate kinase (PGK) protein in all parasitic kinetoplastids, but not the free-living *B. saltans*, with a similarly high isoelectric point in each of the parasites [85]. In many of these kinases, the PAS domain is present together with other signalling-related domains (Tables 1 and S1), often in domain combinations that are usually not observed in kinases from other eukaryotes. For example, the *L. major* and *T. brucei* genes Lmj15.1200 and Tb927.1.1530 both have an unusual apposition of two domains related to cyclic-nucleotide binding and a PAS domain. Other kinases possess, besides the PAS and cyclic nucleotide-binding domain, also domains related to the CheY-like superfamily (TvY486\_0100670) or the histidine kinase-like ATPase, C-terminal domain (Tc\_MARK\_5804) (Table S1). In *Trypanosoma grayi* (Tgr.1039.1000), a putative PAS-containing protein kinase has also a cAMP-dependent protein kinase regulatory subunit and a dimerisation-anchoring domain. Such regulatory proteins with five modules in their structure deserve special interest.

The kinases with PAS/cyclic nucleotide-binding domains are the most frequent ones and seem to be a common feature of these parasites. In eukaryotic pathogens many biological functions also are mediated by cyclic nucleotides [112]. In *Plasmodium* and kinetoplastid protists, the binding of cyclic nucleotides to these modules is apparently involved in regulating progression through the life cycle, pathogenesis and the process of cell invasion [113-115]. In *T. cruzi* and *Trypanosoma vivax* protein kinases, a PAS domain is linked to a domain of the CheY-like superfamily, a response regulator domain which is also present in some plants (*e.g. Arabidopsis*) and prokaryotic two-component systems. In prokaryotes, this domain is associated with many processes such as responding to stimuli, sporulation, regulation of transcription, ethylene detection and signal transduction [116]. Other domains such as the histidine kinase-like ATPase, C-terminal domain, are present in proteins such as



DNA gyrase B, topoisomerases and proteins involved in DNA repair. Thus far, there is little knowledge about the possible physiological role of these domains when present together with PAS in these multidomain protein kinases.

The abundance of alternative domain combinations suggests that fusions between the PAS domain and various output domains is a mechanism that allows these parasites to regulate transcription, enzyme activity, and/or protein-protein interactions more efficiently in response to environmental challenges. PAS domains can perform the same function, but in different protein contexts (*i.e.* with different partner domains), resulting in the domain having acquired a variety of novel overall functions in the physiology of the organism [117]. It is also possible that the presence of a PAS domain in various proteins (possibly involved in different signalling pathways) could serve as a “link-module” of different regulatory processes, serving as a “master regulator” in the biology of these parasites. Synteny of the different genes encoding PAS-domain containing proteins is conserved among the different trypanosomatids, as shown in Table S2 (in the Supplementary Information online).

## **6. PAS domains in *Trypanosoma cruzi***

Several genomic and proteomic analyses of *T. cruzi* have been performed [87,108,118]. The genomic analyses revealed that *T. cruzi* contains a predicted 22,570 protein-encoding genes, of which 12,570 represent allelic pairs. Among them are genes encoding different families of kinases [87,108]. Some of these protein kinases appear to possess different regulatory modules within their structure, including PAS domains (Tables 1 and S1).

Among the *T. cruzi* kinases we previously analysed are isoenzymes of the glycolytic/gluconeogenic enzyme phosphoglycerate kinase (PGK) [119-121]. The tandemly arranged genes Tc00.1047053505999.100 and Tc00.1047053505999.90 encode isoenzymes PGKA (56 kDa) and PGKB (47 kDa) which were located in the glycosomes and cytosol, respectively. In addition, by western blots a 47 kDa form was also detected in glycosomes, probably as a result of dual subcellular distribution of the Tc00.1047053505999.90 translation product. 80% of the PGK activity was associated with the cytosolic cell fraction, 20% with the glycosomes. Of this latter fraction, 23% could be attributed to

PGKA. Surprisingly, the majority (77%) of the glycosomal PGK activity (assumed to be associated with the 47 kDa form called PGKC) displayed kinetics with respect to the substrate ATP very different from that of the cytosolic PGKB: a much lower  $K_m$  (10 *versus* 99  $\mu\text{M}$  and inhibition at ATP concentrations  $>100 \mu\text{M}$ ). This was tentatively attributed to activity regulation of PGKC by posttranslational modification [120].

Additionally, we recently identified a gene for a PGK-like protein in the *T. cruzi* genome, TcCLB.506945.20, and in the genomes of other trypanosomatids, except *T. brucei*. It is present in a locus distinct from that harbouring the *PGKA-PGKB* gene tandem. The encoded TcPGK-like protein is only 44-45% identical to TcPGKA and TcPGKB, but seems to have conserved all residues for PGK substrates binding and catalysis. Interestingly, the sequence of its *pgk-like* gene contains a region encoding a PAS domain at the protein's N-terminal end. This indicates that this PGK isoenzyme of *T. cruzi* is a PAS-PGK-like enzyme. From the amino-acid sequence is predicted that this protein has a molecular weight of about 58 kDa (Table 1). Transcriptional analysis showed that the gene is expressed at different differentiation stages of the parasite [85]. Furthermore, the encoded PAS-PGK-like proteins of the different trypanosomatids possess a C-terminal tripeptide conform with a putatively functional type-1 peroxisomal-targeting signal (PTS1); in the case of the *T. cruzi* protein it is the sequence -PRL. Indeed, the protein was detected in a proteomic analysis of glycosomes purified from epimastigote forms of the parasite (PM, JLC, AC, WQ, unpublished results). It was present in the pellet fraction obtained after an osmotic shock treatment of the organelles, whereas it was found in the soluble fraction after their carbonate treatment with carbonate (PM, JLC, AC, WQ, unpublished results). This strongly suggests that the protein is peripherally associated to the glycosomal membrane. In addition, the protein appeared more abundant in the proteome of trypanosomes sampled in the stationary growth phase than in exponentially growing cells (PM, JLC, AC, WQ, unpublished results). This could be related with a role of this protein associated with the availability of the glucose in the medium which is lowered in the former growth phase [119,122].

In order to determine the possible biological relevance of the PAS domain in the PAS-PGK-like *T. cruzi* protein, we conducted a structure-based sequence alignment (using alignment programs MUSTANG / STACCATO [123,124]) of five PAS domains with

known crystal structures and diverse ligand-binding specificity (they bind FMN, FAD, heme, chromophore and fatty acid) (Fig. 2). In addition, the PAS-PGK sequences from *T. cruzi* and four other trypanosomatids were aligned to the structures using the HHPRED alignment software vs 3ue6. This result shows that the PAS domain in the PAS-PGK-like *T. cruzi* protein is an authentic PAS domain, as do the PAS domains of some *Leishmania* species. As is evident from Figure 2, PAS domains (selected from the five crystal structures) are highly diverse in some regions, yet they exhibit conservation in other areas. The PAS sequences from the PGK-like trypanosomatid proteins (collectively called PAS-Tryp) match these variable and conserved areas. An important aspect of the PAS domains of these parasites is that they not only resemble the PAS domains of proteins from other organisms, but that the sequences within the PAS-Tryp set are well conserved.

One notable difference between the PAS domains is the absence of conservation of some key amino acids for ligand binding among the reference structures (magenta positions in the alignment of Fig. 2), in agreement with their different ligand specificities. The set of trypanosomatid sequences have also residues different from the reference structures at these positions. Assuming that the PAS domains present in these PGKs of the parasites possess a ligand-binding activity implies that these proteins have probably acquired a new specificity. Furthermore, a similar high level of conservation is observed between the PAS domains and the PGK domains (87% and 85%, respectively) of these different PAS-Tryp sequences. This also strengthens the hypothesis that the PAS domain in these PAS-PGK-like protein has some function, *i.e.* ligand binding specificity, instead of being a vestige or useless relic. PAS domains can also function in signalling without small molecule binding, for example in the dimerisation of animal circadian clock proteins and the intramolecular regulation of potassium voltage-gated channels [77,125,126].

The sequences of PAS domains (approximately 75 amino acids long) in different protein kinases and PGKs of kinetoplastids were aligned and then used to perform a phylogenetic analysis. For this analysis, different sequences of parasitic kinetoplastids belonging to the genera *Trypanosoma*, *Leishmania*, *Crithidia*, *Angomonas*, *Strigomonas*, *Leptomonas* and *Endotrypanum* were taken. Sequences of PAS domains of the free-living of kinetoplastid *B. saltans* were also included (Fig. 3).

Inspection of the phylogenetic tree showed that the PAS domains present in proteins of these kinetoplastids can be grouped into two families/lineages, supported by high bootstrap values, one of them comprising the PAS domains present in protein kinases with a transmembrane domain (TMD), MAP kinases (MAPK) and isoforms of the glycolytic enzyme PGK. The other lineage represents the PAS modules of protein kinases containing also the effector domain of CAP family transcription factors (PK cap-EDs). These latter proteins are present in most organisms belonging to the genera *Trypanosoma*, *Leishmania*, *Leptomonas* and *Crithidia*. These PAS domains of PK cap-EDs were excluded from the clade comprising the PAS-containing protein kinases with a TMD, MAPKs and PGKs. This strongly suggests that these proteins acquired these domains separately from the PK cap-EDs.

The tree topology found for the taxa belonging to the Trypanosomatida and Eubodonida orders corresponds to what has been reported previously [127]. The enzyme-specific clustering revealed by this analysis indicates that the PAS domain sequences have evolved characteristics specific for each type of protein in which it is found (PAS kinase TMD, MAPK, PGK and PK cap-EDs, respectively) in the kinetoplastids. Furthermore, in the case of the PAS-PGKs the analysis suggests that this domain was lost from the mammal-infective African trypanosomes, since none of its PGK isoenzymes (A, B and C) possesses a PAS domain, while *Trypanosoma grayi* (a tsetse fly transmitted trypanosome infecting African crocodiles) preserves the PAS-PGK. This latter finding tallies with the result of a phylogenomic analysis [128], showing that *T. grayi* is more closely related to *T. cruzi* than to the African trypanosomes *T. brucei*, *T. congolense* and *T. vivax*. The presence of this PAS-PGK in the free-living organism *B. saltans* indicates that this enzyme has an origin at least prior to the divergence between the bodonids and trypanosomatids. This situation with PAS-PGK is similar as has been found for another glycolytic enzyme, glucokinase (GCK), that is present together with hexokinase in glycosomes of *B. saltans* and all trypanosomatids studied, but was lost from *T. brucei* [129,130].

The PAS domain of the trypanosomatid PAS-PGK-like protein is located at the N-terminal end of the sequence. An N-terminal position with respect to the effector domain has also been reported for the PAS domain in most other proteins [11]. Figure 4 shows that the N- and C-termini of the PGK catalytic unit are in the protein localised at the opposite side to

the catalytic site. There is a very short linker between PAS and PGK domains suggesting that the PAS domain will also be on this opposite side (*i.e.* at the bottom of the structure in the Figure 4 orientation). This hypothetical localisation positions the PAS domain near the flexible hinge region (pink) connecting the two domains of the PGK between which is the deep cleft with the active site. By interfering with the bending of the hinge region, the PAS domain could indirectly cause an allosteric transition of the PGK and hence influence its catalytic activity. It would be a scenario similar to that of the PAS-kinase in yeast (Psk1/Psk2), where ligand-dependent structural alterations occur that are transmitted to the rest of the protein. This structural alteration could induce a change in affinity of this protein for its substrate [20]. In the case of the PAS-PGK-like protein from *T. cruzi*, a change in affinity for the substrate could have a significant effect on the fluxes through the metabolic pathways. Knowing the versatility of the functions of PAS domains, is possible that the role of this domain in the PAS-PGK-like protein is not limited to an allosteric regulation of the catalytic activity. Promoting interaction of this PGK isoenzyme with other proteins is another possibility. Different scenarios can be considered for the biological role of this domain in a protein most likely localized within glycosomes, as we infer from the presence of a PTS1 in its sequence and our (unpublished) proteomics data.

## **7. Possible functions that could be attributed to the PAS domain in the *T. cruzi* PAS-PGK-like protein**

### **7.1. Metabolic adaptation to different environments**

#### **7.1.1. Internal sensor of carbon sources**

Various kinetoplastid protists have complicated life cycles involving successive cell differentiation steps. The differentiation of these different organisms has many aspects in common. The differentiation processes are inscribed in the organisms' nuclear genome, by encoding the proteins involved in the consecutive steps executing them, but each transition is activated in response to environmental cues [131,132]. The signals that activate these morphological and physiological changes act via a variety of complex physiological

signalling pathways comprising cascades of protein kinases [92,93,133]. Various studies have reported that lipid components (such as oleic acid and phorbol esters) present in the intestine of *Triatoma infestans*, an insect vector transmitting *T. cruzi*, induce cell differentiation of the parasite [134]. These free fatty acids (FFAs) act as triggers for metacyclogenesis (*i.e.* differentiation into human infective forms), through different signalling pathways, involving diacylglycerol biosynthesis and protein kinase C (PKC) [134]. It has been demonstrated that these FFAs are efficiently incorporated into *T. cruzi* epimastigotes [135]. This uptake could occur through a spontaneous flip-flop process or be mediated by specific fatty-acid transporters, as they have been reported in other cell types [134,136]. The glycosomal membrane of *T. brucei* has been shown to possess three distinct half-size ABC transporters, called GAT1-3 [137]. One of these transporters, GAT1, is involved in the uptake of long-chain fatty acids such as oleoyl-CoA, from the cytosol toward the glycosomal matrix [138]. Although sequence analysis (Figure 2) failed to identify particular similarities to fatty acid- or oxygen-sensing PAS domains, these remain candidate ligands of trypanosome PAS domains, since alternative recognition modes may have evolved. A further plausible novel candidate, in our opinion, is glucose. These ligands could modulate the activity of the intraglycosomal PAS-PGK-like enzyme through its PAS domain, as has been reported for several PAS proteins in other eukaryotes [44].

The presence of a PAS domain may be linked to an adaptive mechanism that allows these parasites to survive and deal with the environmental differences (*e.g.* a different availability of nutrients such as glucose, amino acids or fatty acids) encountered during the transfer from vertebrate to invertebrate hosts and *vice versa* [139]. Molecules such as fatty acids or glucose could make this protein function as a “switch” between the different metabolic pathways present in the glycosomes, such as those for  $\beta$ -oxidation of fatty acids, gluconeogenesis, and glycolysis. Binding of these molecules could mediate activation or inhibition of this enzyme through the PAS domain. Binding of fatty acids to this PAS-PGK-like protein could be responsible for inducing inhibition or modulation of the activity of this enzyme, similarly as the inhibition by acyl-CoAs reported for glucose-6-phosphate dehydrogenase and hexokinase in partially purified extracts of *T. cruzi* [140]. Activation of  $\beta$ -oxidation will allow epimastigotes to adapt and grow in environments poor in glucose, as the gut of insects [138], by obtaining energy through an alternative route to glycolysis.

Interestingly, molecules such as fatty acids could direct, in a coordinated manner, both the morphological and physiological changes necessary for the parasite to adapt to the environment of the insect gut. Something similar could happen in intracellular amastigotes. These live in the cytosol of mammalian cells, where the concentration of free glucose is very low, thus rendering it likely that they obtain their energy, at least to a considerable extent, through the oxidation of fatty acids [139]. In this respect, it is perhaps relevant that in the mammalian host, *T. cruzi* replicates and persists in cells such as cardiac muscle, smooth muscle and fat cells [141], where the availability of fatty acids is high and their oxidative metabolism is favoured over glucose utilization. There is probably a link between fatty-acid metabolism of the parasites with the metabolism of its host cell, through the utilization of fatty acids generated by the host cell (by peroxisomal and/or mitochondrial pathways) [142]. Maybe the linkage of the metabolism of the parasites with the metabolism of its host cell could occur in a PAS-dependent manner.

The progression from slender trypanomastigote to epimastigote could be driven by a PAS-dependent process. Some authors [143] have suggested that this transformation represents a progressive change from an environment rich in glucose (the bloodstream of the vertebrate host) to an environment poor in monosaccharides (invertebrate host). In bloodstream trypanomastigotes, the association of a glucose molecule to a PAS domain could induce activation of the PAS-kinase (as in yeast and mammals) [20], while the absence of monosaccharides could trigger an inhibition of the function of this enzyme, and the subsequent induction of the morphological and physiological changes that allow it to adapt to the shortage of glucose.

### **7.1.2. Internal sensor of oxygen**

The ability to sense and adapt to changes in oxygen tension can be very important for many organisms, both eukaryotes, from mammals to protists, and prokaryotes. Facultative anaerobic bacteria and lower eukaryotes (often pathogens), can drastically change their metabolism in response to the available electron acceptor [144,145]. For both mammals and some bacteria it has been shown that a way to detect changes in the concentration of oxygen is through PAS domains [144,146,147]. In trypanosomes, something similar might

occur. During their life cycle they are exposed to variable molecular oxygen tensions. This could suggest the need for an oxygen-sensing mechanism that can contribute to the regulation of their metabolism. It is possible that the presence of the glycosomal PAS-PGK-like protein could be part of a mechanism for rapid adaptation to the environment encountered by *T. cruzi*. One could imagine that this metabolic switching occurs through binding of oxygen to the PAS domain, as has been reported for other PAS-domain proteins in prokaryotes [33]. The binding of molecules (such as glucose, fatty acid, but also oxygen, or any other molecule) could induce a change in the activity of this enzyme, leading to a change in catalytic state, from an “on” to an “off” state. Another possibility is that the binding of specific ligands could facilitate the catalytic rate in one or the other direction of a reversible reaction (*e.g.* glycolytic *versus* gluconeogenic) by changing the kinetic properties.

## 7.2. Protein phosphorylation inside glycosomes

Another function that may be associated with the PAS domain of the PAS-PGK-like protein is phosphorylation. The protein might phosphorylate other proteins inside the glycosomes. In yeast, PAS-kinases are involved in regulating glucose partitioning and translation of signalling information about glucose concentrations into a physiological response through the phosphorylation of several proteins. Phosphorylation of these proteins by these PAS-kinases may have different consequences, such as modification of their cellular localisation, enzymatic activity or specificity for the target protein to which they interact [48]. Until now very little is known about the presence of protein kinases inside glycosomes. However, Gonzalez-Marcano *et al.* reported post-translational modification of *T. cruzi* PPDK [97], an enzyme in an auxiliary branch of glycolysis, located in the glycosomes of trypanosomatids. Phosphorylation of this protein leads to proteolytic cleavage and concomitant inactivation of the enzyme. In some eukaryotic organisms such as *Zea mays*, PPDK is reversibly phosphorylated by a bifunctional serine/threonine kinase called PPDK regulatory protein (PDRP) through a light-dependent reaction [148]. Since PDRP is apparently absent in all protists including trypanosomes, the protein and the mechanism by which this post-translational modification occurs in PPDK are unknown.



The PAS-PGK-like protein could be a good candidate for this function (see below). Similar to the PAS-kinases in yeast, this protein may act in response to a stimulus (such as a carbon source) to phosphorylate PDK and so to regulate its enzymatic activity. This notion is in agreement with the results found by González-Marcano *et al.* [97]. They observed changes in the expression of the different PDK forms (phosphorylated/non-phosphorylated) when the metabolism changes from glucose-based to amino acid-based energy and carbon source. It suggests that it functions as a nutrient-dependent regulatory system of metabolism in this parasite. Such a role by PDK in a nutrient-dependent regulatory system was supported by observing that PDK (phosphorylated/non-phosphorylated) levels did not change when a nutrient such as glucose was replenished by adding it regularly to the medium [97]. Regulation by phosphorylation of PDK through a PAS-kinase could be an elegant mechanism for metabolic regulation.

Despite the inability to detect any protein kinases so far inside glycosomes, the finding of a protein phosphatase in the organelles of *T. brucei* renders the necessity for the presence of such kinases very likely. These trypanosomes possess the phosphatase PIP39 with a PTS1 that, during the differentiation from non-proliferating short-stumpy bloodstream forms into insect-stage procyclic forms, translocates from the cytosol into the glycosomes [149]. PIP39 is clearly part of a differentiation-linked signalling pathway involving a cascade of protein kinases; it is itself phosphorylated during the process. Although the intraglycosomal targets of PIP39, as well as the nature of the kinases phosphorylating such targets remain to be determined, it underscores the importance of protein-phosphorylation dependent signalling pathways in trypanosomes. A gene putatively encoding glycosomal PAS-PGK-like protein has been detected in all trypanosomatids except *T. brucei*. But it could be imagined that in the other parasites, the PAS-PGK plays a role similar to the still unidentified kinase of *T. brucei*.

Finally, the likelihood of the presence of protein kinases within glycosomes is also supported by the finding that several glycosomal enzymes appeared to be differentially phosphorylated in a comparative phosphoproteomics analysis of procyclic and bloodstream form life-cycle stages of *T. brucei* [96]. Examples are the glycosomal NADH-dependent fumarate reductase (Tb927.5.930) and glycerol kinase (Tb09.211.3550), with three and two apparent physiological phosphorylation sites, respectively [96].

### 7.3. Regulation of subcellular localization of proteins

The presence of a PAS domain in the PAS-PGK-like protein may be related to trafficking of this protein between cytosol and glycosomes, and so to exert functions in different subcellular compartments. Studies with other organisms have shown modulation of the activity of proteins by changes of their subcellular localization directed by PAS domains. For example, in *Drosophila*, the methoprene-tolerant protein (MET) functions both as a transcription factor and juvenile hormone (JH) receptor, regulating the expression of genes involved in the development and diverse physiological processes. This transcription factor is characterized by having two topogenic sequences, a nuclear export sequence (NES) and a nuclear localization sequence (NLS), located in its PAS-A and PAS-B domains, respectively [150]. This feature allows this MET protein to undergo translocation from the cytosol to the nucleus, where it fulfils its function as a recruiter of co-activators of transcription. In the absence of JH, the MET protein is inactive in the cytosol. When the cellular concentration of JH increases, it binds to MET through its PAS-B domain. This binding induces a conformational change allowing the NLS (located within the PAS-B) in MET to be exposed, resulting in nuclear routing [151]. In plants, this type of PAS domain-mediated regulation has also been observed [152].

### 7.4. Does PAS-PGK-like protein act as a protein kinase?

The trypanosomatid's PAS-PGK protein might even fulfil multiple roles. Recent studies with mammalian cells have shown that some glycolytic kinases, like pyruvate kinase and PGK, not only phosphorylate their metabolite substrates, but under some conditions also proteins [153-155]. For example, Li and coworkers described for cells involved in tumorigenesis that, under certain conditions, translocation of isoenzyme 1 of PGK (PGK1) to mitochondria was induced. Within the mitochondria, this PGK1 functions as a protein kinase, phosphorylating pyruvate dehydrogenase kinase (PDHK1). The subsequent phosphorylation of the pyruvate dehydrogenase (PDH) complex by this activated PDHK1

suppresses mitochondrial pyruvate metabolism and promotes the Warburg effect [154]. This effect results in increased production of lactate.

Although the human PGK1 does not contain a PAS domain, the observation that it possesses protein kinase activity suggests that trypanosomatid PAS-PGK may also provide such activity inside glycosomes triggered by a cytosolic signalling system. *T. brucei* and *T. cruzi* contain homologues of the MAP kinase ERK and the peptidyl-prolyl cis/trans isomerase (PPIase) PIN1 which in the human tumour cells are involved in the signalling resulting in translocation of PGK1 to mitochondria and exhibiting protein kinase activity [156-159]. The *T. brucei* and *T. cruzi* ERK homologues are called Erk-like (TbECK1) and Erk2 (TcMAPK2), respectively [156,157]. These proteins are associated with functions related to phenotype, karyotype, cell cycle and growth modulation. The homologue of the PIN1 protein in *T. brucei* is known as TbPIN1, whereas in *T. cruzi* three homologues have been found called TcPIN1 [158], TcPAR14, and TcPAR45 [160]. These trypanosome PIN1 proteins appear to be involved in cell proliferation and growth. Like PAS-PGK, these proteins are expressed in all developmental stages of these parasites, but are localized in the cytosol. The presence of these signalling pathway homologues could support the idea that they interact with PAS-PGK to endow it with protein kinase activity. Moreover, the identification of LXL-like motifs in the PAS domain of PAS-PGK could support this notion. LXL motifs are clusters of basic amino acids present in the substrates of ERK kinase. These motifs allow the ERK substrate to interact with this kinase, through a slot or pocket coupling [154].

Interestingly, human PGK1 with protein kinase activity plays also another role in cancer development. Tumorigenesis inducing conditions result in its acetylation, rendering it into a kinase that phosphorylates Beclin1 [155]. This stimulates autophagy that is required to support tumour metabolism. Autophagy plays also a major role in trypanosomatid biology. The process is strongly upregulated during differentiation steps when the parasites undergo drastic metabolic and morphological changes to degrade redundant proteins and organelles such as glycosomes [161-166]. The possibility that PAS-PGK might act not only as a glycolytic enzyme but also as a protein kinase perhaps with an involvement in signalling pathways for differentiation steps triggered by environmental cues, like pH dependent

activation of *T. cruzi* metacyclogenesis [167] is an intriguing possibility that deserves further study.

## 8. Conclusions and future trends

PAS domains are highly conserved regulatory modules, widely distributed in Archaea, Bacteria and eukaryotes. They are present in different types of proteins and participate in a wide range of cellular functions, such as virulence, aerotaxis, metabolism and adaptation of microorganisms to their environment. Little is known about the presence of these regulatory domains in parasitic organisms, specifically in kinetoplastids. Genome analysis of some of these parasites revealed the presence of PAS domains in various protein kinases, where they are associated with other regulatory domains. Since the PAS domains have been overlooked so far in the kinetoplastids, because of their low abundance in the parasite genomes and their high sequence divergence, we cannot discard the possibility that this has also been the case in other parasites, notably unicellular ones. Therefore, our assessment of these overlooked domains in kinetoplastid genomes provides a paradigm for assessing possibilities for PAS domains in other parasites. In *T. cruzi* a PAS-PGK-like protein was identified. Analysis of its amino-acid sequence suggests that this PAS domain can have an important function in the modulation of the catalytic activity of the glycolytic enzyme. However, due to the multiple functions that these domains can have, several other possible functions of this domain in this glycosomal enzyme could also be considered. Currently, it is difficult to establish the role of this PAS domain in the PGK-like protein of *T. cruzi*. Therefore, we set out to clone and overexpress this protein. Future molecular and biochemical studies of this recombinant protein could give insight into the physiological significance of this likely regulatory PAS module in a glycolytic enzyme of an important human parasite.

## 9. Acknowledgement

This work was financially supported by the 'Fondo Nacional de Ciencia, Tecnología e Innovación' (FONACIT) in Project MC-2007001425 (to J.L. Concepción).

## 10. References

- [1] S.J. Wheelan, A. Marchler-Bauer, S.H. Bryant, Domain size distributions can predict domain boundaries, *Bioinformatics* 16 (2000) 613-618, <http://dx.doi.org/10.1093/bioinformatics/16.7.613>.
- [2] L. Brocchieri, S. Karlin, Protein length in eukaryotic and prokaryotic proteomes, *Nucleic Acids Res.* 33 (2005) 3390-3400, <http://dx.doi.org/10.1093/nar/gki615>.
- [3] A. Andreeva, D. Howorth, S.E. Brenner, T.J. Hubbard, C. Chothia, C.C. Murzin, SCOP database in 2004: refinements integrate structure and sequence family data, *Nucleic Acids Res.* 32 (2004) 226-229, <http://dx.doi.org/10.1093/nar/gkh039>.
- [4] M.K. Basu, L. Carmel, I.B. Rogozin, E.V. Koonin, Evolution of protein domain promiscuity in eukaryotes, *Genome Res.* 18 (2008) 449-461, <http://dx.doi.org/10.1101/gr.6943508>.
- [5] E. Pang, T. Tan, K. Lin, Promiscuous domains: facilitating stability of the yeast protein-protein interaction network, *Mol. BioSyst.* 8 (2012) 766-771, <http://dx.doi.org/10.1039/clmb05364g>.
- [6] R.D. Finn, J. Mistry, B. Schuster-Böckler, S. Griffiths-Jones, V. Hollich, T. Lasmann et al., Pfam: clans, web tools and services, *Nucleic Acids Res.* 34 (2006) D247-251, <http://dx.doi.org/10.1093/nar/gkj149>.
- [7] J.R. Nambu, J.O. Lewis, K.A. Wharton, S.T. Crews, The *Drosophila* single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development, *Cell* 67 (1991) 1157-1167, <http://dx.doi.org/10.1093/nar/gkj149>.
- [8] Z. Salamon, T.E. Meyer, G. Tollin, Photobleaching of the photoactive yellow protein from *Ectothiodospira halophila* promotes binding to lipid bilayers: evidence from surface plasmon resonance spectroscopy, *Biophys. J.* 68 (1995) 648-654, [http://dx.doi.org/10.1016/S0006-3495\(95\)80225-0](http://dx.doi.org/10.1016/S0006-3495(95)80225-0).

- [9] Y. Imamoto, M. Kataoka, Structure and photoreaction of photoactive yellow protein, a structural prototype of the PAS domain superfamily, *Photochem. Photobiol.* 83 (2007) 40-49, <http://dx.doi.org/10.1562/2006-02-28-IR-827>.
- [10] J.L. Pellequer, K.A. Wager-Smith, S.A. Kay, E.D. Getzoff, Photoactive yellow protein: a structural prototype for the three-dimensional fold of the PAS domain superfamily, *Proc. Natl. Acad. Sci. U.S.A.* 95 (1998) 5884-5890.
- [11] A. Möglich, R.A. Ayers, K. Moffat, Structure and signaling mechanism of Per-ARNT-Sim domains, *Structure* 17 (2009) 1282-1294, <http://dx.doi.org/10.1016/j.str.2009.09.011>.
- [12] Q. Ma, F. Roy, S. Herrmann, B.L. Taylor, M.S. Johnson, The Aer protein of *Escherichia coli* forms a homodimer independent of the signaling domain and flavin adenine dinucleotide binding, *J. Bacteriol.* 186 (2004) 7456-7459, <http://dx.doi.org/10.1128/JB.186.21.7456-7459.2004>.
- [13] Z. Xie, L.E. Ulrich, I.B. Zhulin, G. Alexandre, PAS domain containing chemoreceptor couples dynamic changes in metabolism with chemotaxis, *Proc Natl. Acad. Sci. U.S.A.* 107 (2010) 2535-2240, <http://dx.doi.org/10.1073/pnas.0910055107>.
- [14] M.A. Gilles-Gonzalez, G. Gonzalez, Signal transduction by heme-containing PAS-domain proteins, *J. Appl. Physiol.* 96 (2004) 774-783, <http://dx.doi.org/10.1152/jappphysiol.00941.2003>.
- [15] D. Georgellis, O. Kwon, E.C.C. Lin, S.M. Wong, B.J. Akerley, Redox signal transduction by the ArcB sensor kinase of *Haemophilus influenzae* lacking the PAS domain, *J. Bacteriol.* 183 (2001) 7206-7212, <http://dx.doi.org/10.1128/JB.183.24.7206-7212.2001>.
- [16] M. Etzkorn, H. Kneuper, P. Dünwald, V. Vijayan, J. Krämer, C. Griesinger, et al., Plasticity of the PAS domain and a potential role for signal transduction in the histidine kinase DcuS, *Nat. Struct. Mol. Biol.* 15 (2008) 1031-1039, <http://dx.doi.org/10.1038/nsmb.1493>.

- [17] B.L. Taylor, M.S. Johnson, K.J. Watts, Signal transduction in prokaryotic PAS domains, in: S.T. Crews (Eds.), PAS proteins: regulators and sensors of development and physiology, Springer., U.S.A., 2004, pp. 17-50.
- [18] R. Little, P. Salinas, P. Slayny, T.A. Clarke, R. Dixon, Substitutions in the redox-sensing PAS domain of the NifL regulatory protein define an inter-subunit pathway for redox-signal transmission, *Mol. Microbiol.* 82 (2011) 222-235, <http://dx.doi.org/10.1111/j.1365-2958.2011.07812.x>.
- [19] X. Ma, N. Sayed, P. Baskaran, A. Beuve, F. Akker, PAS-mediated dimerization of soluble guanylyl cyclase revealed by signal transduction histidine kinase domain crystal structure, *J. Biol. Chem.* 283 (2008) 1167-1178, <http://dx.doi.org/10.1074/jbc.M706218200>.
- [20] D. DeMille, J.H. Grose, PAS kinase: a nutrient sensing regulator of glucose homeostasis, *IUBMB Life* 65 (2013) 921-929, <http://dx.doi.org/10.1002/iub.1219>.
- [21] R.J. Kewley, M.L. Whitelaw, A. Chapman-Smith, The mammalian basic helix-loop-helix/PAS family of transcriptional regulators, *Int. J. Biochem. Cell Biol.* 36 (2004) 189-204, [http://dx.doi.org/10.1016/S1357-2725\(03\)00211-5](http://dx.doi.org/10.1016/S1357-2725(03)00211-5).
- [22] J. Herrou, S. Crosson, Function, structure and mechanism of bacterial photosensory LOV proteins, *Nat. Rev. Microbiol.* 9 (2011) 713-723, <http://dx.doi.org/10.1038/nrmicro2622>.
- [23] J.H. Grose, J. Rutter, The role of PAS kinase in PASSing the glucose signal, *Sensors* 10 (2010) 5668-5682, <http://dx.doi.org/10.3390/s100605668>.
- [24] M.C. Lindebro, L. Poellinger, M.L. Whitelaw, Protein-protein interaction via PAS domains: role of the PAS domain in positive and negative regulation of the bHLH/PAS dioxin receptor-Arnt transcription factor complex, *EMBO J.* 14 (1995) 3528-3539.
- [25] J. Lee, D.R. Tomchick, C.A. Brautigam, M. Machius, R. Kort, K.J. Hellingwerf, et al., Changes at the KinA PAS-A dimerization interface influence histidine kinase function, *Biochemistry* 47 (2008) 4051-4064, <http://dx.doi.org/10.1021/bi7021156>.

- [26] U. Heintz, A. Meinhart, A. Winkler, Multi-PAS domain-mediated protein oligomerization of PpsR from *Rhodobacter sphaeroides*, *Acta Crystallogr. D. Biol. Crystallogr.*, 70 (2014) 863-876, <http://dx.doi.org/10.1107/S1399004713033634>.
- [27] K. Hashimoto, A.R. Panchenko, Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proc Natl. Acad. Sci. U.S.A.* 107 (2010) 20352-20357, <http://dx.doi.org/10.1073/pnas.1012999107>.
- [28] B.L. Taylor, I.B. Zhulin, PAS domains: internal sensors of oxygen, redox potential, and light, *Microbiol. Mol. Biol. Rev.* 63 (1999) 479-506.
- [29] Q. Mei, V Dyornyk, Evolution of PAS domains and PAS-containing genes in eukaryotes, *Chromosoma* 123 (2014) 385-405, <http://dx.doi.org/10.1007/s00412-014-0457-x>.
- [30] S. Herrmann, Q. Ma, M.S. Johnson, A.V. Repik, B.L. Taylor, PAS domain of the Aer redox sensor requires C-terminal residues for native fold formation and flavin adenine dinucleotide binding, *J. Bacteriol.* 186 (2004) 6782-6791, <http://dx.doi.org/10.1128/JB.186.20.6782-6791.2004>.
- [31] R.A. Schmitz, K. Klopprogge, R. Grabbe, Regulation of nitrogen fixation in *Klebsiella pneumoniae* and *Azotobacter vinelandii*: NifL, transducing two environmental signals to the nif transcriptional activator NifA, *J. Mol. Microbiol. Biotechnol.* 4 (2002) 235-242.
- [32] K Krauss, B.Q. Minh, A. Losi, W Gärtner, T Eggert, A. von Haeseler, KE Jaeger, Distribution and phylogeny of light-oxygen-voltage-blue-light-signaling proteins in the three kingdoms of life, *J. Bacteriol.* 191 (2009) 7234-7242, <http://dx.doi.org/10.1128/JB.00923-09>.
- [33] F. Rao, Q. Ji, I. Soehano, Z.X. Liang, Unusual heme-binding PAS domain from YybT family proteins, *J. Bacteriol.* 193 (2011) 1543-1551, <http://dx.doi.org/10.1128/JB.01364-01>.
- [34] Y.Y. Londer, I.S. Dementieva, C.A. D'Ausilio, P.R. Pokkuluri, M. Schiffer, Characterization of a c-type heme-containing PAS sensor domain from *Geobacter*



*sulfurreducens* representing a novel family of periplasmic sensors in Geobacteraceae and other bacteria, FEMS Microbiol. Lett. 258 (2006) 173-181, <http://dx.doi.org/10.1111/j.1574-6968.2006.00220.x>.

[35] K.R. Rodgers, G.S. Lukat-Rodgers, Insights into heme-based O<sub>2</sub> sensing from structure-function relationship in the FixL proteins, J. Inorgan. Biochem. 99 (2005) 963-977, <http://dx.doi.org/10.1016/j.jinorgbio.2005.02.016>.

[36] Y. Sasakura, T. Yoshimura-Suzuki, H. Kurokawa, T. Shimizu, Structure-function relationships of EcDOS, a heme-regulated phosphodiesterase from *Escherichia coli*, Acc. Chem. Res. 39 (2006) 37-43, <http://dx.doi.org/10.1021/ar0501525>.

[37] R. Purohit, A. Weichsel, W.R. Montfort, Crystal structure of the alpha subunit PAS domain from soluble guanylyl cyclase, Protein Sci. 22 (2013) 1439-1444, <http://dx.doi.org/10.1002/pro.2331>.

[38] H.M. Girvan, A.W. Munro, Heme sensor protein, J. Biol. Chem. 288 (2013) 13194-13203, <http://dx.doi.org/10.1074/jbc.R112.422642>.

[39] V. García, J.A. Reyes-Darias, D. Martín-Mora, B. Morel, M.A. Matilla, T. Krell, Identification of a chemoreceptor for C<sub>2</sub> and C<sub>3</sub> carboxylic acids, Appl. Environ. Microbiol. 81 (2015) 5449-5457, <http://dx.doi.org/10.1128/AEM.01529-15>.

[40] J. Krämer, J.D. Fischer, E. Zientz, V. Vijayan, C. Griesinger, A. Lupas, et al., Citrate sensing by the C<sub>4</sub>-dicarboxylate/citrate sensor kinase DcuS of *Escherichia coli*: binding site and conversion of DcuS to a C<sub>4</sub>-dicarboxylate- or citrate-specific sensor, J. Bacteriol. 189 (2007) 4290-4298, <http://dx.doi.org/10.1128/JB.00168-07>.

[41] C. Monzel, P. Degreif-Dünnwald, C. Gröpper, C. Griesinger, G. Unden, The cytoplasmic PAsc domain of the sensor kinase DcuS of *Escherichia coli*: role in signal transduction, dimer formation, and DctA interaction, Microbiologyopen 2 (2013) 912-926, <http://dx.doi.org/10.1002/mbo3.127>.

- [42] Y.F. Zhou, B. Nan, J. Nan, Q. Ma, S. Panjikar, Y.H. Liang, Y. Wang, X.D. Su, C4-dicarboxylates sensing mechanism revealed by the crystal structure of DctB sensor domain, *J. Mol. Biol.* 383 (2008) 49-61, <http://dx.doi.org/10.1016/j.jmb.2008.08.010>.
- [43] J. King-Scott, P.D. Konarev, S. Panjikar, R. Jordanova, D.J. Svergun, P.A. Tucker, Structural characterization of the multidomain regulatory protein Rv1364 from *Mycobacterium tuberculosis*, *Structure* 19 (2011) 59-69, <http://dx.doi.org/10.1016/j.str.2010.11.010>.
- [44] A.M. Fala, J.F. Oliveira, D. Adamoski, J.A. Aricetti, M.M. Dias, V.B. Marcio, et al., Unsaturated fatty acids as high-affinity ligands of the C-terminal Per-ARNT-Sim domain from the hypoxia-inducible factor 3 $\alpha$ , *Sci. Rep.* 5 (2015) 12698, <http://dx.doi.org/10.1038/srep12698>.
- [45] R.K. Ernst, T. Guina, S.I. Miller, *Salmonella typhimurium* outer membrane remodeling: role in resistance to host innate immunity, *Microbes Infect.* 3 (2001) 1327-1334, [http://dx.doi.org/10.1016/S1286-4579\(01\)01494-0](http://dx.doi.org/10.1016/S1286-4579(01)01494-0).
- [46] E. García-Véscovi, Y.M. Ayala, E. Di Cera, E.A. Groisman, Characterization of the bacterial sensor protein PhoQ. Evidence for distinct binding sites for Mg<sup>2+</sup> and Ca<sup>2+</sup>, *J. Biol. Chem.* 272 (1997) 1440-1443, <http://dx.doi.org/10.1074/jbc.272.3.1440>.
- [47] Z. Ma, F.E. Jacobsen, D.P. Giedroc, Coordination chemistry of bacterial metal transport and sensing, *Chem. Rev.* 109 (2009) 4644-4681, <http://dx.doi.org/10.1021/cr900077w>.
- [48] J.H. Grose, T.L. Smith, H. Sabic, H. Rutter H, Yeast PAS kinase coordinates glucose partitioning in response to metabolic and cell integrity signalling, *EMBO J.* 26 (2010) 4824-4830, <http://dx.doi.org/10.1038/sj.emboj.7601914>.
- [49] S.P. Hong, F.C. Leiper, A. Woods, D. Carling, M. Carlson, Activation of yeast Snf1 and mammalian AMP-activated protein kinase by upstream kinases, *Proc. Natl. Acad. Sci. U.S.A.* 100 (2003) 8839-8843, <http://dx.doi.org/10.1073/pnas.1533136100>.

- [50] C.M. Sutherland, S.A. Hawley, R.R. McCartney, A. Leech, M.J. Stark, M.C. Schmidt, D.G. Hardie, Elm1p is one of three upstream kinases for the *Saccharomyces cerevisiae* SNF1 complex, *Curr. Biol.* 13 (2003) 1299-1305.
- [51] K. Hedbacker, S.P. Hong, M. Carlson, Pak1 protein kinase regulates activation and nuclear localization of Snf1-Gal83 protein kinase, *Mol. Cell. Biol.* 24 (2004) 8255-8263, <http://dx.doi.org/10.1128/MCB.24.18.8255-8263.2004>.
- [52] M. Semache, B. Zarrouki, G. Fontes, S. Fogarty, C. Kikani, M.B. Chawki, J. Rutter, V. Poitout, Per-Arnt-Sim kinase regulates pancreatic duodenal homeobox-1 protein stability via phosphorylation of glycogen synthase kinase 3 beta in pancreatic beta-cells, *J. Biol. Chem.* 288 (2013) 24825-24833, <http://dx.doi.org/10.1074/jbc.M113.495945>.
- [53] D. Huang, I. Farkas, P.J. Roach, Pho85p, a cyclin-dependent protein kinase, and the Snf1p protein kinase act antagonistically to control glycogen accumulation in *Saccharomyces cerevisiae*, *Mol. Cell. Biol.* 16 (1996) 4357-4365.
- [54] K. Eckhardt, J. Troger, J. Reissmann, D.M. Katschinski, K.F. Wagner, et al., Male germ cell expression of the PAS domain kinase PASKIN and its novel target eukaryotic translation elongation factor eEF1A1, *Cell. Physiol. Biochem.* 20 (2007) 227-240, <http://dx.doi.org/10.1159/000104169>.
- [55] E.F. Blommaart, J.J. Luiken, P.J. Blommaart, G.M. van Woerkom, A.J. Meijer, Phosphorylation of ribosomal protein S6 is inhibitory for autophagy in isolated rat hepatocytes, *J. Biol. Chem.* 270 (1995) 2320-2326.
- [56] J. Rutter, B.L. Probst, S.L. McKnight, Coordinate regulation of sugar flux and translation by PAS kinase, *Cell* 111 (2002) 17-28.
- [57] D. DeMille, B.D. Badal, J.B. Evans, A.D. Mathis, J.F. Anderson, J.H. Grose, PAS kinase is activated by direct SNF1-dependent phosphorylation and mediates inhibition of TORC1 through the phosphorylation and activation of Pbp1, *Mol. Biol. Cell.* 2015 26 (2015) 569-82, <http://dx.doi.org/10.1091/mbc.E14-06-1088>.

- [58] B. Winnen, E. Anderson, J.L. Cole, G.F King, S.L. Rowland, Role of the PAS sensor domains in the *Bacillus subtilis* sporulation kinase KinA, *J. Bacteriol.* 195 (2013) 2349-2358, <http://dx.doi.org/10.1128/JB.00096-13>.
- [59] K. Stephenson, J.A. Hoch, PAS-A domain of phosphorelay sensor kinase A: a catalytic ATP-binding domain involved in the initiation of development in *Bacillus subtilis*, *Proc. Natl. Acad. Sci. U.S.A.* 98 (2001) 15251-15256, <http://dx.doi.org/10.1073/pnas.25140898>.
- [60] D. Roymans, R. Willems, D.R. Van Blockstaele, H. Slegers, Nucleoside diphosphate kinase (NDPK/NM23) and the waltz with multiple partners: possible consequences in tumor metastasis, *Clin. Exp. Metastasis* 19 (2002) 465-476, <http://dx.doi.org/10.1023/A:1020396722860>.
- [61] C.A. Pereira, L.A. Bouvier, M. de los Milagros-Cámara, M.R. Miranda, Singular features of trypanosomatids` phosphotransferases involved in cell energy management, *Enzyme Res.* 2011 (2011) 576483, <http://dx.doi.org/10.4061/2011/576483>.
- [62] L. Wang, C. Fabret, K. Kanamura, K. Stephenson, V. Dartois, M. Perego, et al., Dissection of the functional and structural domains of phosphorelay histidine kinase A of *Bacillus subtilis*, *J. Bacteriol.* 183 (2001) 2795-2802, <http://dx.doi.org/10.1128/JB.183.9.2795.2802.2001>.
- [63] E. Magnani, M.K. Barton, A Per-ARNT-Sim like sensor domain uniquely regulates the activity of the homeodomain leucine zipper transcription factor REVOLUTA in *Arabidopsis*, *Plant Cell* 23 (2011) 567-582, <http://dx.doi.org/10.1105/tpc.110.080754>.
- [64] B. Greb-Markiewicz, M. Orłowski, J. Dobrucki, A. Ozyhar, Sequences that direct subcellular traffic on the *Drosophila* methoprene-tolerant protein (MET) are located predominantly in the PAS domains, *Mol. Cell. Endocrinol.* 354 (2011) 16-26, <http://dx.doi.org/10.1016/j.mce.2011.06.035>.
- [65] R. Dixon, D. Kahn, Genetic regulation of biological nitrogen fixation, *Nat. Rev. Microbiol.* 2 (2004) 621-631, <http://dx.doi.org/10.1038/nrmicro95>.

- [66] R.K. Jaiswal, G. Manjeera, B. Gopal B, Role of a PAS sensor domain in the *Mycobacterium tuberculosis* transcription regulator Rv1364c, *Biochem. Biophys. Res. Commun.* 398 (2010) 342-349, <http://dx.doi.org/10.1016/j.bbrc.2010.06.027>.
- [67] C.P. Ponting, L. Aravind, PAS: a multifunctional domain family comes to light, *Curr. Biol.* 7 (1997) R674-R677, [http://dx.doi.org/10.1016/s0906-9822\(06\)00352-6](http://dx.doi.org/10.1016/s0906-9822(06)00352-6).
- [68] A. Mooney, P.G. Ward, K.E. O'Connor, Microbial degradation of styrene: biochemistry, molecular genetics, and perspectives for biotechnological applications, *Appl. Microbiol. Biotechnol.* 72 (2006) 1-10, <http://dx.doi.org/10.1007/s00253-006-0443-1>.
- [69] S. Sanselicio, M. Bergé, L. Théraulaz, S.K. Radhakrishnan, P.H. Viollier. Topological control of the *Caulobacter* cell cycle circuitry by a polarized single-domain PAS protein, *Nat. Commun.* 6 (2015) 7005, <http://dx.doi.org/10.1038/ncomms8005>.
- [70] E. Dupré, A. Wohlkonig, J. Herrou, C. Locht, F. Jacob-Dubuisson, R. Antoine, Characterization of the PAS domain in the sensor-kinase BvgS: mechanical role in signal transmission, *BMC Microbiol.* 13 (2013) 172, <http://dx.doi.org/10.1186/1471-2180-13-172>.
- [71] L. Rickman, J.W. Saldanha, D.M. Hunt, D.N. Hoar, M.J. Colston, J.B. Millar, R.S. Buxton, A two-component signal transduction system with a PAS domain-containing sensor is required for virulence of *Mycobacterium tuberculosis* in mice, *Biochem. Biophys. Res. Commun.* 314 (2004) 259-267, <http://dx.doi.org/10.1016/j.bbrc.2003.12.082>.
- [72] W.R. Briggs, M.A. Olney, Photoreceptors in plant photomorphogenesis to date. Five phytochromes, two cryptochromes, one phototropin, and one superchrome, *Plant Physiol.* 125 (2001) 85-88, <http://dx.doi.org/10.1104./pp:125.1.85>.
- [73] F.W. Li, M. Melkonian, C.J. Rothfels, J.C. Villarreal, D.W. Stevenson, S.W. Graham, G.K.S Wong, K.M. Pryer, S. Mathews. Phytochrome diversity in green plants and the origin of canonical plant phytochromes, *Nat. Commun.* 6 (2015) 7852, <http://dx.doi.org/10.1038/ncomms8852>.

- [74] J.A. Jarillo, J. Capel, R.H. Tang, H.Q. Yang, J.M. Alonso, J.R. Ecker, A.R. Cashmore, An *Arabidopsis* circadian clock component interacts with both CRY1 and phyB, *Nature* 410 (2001) 487-490, <http://dx.doi.org/10.1038/35068589>.
- [75] K. Mukherjee, T.R. Bürglin, MEKLA, a novel domain with similarity to PAS domains, is fused to plant homeodomain-leucine zipper III proteins, *Plant Physiol.* 140 (2006) 1142-1150, <http://dx.doi.org/10.1104/pp.105.073833>.
- [76] J.H. Vogt, J.H. Schippers, Setting the PAS, the role of circadian PAS domain proteins during environmental adaptation in plants, *Front. Plant Sci.* 6 (2015) 513, <http://dx.doi.org/10.3389/fpls.2015.00513>.
- [77] AK. Michael, C.L. Partch, bHLH-PAS proteins: functional specification through modular domain architecture, *OA Biochemistry* 1 (2013) 16-22, <http://dx.doi.org/10.13172/2052-9651-1-2-1123>.
- [78] M.E. Hahn, S.I. Karchner, M.A. Shapiro, S.A. Perera, Molecular evolution of two vertebrate aryl hydrocarbon (dioxin) receptors (AHR1 and AHR2) and the PAS family, *Proc. Natl. Acad. Sci. U.S.A.* 94 (1997) 13743-13748.
- [79] N. Wang, G. Shaulsky, R. Escalante, W.F. Loomis, A two-component histidine kinase gene that functions in *Dictyostelium* development, *EMBO J.* 15(1996) 3890-3898.
- [80] M.J. Zinda, C.K. Singleton, The hybrid histidine kinase *dhkB* regulates spore germination in *Dictyostelium discoideum*, *Dev. Biol.* 196 (1998) 171-183, <http://dx.doi.org/10.1006/dbio.1998.8854>.
- [81] F. Oehme, S.C. Schuster SC, Osmotic stress-dependent serine phosphorylation of the histidine kinase homologue DokA, *BMC Biochem.* 2 (2001) 2, <http://dx.doi.org/10.1186/1471-2091-2-2>.
- [82] O. Arnaiz, L. Sperling, ParameciumDB in 2011: new tools and new data for functional and comparative genomics of the model ciliate *Paramecium tetraurelia*, *Nucleic Acids Res.* 39 (2011) D632-D636, <http://dx.doi.org/10.1093/nar/gkq918>.

- [83] L.K. Fritz-Laylin, S.E. Prochnik, M.L. Ginger, J.B. Dacks, M.L. Carpenter, M.C. Field, et al., The genome of *Naegleria gruberi* illuminates early eukaryotic versatility, *Cell* 140 (2010) 631-642, <http://dx.doi.org/10.1016/j.cell.2010.01.032>.
- [84] M.Y. Galperin, A.N. Nikolskaya, E.V. Koonin, Novel domains of the prokaryotic two-component signal transduction systems, *FEMS Microbiol. Lett.* 203 (2001) 11-21, [http://dx.doi.org/10.1016/S0378-1097\(01\)00326-3](http://dx.doi.org/10.1016/S0378-1097(01)00326-3).
- [85] M. Aslett, C. Aurrecochea, M. Berriman, J. Brestelli, B.P. Brunk, M. Carrington et al., TriTrypDB: a functional genomic resource for the Trypanosomatidae, *Nucleic Acids Res.* 38 (2010) D457-D62, <http://dx.doi.org/10.1093/nar/gkp851>.
- [86] K. Stuart, R. Brun, S. Croft, A. Fairlamb, R.E. Gürtler, J. McKerrow, et al., Kinetoplastids: related protozoan pathogens, different diseases, *J. Clin. Invest.* 118 (2008) 1301-1310, <http://dx.doi.org/10.1172/JCI33945>.
- [87] M. Parsons, E.A. Worthey, P.N. Ward, J.C. Mottram, Comparative analysis of the kinomes of three pathogenic trypanosomatids: *Leishmania major*, *Trypanosoma brucei* and *Trypanosoma cruzi*, *BMC Genomics* 6 (2005) 127, <http://dx.doi.org/10.1186/1471-2164-6-127>.
- [88] O.S. Adeyemi, F.A. Sulaiman, Biochemical and morphological changes in *Trypanosoma brucei brucei* infected rats treated with homidium chloride and diminazene aceturate, *J. Basic. Clin. Physiol. Pharmacol.* 23 (2012) 179-183, <http://dx.doi.org/10.1515/jbcpp-2012-0018>.
- [89] B.A. Eyford, T. Sakurai, D. Smith, B. Loveless, C. Hertz-Fowler, J.E. Donelson, N. Inoue N, T.W. Pearson, Differential protein expression throughout the life cycle of *Trypanosoma congolense*, a major parasite of cattle in Africa, *Mol. Biochem. Parasitol.* 177 (2011) 116-125, <http://dx.doi.org/10.1016/j.molbiopara.2011.02.009>.
- [90] A.J. Mejía, M.T. Palaú, C.A. Zúñiga, Protein profile of *Trypanosoma cruzi* and *Trypanosoma rangeli*, *Parasitol. Latinoam.* 59 (2004) 142-147, <http://dx.doi.org/10.4067/s0717-77122004000300009>.

- [91] M.D. Urbaniak, M.L. Güther, M.A. Ferguson, Comparative SILAC proteomic analysis of *Trypanosoma brucei* bloodstream and procyclic lifecycle stages, PLoS One 7 (2012) e36619, <http://dx.doi.org/10.1371/journal.pone.0036619>.
- [92] B.M. Mony, P. MacGregor, A. Ivens, F. Rojas, A. Cowton, J. Young, D. Horn, K. Matthews, Genome-wide dissection of the quorum sensing signalling pathway in *Trypanosoma brucei*, Nature 505 (2015) 681-685, <http://dx.doi.org/10.1038/nature12864>.
- [93] B.M. Mony, K.R. Matthews, Assembling the components of the quorum sensing pathway in African trypanosomes, Mol. Microbiol. 96 (2015) 220-232, <http://dx.doi.org/10.1111/mmi.12949>.
- [94] T. Aboagye-Kwarteng, O.K. ole-MoiYoi, J.D. Lonsdale-Eccles, Phosphorylation differences among proteins of bloodstream developmental stages of *Trypanosoma brucei brucei*, Biochem. J. 275 (1991) 7-14, <http://dx.doi.org/10.1042/bj2750007>.
- [95] M. Parsons, M. Valentine, J. Deans, G.L. Schieven, J.A. Ledbetter, Distinct patterns of tyrosine phosphorylation during the life cycle of *Trypanosoma brucei*, Mol. Biochem. Parasitol. 45 (1991) 241-248, [http://dx.doi.org/10.1016/0166-6851\(91\)90091-J](http://dx.doi.org/10.1016/0166-6851(91)90091-J).
- [96] M.D. Urbaniak, D.M. Martin, M.A. Ferguson, Global quantitative SILAC phosphoproteomics reveals differential phosphorylation is widespread between the procyclic and bloodstream form lifecycle stages of *Trypanosoma brucei*, J. Proteome Res. 12 (2013) 2233-2244, <http://dx.doi.org/10.1021/pr400086y>.
- [97] E. González-Marcano, A. Mijares, W. Quiñones, A. Cáceres, J.L. Concepción, Post-translational modification of the pyruvate phosphate dikinase from *Trypanosoma cruzi*, Parasitol. Int. 63 (2014) 80-86, <http://dx.doi.org/10.1016/j.parint.2013.09.007>.
- [98] C. Doerig, J.C. Rayner, A. Scherf, A.B. Tobin, Post-translational protein modifications in malaria parasites, Nat. Rev. Microbiol. 13 (2015) 160-172, <http://dx.doi.org/10.1038/nrmicro3402>.
- [99] J. Ellis, M. Sarkar, E. Hendriks, K. Matthews, A novel ERK-like, CRK-like protein kinase that modulates growth in *Trypanosoma brucei* via an autoregulatory C-terminal



extension, *Mol. Microbiol.* 53 (2004) 1487–1499, <http://dx.doi.org/10.1111/j.1365-2958.2004.04218.x>

[100] E.B. Gómez, M.I. Santori, S. Laría, J.C. Engel, J. Swindle, H. Eisen, P. Szankasi, M.T. Téllez-Iñón, Characterization of the *Trypanosoma cruzi* Cdc2p-related protein kinase 1 and identification of three novel associating cyclins, *Mol. Biochem. Parasitol.* 113 (2001) 97-108.

[101] S. Li, M.E. Wilson, J.E. Donelson, *Leishmania chagasi*: A Gene Encoding a Protein Kinase with a Catalytic Domain Structurally Related to MAP Kinase Kinase, *Exp. Parasitol.* 82 (1996) 87-96.

[102] D. Portal, G.S. Lobo, S. Kadener, J. Prasad, J.M. Espinosa, C.A. Pereira, et al., *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 (2003) 9-21.

[103] M. Parsons, J.A. Ledbetter, G.L. Schieven, A.E. Nel, S.B. Kanner, Developmental regulation of pp44/46, tyrosine-phosphorylated proteins associated with tyrosine/serine kinase activity in *Trypanosoma brucei*, *Mol. Biochem. Parasitol.* 63 (1994) 69-78.

[104] M. Parsons, M. Valentine, V. Carter, Protein kinases in divergent eukaryotes: Identification of protein kinase activities regulated during trypanosome development, *Proc. Natl. Acad. Sci. U.S.A.* 90 (1993) 2656-2660.

[105] B. Pils, J. Schultz, Inactive enzyme-homologues find new function in regulatory processes, *J. Mol. Biol.* 340 (2004) 399-404, <http://dx.doi.org/10.1016/j.jmb.2004.04.063>.

[106] C. Adrain, M. Freeman, New lives for old: evolution of pseudoenzyme function illustrated by iRhoms, *Nat. Rev. Mol. Cell Biol.* 13 (2012) 489-498, <http://dx.doi.org/10.1038/nrm3392>.

[107] P.H. Stoco, G. Wagner, C. Talavera-Lopez, A. Gerber, A. Zaha, C.E. Thompson, et al., Genome of the avirulent human- infective trypanosome-*Trypanosome rangeli*, *PLoS Negl. Trop. Dis.* 8 (2014) e3716, <http://dx.doi.org/10.1376/journal.pntd.0003176>.

- [108] N.M El Sayed, P.J Myler, D.C. Bartholomeu, D. Nilsson, G. Aggarwal, A.N. Tran, et al., 2005, The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease, *Science* 309 (2005) 409-415, <http://dx.doi.org/10.1126/science.1112631>.
- [109] B.M. Porcel, F. Denoed, F. Opperdoes, B. Noel, M.A. Madoui, T.C. Hammarton, et al., The streamlined genome of *Phytomonas* spp. relative to human pathogenic kinetoplastids reveals a parasite tailored for plants, *PLoS Genet.* 10 (2014):e1004007, <http://dx.doi.org/10.1371/journal.pgen.1004007>.
- [110] A.P. Jackson, T.D. Otto, M. Aslett, S.D. Armstrong, F. Bringaud, A. Schlacht, C. Hartley, et al., Kinetoplastid phylogenomics reveals the evolutionary innovations associated with the origins of parasitism, *Curr. Biol.* 26 (2016) 161-172, <http://dx.doi.org/10.1016/j.cub.2015.11.055>.
- [111] G. Tanifuji, U. Cenci, D. Moog, S. Dean, T. Nakayama, et al., Genome sequencing reveals metabolic and cellular interdependence in an amoeba-kinetoplastid symbiosis, *Sci. Rep.* 7 (2017) 11688, <http://dx.doi.org/10.1038/s41598-017-11866-x>.
- [112] S. Mohanty, E.J. Kennedy, F.W. Herberg, R. Hui, S.S Taylor, G. Langsley, et al., Structural and evolutionary divergence of cyclic nucleotide binding domains in eukaryotic pathogens: implications for drug design, *Biochim. Biophys. Acta* 1854 (2015) 1575-1585, <http://dx.doi.org/10.1016/j.bbapap.2015.03.012>.
- [113] D.A Baker, Cyclic nucleotide signalling in malaria parasites, *Cell. Microbiol.* 13 (2011) 331-339, <http://dx.doi.org/10.1111/j.1462-5822.2010.01561.x>.
- [114] L. Makin, E. Gluenz, cAMP signalling in trypanosomatids: role in pathogenesis and as a drug target, *Trends Parasitol.* 31 (2015) 373-379, <http://dx.doi.org/10.1016/j.pt.2015.04.014>.
- [115] D.N.A Tagoe, T.D. Kalejaiye, H.P. de Koning, The ever unfolding story of cAMP signaling in trypanosomatids: vive la difference!, *Front. Pharmacol.* 6 (2015) 185, <http://dx.doi.org/10.3389/fphar.2015.00185>.

- [116] M.Y. Galperin, Structural classification of bacterial response regulators: diversity of output domains combination, *J. Bacteriol.* 188 (2006) 4169-4182, <http://dx.doi.org/10.1128/JB.01887-05>.
- [117] M. Vogel, M. Bashton, N.D. Kerrison, C. Chothia, S.A. Teichmann, Structure, function and evolution of multidomain proteins. *Curr. Opin. Struct. Biol.* 14 (2004) 208-216, <http://dx.doi.org/10.1016/j.sbi.2004.03.011>.
- [118] S.B. Roberts, J.L. Robichaux, A.K. Chavali, P.A. Manque, V. Lee, A.M. Lara, et al., Proteomic and network analysis characterize stage-specific metabolism in *Trypanosoma cruzi*, *BMC Syst. Biol.* 3 (2009) 52, <http://dx.doi.org/10.1186/1752-0509-3-52>.
- [119] J.L. Concepción, C.A. Adjé, W. Quiñones, N. Chevalier, M. Dubourdiou, P.A. Michels, The expression and intracellular distribution of phosphoglycerate kinase isoenzymes in *Trypanosoma cruzi*, *Mol. Biochem. Parasitol.* 118 (2001) 111-121, [http://dx.doi.org/10.16/S0166\(01\)00381-4](http://dx.doi.org/10.16/S0166(01)00381-4).
- [120] X. Barros-Álvarez, A.J. Cáceres, M.T. Ruiz, P.A. Michels, J.L. Concepción, W. Quiñones, The phosphoglycerate kinase isoenzymes have distinct roles in the regulation of carbohydrate metabolism in *Trypanosoma cruzi*, *Exp. Parasitol.* 143 (2014) 39-47, <http://dx.doi.org/10.1016/j.exppara.2014.05.010>.
- [121] X. Barros-Álvarez, A.J. Cáceres, M.T. Ruiz, P.A. Michels, J.L. Concepción, W. Quiñones W, The glycosomal-membrane associated phosphoglycerate kinase isoenzyme A plays a role in sustaining the glucose flux in *Trypanosoma cruzi* epimastigotes, *Mol. Biochem. Parasitol.* 200 (2015) 5-8, <http://dx.doi.org/10.1016/j.molbiopara.2015.04.003>.
- [122] F.J. Adroher, A. Osuna, J.A. Lupiáñez, Differential energetic metabolism during *Trypanosoma cruzi* differentiation. II. Hexokinase, phosphofructokinase and pyruvate kinase, *Mol. Cell. Biochem.* 94 (1990) 71-82.
- [123] A. Konagurthu, J. Whisstock, P.J. Stuckey, A.M. Lesk, MUSTANG: A multiple structural alignment algorithm, *Proteins: Struct. Funct. Bioinf.* 64 (2006) 559-574, <http://dx.doi.org/10.1002/prot.20921>.

- [124] M. Shatsky, R. Nussinov, H.J. Wolfson, Optimization of multiple-sequence alignment based on multiple-structure alignment, *Proteins: Struct. Funct. Bioinf.* 62 (2006) 209-217, <http://dx.doi.org/10.1002/prot.20665>.
- [125] E. Zelzer, P. Wappner, B.Z. Shilo, The PAS domain confers target gene specificity of *Drosophila* bHLH/PAS proteins, *Genes Dev.* 11 (1997) 2079-2089, <http://dx.doi.org/10.1101/gad.11.16.2079>.
- [126] R. Adaxo, C.A. Harley, A.F. Castro-Rodrigues, J.H. Morais-Cabral, Structural properties of PAS domains from the KCNH potassium channels, *PLoS ONE* 8 (2013) e59265, <http://dx.doi.org/10.1371/journal.pone.0059265>.
- [127] P. Deschamps, E. Lara, E. Marande, P. Lopez-García, F. Ekelund, D. Moreira, Phylogenomic analysis of kinetoplastids supports that trypanosomatids arose from within bodonids, *Mol. Biol. Evol.* 28 (2011) 53-58, <http://dx.doi.org/10.1093/molbev/msq289>.
- [128] S. Kelly, A. Ivens, P.T. Manna, W. Gibson, M.C. Field, A draft genome for the African crocodylian trypanosome *Trypanosoma grayi*, *Sci. Data* 1 (2014) 140024, <http://dx.doi.org/10.1038/sdata.2014.24>.
- [129] F.R. Opperdoes, A. Butenko, P. Flegontov, V. Yurchenko, J. Lukeš, Comparative metabolism of free-living *Bodo saltans* and parasitic trypanosomatids. *J. Eukaryot. Microbiol.* 63 (2015) 657-678, <http://dx.doi.org/10.1111/jeu.12315>.
- [130] A.J. Cáceres, W. Quiñones, M. Gualdrón, A. Cordeiro, L. Avilán, P.A. Michels, et al., Molecular and biochemical characterization of novel glucokinases from *Trypanosoma cruzi* and *Leishmania* spp., *Mol. Biochem. Parasitol.* 156(2007) 235-245, <http://dx.doi.org/10.1016/j.molbiopara.2007.08.007>.
- [131] K.R. Matthews, The developmental cell biology of *Trypanosoma brucei*, *J. Cell Sci.* 118 (2005) 283-290, <http://dx.doi.org/10.1242/jcs.01649>.
- [132] M.R. Domingo-Sananes, B. Szöör, M.A. Ferguson, M.D. Urbaniak, K.R. Matthews, Molecular control of irreversible bistability during trypanosome developmental commitment, *J. Cell Biol.* 211 (2015) 455-468, <http://dx.doi.org/10.1083/jcb.201506114>.

- [133] Y. Nishizuka, The molecular heterogeneity of protein kinase C and its implications for cellular regulation, *Nature* 334 (1988) 661–665, <http://dx.doi.org/10.1038/334661a0>.
- [134] M.J. Wainszelbaum, M.L. Belaunzaran, E.M. Lammel, M. Florin-Christensen, J. Florin-Christensen, E.L.D. Isola, Free fatty acids induce cell differentiation to infective forms in *Trypanosoma cruzi*, *Biochem. J.* 305 (2003) 705-712, <http://dx.doi.org/10.1042/BJ20021907>.
- [135] M. García de Lema, E.E. Aeberhard, Desaturation of fatty acids in *Trypanosoma cruzi*, *Lipids* 21 (1986) 718-720, <http://dx.doi.org/10.1007/BF02537247>.
- [136] J.F. Glatz, J. Storch, Unravelling the significance of cellular fatty acid-binding proteins, *Curr. Opin. Lipidol.* 12 (2001) 267-274, <http://dx.doi.org/10.1097/00041433-200106000-00005>.
- [137] C. Yernaux, M. Fransen, C. Brees, S. Lorenzen, P.A.M. Michels, *Trypanosoma brucei* glycosomal ABC transporters: identification and membrane targeting, *Mol. Membr. Biol.* 23 (2006) 157-172, <http://dx.doi.org/10.1080/09687860500460124>.
- [138] M. Igoillo-Esteve, M. Mazet, G. Deumer, P. Wallemacq, P.A.M. Michels, Glycosomal ABC transporters of *Trypanosoma brucei*: characterisation of their expression, topology and substrate specificity, *Int. J. Parasitol.* 41 (2011) 429-438, <http://dx.doi.org/10.1016/j.ijpara.2010.11.00>.
- [139] A.M. Silber, R.R. Tonelli, C.G. Lopes, N. Cunha-e-Silva, A.C.T. Torrecilhas, R.I. Schumacher, W. Colli, M.J.M. Alves, Glucose uptake in the mammalian stages of *Trypanosoma cruzi*, *Mol. Biochem. Parasitol.* 168 (2009) 102-108, <http://dx.doi.org/10.1016/j.molbiopara.2009.07.006>.
- [140] M. García de Lema, G. Lucchesi, G. Racagni, E.E Machado-Domenech, Changes in enzymatic activities involved in glucose metabolism by acyl-CoAs in *Trypanosoma cruzi*, *Can. J. Microbiol.* 47 (2001) 49-54, <http://dx.doi.org/10.1139/w00-120>.
- [141] F. Nagajyothi, F.S. Machado, B.A. Burleigh, L.A. Jelicks, P.E. Scherer, S. Mukherjee, M.P. Lisanti, L.M. Weiss, N.J. Garg, H.B. Tanowitz, Mechanisms of

*Trypanosoma cruzi* persistence in Chagas disease, *Cell. Microbiol.* 14 (2012) 634–643, <http://dx.doi.org/10.1111/j.1462-5822.2012.01764.x>.

[142] K.L. Caradonna, J.C. Engel, D. Jacobi, C.H. Lee, B.A. Burleigh, Host metabolism regulates intracellular growth of *Trypanosoma cruzi*, *Cell Host Microbe* 13 (2013) 108–117, <http://dx.doi.org/10.1016/j.chom.2012.11.011>.

[143] K.M. Tyler, C.L. Olson, D.M. Engman, The life-cycle of *Trypanosoma cruzi*. In: K.M. Tyler, M.A. Miles (Eds.), *American trypanosomiasis. World class parasites*, Kluwer Academic Publishers, Boston, 2003, pp. 1–11.

[144] C.Y. Taabazuing, J.A. Hangasky, M.J. Knapp, Oxygen sensing strategies in mammals and bacteria, *J. Inorg. Biochem.* 133 (2014) 63–72, <http://dx.doi.org/10.1016/j.jinorbio.2013.12.010>.

[145] M. Müller, M. Mentel, J.J. Van Hellemond, K. Henze, C. Woehle, S.B. Gould, et al., Biochemistry and evolution of anaerobic energy metabolism in eukaryotes, *Microbiol. Mol. Biol. Rev.* 76 (2012) 444–495, <http://dx.doi.org/10.1128/MMBR.05024-11>.

[146] T. Uchida, E. Sato, A. Sato, I. Sagami, T. Shimizu, T. Kitagawa, CO-dependent activity-controlling mechanism of heme-containing CO-sensor protein, neuronal PAS domain protein 2, *J. Biol. Chem.* 280 (2005) 21358–21368, <http://dx.doi.org/10.1074/jbc.M412350200>.

[147] T. Shimizu, The heme-based oxygen sensor phosphodiesterase Ec-Dos (DosP): structure-function relationships, *Biosensors* 3 (2013) 211–237, <http://dx.doi.org/10.3390/bios3020211>.

[148] Y.B. Chen, T.C. Lu, H.X. Wang, J. Shen, T.T. Bu, Q. Chao, et al., Posttranslational modification of maize chloroplast pyruvate orthophosphate dikinase reveals the precise regulatory mechanism of its enzymatic activity, *Plant Physiol.* 165 (2014) 534–549, <http://dx.doi.org/10.1104/pp.113.231993>.

- [149] B. Szöör, I. Ruberto, R. Burchmore, K.R. Matthews, A novel phosphatase cascade regulates differentiation in *Trypanosoma brucei* via a glycosomal signaling pathway, *Genes Dev.* 24 (2010) 1306-1316, <http://dx.doi.org/10.1101/gad.570310>.
- [150] B. Greb-Markiewicz, D. Sadowska, N. Surgut, J. Godlewski, M. Zarebski, A. Ozyhar, Mapping of the sequences directing localization of the *Drosophila* germ cell-expressed protein (GCE), *PLoS ONE* 10 (2015) e0133307, <http://dx.doi.org/1371/journal.pone.0133307>.
- [151] T.J. Bernardo, E.B. Dubrovsky, Molecular mechanisms of transcription activation by juvenile hormone: a critical role for bHLH-PAS and nuclear receptor proteins, *Insects* 3 (2012) 324-338, <http://dx.doi.org/10.3390/insects3010324>.
- [152] M Chen, Y. Tao, J. Lim, A. Shaw, J. Chory, Regulation of phytochrome B nuclear localization through light-dependent unmasking of nuclear-localization signals, *Curr. Biol.* 15 (2005) 637-642, <http://dx.doi.org/10.1016/j.cub.2005.02.028>.
- [153] W. Yang, Z. Lu, Pyruvate kinase M2 at a glance, *J. Cell Sci.* 128 (2015) 1655-1660, <http://dx.doi.org/10.1242/jcs.166629>.
- [154] X. Li, Y. Jiang, J. Meisenhelder, W. Yang, D.H. Hawke, Y. Zheng, Mitochondria-translocated PGK1 functions as a protein kinase to coordinate glycolysis and the TCA cycle in tumorigenesis, *Mol. Cell* 61 (2016) 705-719, <http://dx.doi.org/10.1016/j.molcel.2016.02.009>.
- [155] X. Qian, X. Li, Q. Cai, C. Zhang, Q. Yu, Y. Jiang, et al., Phosphoglycerate kinase 1 phosphorylates Beclin1 to induce autophagy, *Mol. Cell* 65 (2017) 917-931, <http://dx.doi.org/10.1016/j.molcel.2017.01.027>.
- [156] J. Ellis, M. Sarkar, E. Hendriks, K. Matthews, A novel ERK-like, CRK-like protein kinase that modulates growth in *Trypanosoma brucei* via an autoregulatory C-terminal extension, *Mol. Microbiol.* 53 (2004) 1487-1499, <http://dx.doi.org/10.1111/j.1365-2958.2004.04218.x>.

- [157] E.C. Mattos, R.I. Schumacher, W. Colli, M.J.M. Alves, Adhesion of *Trypanosoma cruzi* trypomastigotes to fibronectin or laminin modifies tubulin and paraflagellar rod protein phosphorylation, PLoS ONE 7 (2012) e46767, <http://dx.doi.org/10.1371/journal.pone.0046767>.
- [158] E.D. Erben, S. Daum, M.T. Téllez-Iñón, The *Trypanosoma cruzi* PIN1 gene encodes a parvulin peptidyl-prolyl cis/trans isomerase able to replace the essential ESS1 in *Saccharomyces cerevisiae*, Mol. Biochem. Parasitol. 153 (2007) 186-193, <http://dx.doi.org/10.1016/j.molbiopara.2007.03.004>.
- [159] J.Y. Goh, C.Y. Lai, L.C. Tan, D. Yang, C.Y. He, Y.C. Liou, Functional characterization of two novel parvulins in *Trypanosoma brucei*, FEBS Lett. 584 (2010) 2901-2908, <http://dx.doi.org/10.1016/j.febslet.2010.04.077>.
- [160] E.D. Erben, E. Valguarnera, S. Nardelli, J. Chung, S. Daum, M. Potenza, et al., Identification of an atypical peptidyl-prolyl cis/trans isomerase from trypanosomatids, Biochim. Biophys. Acta 1803 (2010) 1028–1037, <http://dx.doi.org/10.1016/j.bbamcr.2010.05.006>.
- [161] M. Herman, D. Pérez-Morga, N. Schtickzelle, P.A. Michels, Turnover of glycosomes during life-cycle differentiation of *Trypanosoma brucei*, Autophagy 4 (2008) 294-308, <http://dx.doi.org/10.4161/auto.5443>.
- [162] M. Duszenko, M.L. Ginger, A. Brennand, M. Gualdrón-López, M.I. Colombo, G.H. Coombs, et al., Autophagy in protists, Autophagy 7 (2011) 127-158, <http://dx.doi.org/10.4161/auto.7.2.13310>.
- [163] S. Besteiro, R.A. Williams, L.S. Morrison, G.H. Coombs, J.C. Mottram, Endosome sorting and autophagy are essential for differentiation and virulence of *Leishmania major*, J. Biol. Chem. 281(2006) 11384-11396, <http://dx.doi.org/10.1074/jbc.M512307200>.
- [164] A. Barquilla, M. Navarro, Trypanosome TOR as a major regulator of cell growth and autophagy, Autophagy 5(2009) 256-258.



- [165] F.J. Li, Q. Shen, C. Wang, Y. Sun, A.Y. Yuan, C.Y. He, A role of autophagy in *Trypanosoma brucei* cell death, *Cell. Microbiol.* 14 (2012) 1242-1256, <http://dx.doi.org/10.1111/j.1462-5822.2012.01795.x>.
- [166] B. Cull, J.L. Prado Godinho, J.C. Fernandes Rodrigues, B. Frank, U. Schurigt, R.A. Williams, G.H. Coombs, J.C. Mottram, Glycosome turnover in *Leishmania major* is mediated by autophagy, *Autophagy* 10 (2014) 2143-2157, <http://dx.doi.org/10.4161/auto.36438>.
- [167] M. Hashimoto, J. Morales, H. Uemura, K. Mikoshiba, T. Nara, A novel method for inducing amastigote-to-trypomastigote transformation *in vitro* in *Trypanosoma cruzi* reveals the importance of inositol 1,4,5-trisphosphate receptor, *PLoS ONE* 10 (2015) e0135726, <http://dx.doi.org/10.1371/journal.pone.0135726>.
- [168] J.D. Thompson, T.J. Gibson, D.G. Higgins, Multiple sequence alignment using ClustalW and ClustalX, *Curr Protoc Bioinformatics* 2002; Chapter 2:Unit 2.3, <http://dx.doi.org/10.1002/0471250953.bi0203s00>.
- [169] A.M. Waterhouse, J.B. Procter, D.M. Martin, M. Clamp, G.J Barton, Jalview Version 2--a multiple sequence alignment editor and analysis workbench, *Bioinformatics* 25 (2009) 1189-91, <http://dx.doi.org/10.1093/bioinformatics/btp033>.

## Figure legends

**Fig. 1.** Structure of photoactive yellow protein (PYP) [PDB code: 2phy].

The structure of this photoactive protein is divided in four segments: N-terminal cap (purple), PAS core (orange), helical connector (green) and  $\beta$ -scaffold (blue) [10].

**Fig. 2.** Structure-based alignment (MUSTANG/STACCATO).

Structure-based alignment (MUSTANG/STACCATO) of trypanosomatid PGK PAS domains with five crystal structures of PAS domains with diverse ligand binding specificities (FMN [PDB code: 3ue6], FAD [2gj3], haem [3vol], chromophore [1nwz] and fatty acid [3k3c]). The alignment is coloured using the ClustalX [168] scheme as implemented in the program Jalview [169]. White on magenta is used to highlight ligand-binding residues in the crystal structures. The upper five sequences are labelled according to their identifiers in the TriTryp database [85] while the lower five structures are labelled with PDB IDs (first four characters) followed by chain ID (fifth character).

**Fig. 3.** Phylogenetic analysis of PAS domains in protein sequences from kinetoplastids.

Sequences of PAS domains were identified by a BLAST search in different protein kinases and PGKs of different kinetoplastids (trypanosomatids and bodonids) as listed in Table 1, and aligned using MUSCLE with default parameter setting. All positions containing gaps in the alignment and missing data were eliminated. There were a total of 48 positions in the final dataset. The phylogenetic relationships of the PAS sequences were subsequently inferred using the Neighbor-Joining method. Evolutionary analyses were conducted in MEGA7. Numbers at the individual nodes of the tree represent bootstrap support (500 replicates). The horizontal bar represents the units of the number of amino-acid substitutions per site.

**Fig. 4.** Crystal structure of cytosolic *T. brucei* PGK coloured from blue to red, N- to C-terminus.

Substrates in this ternary complex (PDB code: 13pk) are shown as balls and sticks. The N-terminus (blue) is far from the catalytic site but near the hinge region (pink).

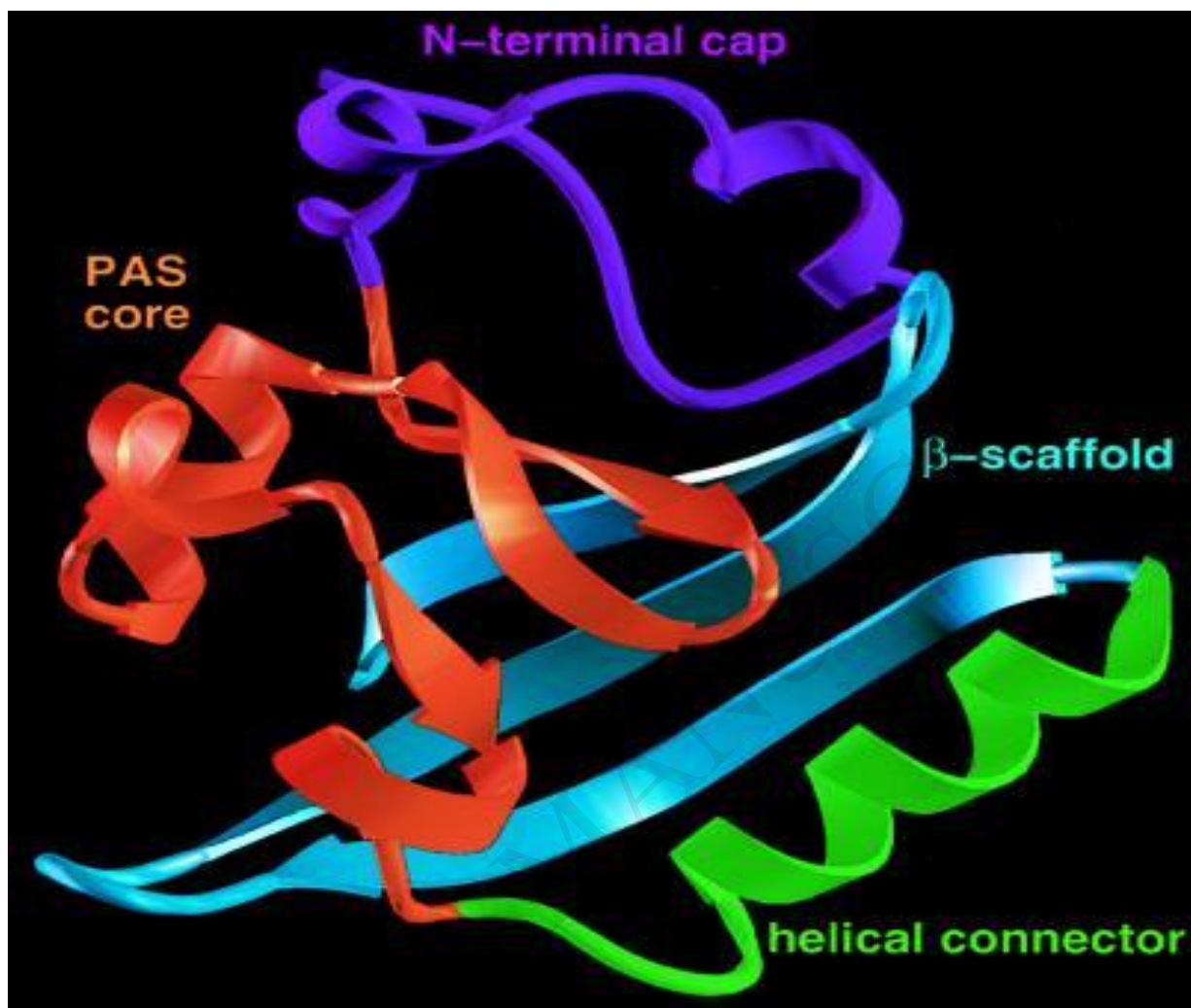


Figure 1

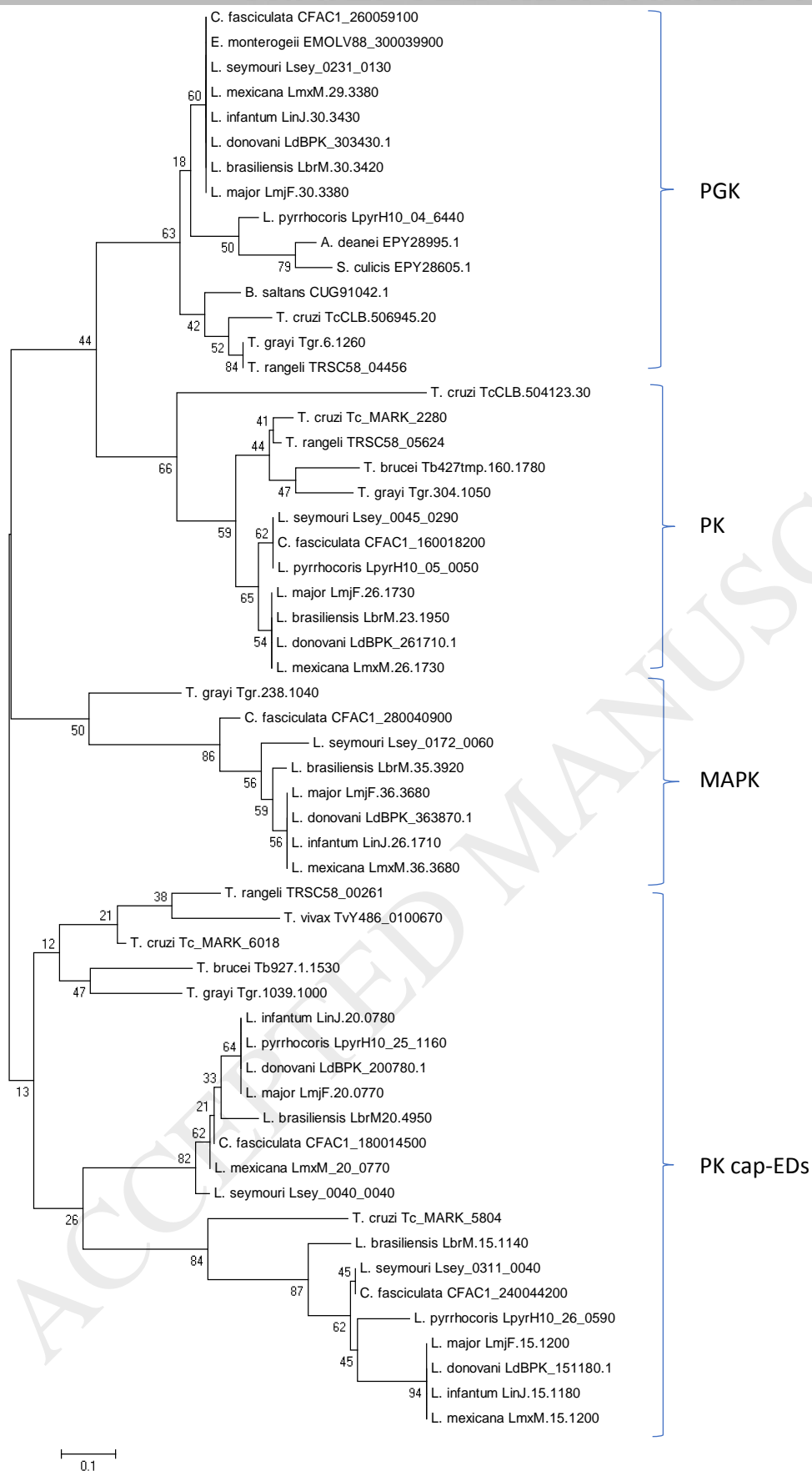
ACCEPTED

Figure 2.

*Tc00.1047053506945.20* 1 ----- MLQHSKIDM---- IFKS LRDVVVVGD --- NDMVI LDMNRSACDFFG ---- 39  
*EmtLcontig3-4-120038-118446* 1 --- ARGSMYQESKISM---- IFRSLRDMVVI GD --- NDMVITGMNDPAASFFG ---- 43  
*LbrM.30.3420* 1 ----- MYQESKIFM---- IFKS LRDIVVIGD --- NDMVITDMNAAAFAFFG ---- 39  
*LtaP30.3420* 1 ----- MYEDSKILM---- IFKS LRDIVVIGD --- NDMVITDMNAAAFAFFG ---- 39  
*LmjF.30.3380* 1 ----- MYQDSKILM---- IFKS LRDIVVIGD --- NDMVITDMNAAAFAFFG ---- 39  
*3ue6A* 1 ----- EDPDYSLVK--- ALQMAQQNFVITDASLPDNP IYVASRGFLTITG ---- 42  
*2gj3A* 1 ----- LLPEI - FRQ--- TVEHAPIAISITDL --- KANILYANRAFRITIG ---- 38  
*3volA* 1 ----- GSHMARIKSALDNVSANVM IADN --- DLNI IYMNRTVSEMLGRAEA 43  
*1nwzA* 1 MEHVAFGSEDIENTLAKMDDGQLDGLAFGA IQLDG --- DGNILQYNAAEFGDITG ---- 51  
*3k3cA* 1 GAMA AEMDWDKTVGAAEDVRRIFEHIPAILVGLLEG --- PDHRFVAVNAAYRGFSP ---- 52

*Tc00.1047053506945.20* 40 ----- WSVEEVRGKSI TFLIPKYP L ----- DVSDGTVTLVARLKN SVDIPVVIQ 83  
*EmtLcontig3-4-120038-118446* 44 ----- WTADDVKGKSI TFLVPTHP L ----- DVQDNTVTLMARLANSTDPVVIQ 87  
*LbrM.30.3420* 40 ----- WTADDVKGKSI AFLVPTHP L ----- DVQDNTVTLMAHLANSTDPVVIQ 83  
*LtaP30.3420* 40 ----- WKIDDVKGKSI TFLVPTHP L ----- DVQDNTVTLMAHLANSTDPVVIQ 83  
*LmjF.30.3380* 40 ----- WKVDDVKGKSI TFLVPTHP L ----- DVQDNTVTLMAHLANSTDPVVIQ 83  
*3ue6A* 43 ----- YSLDQILGRNC - RFLQGPETDPRAVDKIRNAITKGVDT SVC ----- LLN 85  
*2gj3A* 39 ----- YGSEEVLGKNE - SILSNGTTPRLVYQALWGR LAQKPPWSGV ----- LVN 81  
*3volA* 44DIRKQLPNFDAGR LMGANIDV FHKN ----- PATQRHLLANLTGVHKA ----- ELN 88  
*1nwzA* 52 ----- RDPKQVIGKNF FFKDVAPCTDSPEFYKFKEGVASGNLNTMF ----- EYT 95  
*3k3cA* 53 ----- LLDT - VGQPA - REVYPELEGQIYEMLDRVYQTGE PQSGSEW ----- RLO 95

*Tc00.1047053506945.20* 84 STRDPH --- ASMVGWTILPIILPRVDFFGVHMN LRA ----- 116  
*EmtLcontig3-4-120038-118446* 88 TTRDPH --- ATMVAWTILPIRLPRVYFEGVHMN LRA ----- 120  
*LbrM.30.3420* 84 TTRDPH --- ATMVAWTILPIRLPRVYFEGVHMN LRA ----- 116  
*LtaP30.3420* 84 TTRDPH --- ATMVAWTILPIRLPCVYFEGVIN LRA ----- 116  
*LmjF.30.3380* 84 TTRDPH --- ATVVAWTILPIRLPRVYFEGVIN LRA ----- 116  
*3ue6A* 86 YRQDGT --- TFWNLFFVAGLRDSKGNIVNYVGVQSKVSEYAKLLVNEQNI EY ---- 135  
*2gj3A* 82 RRKDKT --- LYLAELTVAPVLNEAGETIYYLGMH RDTSELH ----- 119  
*3volA* 89 LG ----- GRRFSLDVVPVFNDANERLGS AVQWTDRTTEEHRAEQEVSQLVQAAAAG 138  
*1nwzA* 96 FDYQMT --- PTKVKVHMKKALS --- GDSYWFVKRV ----- 125  
*3k3cA* 96 IDYDGS GVEERYFDVVT PRRRADGSI EG VQLIVDQVTSRVRARQAAEARVEELSER 152



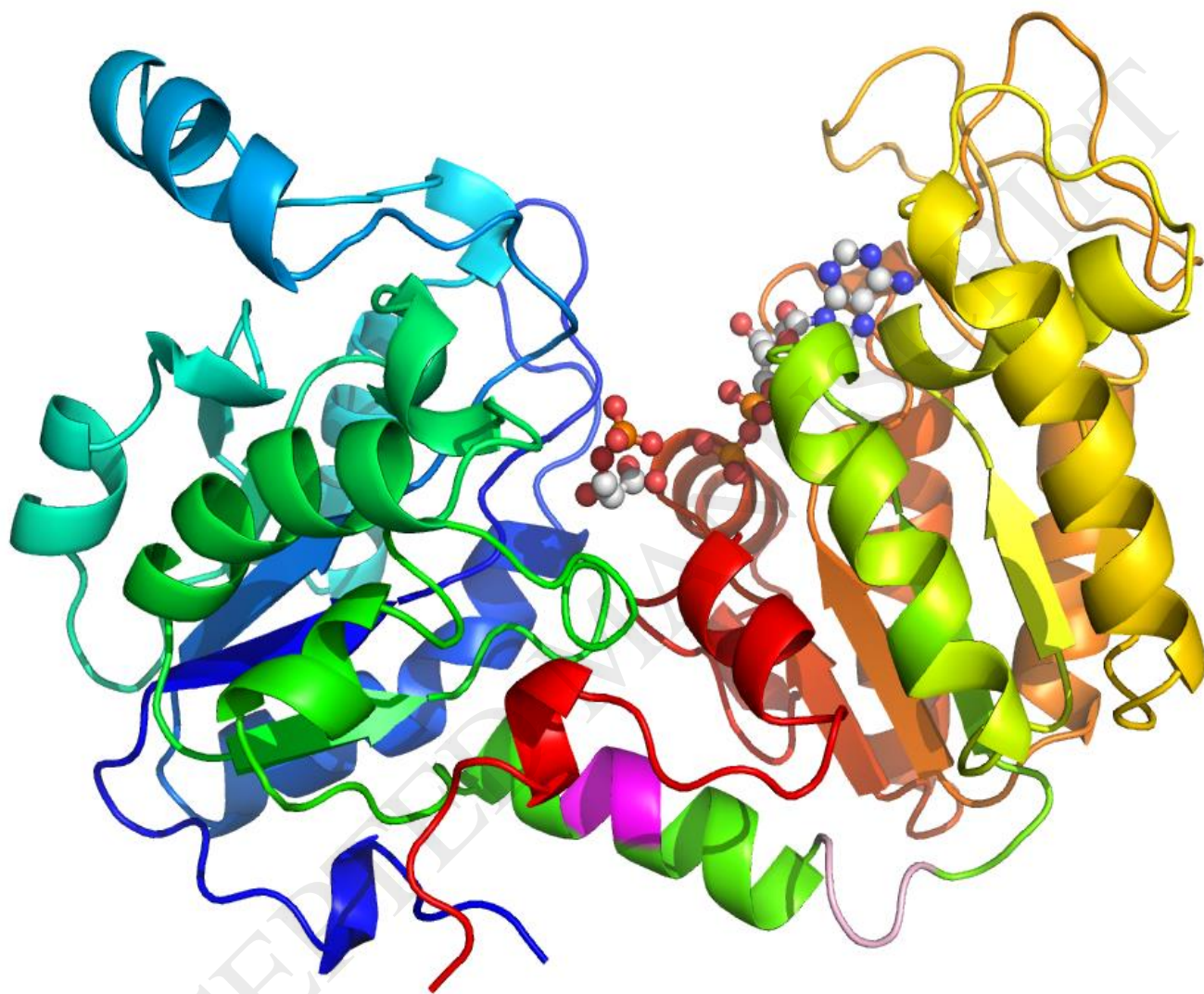


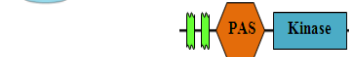
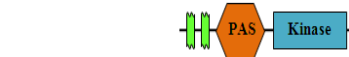

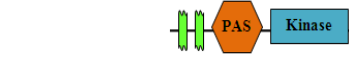



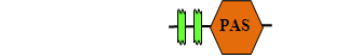



**Figure 4.**





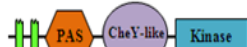










Table 1. Detectable proteins containing PAS domains in complete kinetoplastid genomes

Organism	Product	Localization Genome (Chromosome)	Associated domains	Mass (Da)	pI	Gene ID
<b>Free-living</b>						
<i>Bodo</i>						
<i>Bodo saltans</i>	PAS PAC sensor-like protein, putative	Not assigned		258487	5.50	BSAL_46115
	Phosphoglycerate kinase, putative	Not assigned		57000	9.10	CUG91042.1
<b>Obligate Parasites</b>						
<i>Cryptobia</i>						
<i>Trypanoplasma borreli</i>	Unspecified product	Not assigned		-----	-----	NODE_142343_length_19892_cov_7.208827
<b>Trypanosoma</b>						
<i>Stercorarian and reptilian Trypanosomes</i>						
<i>T. cruzi</i> CL Brener Esmeraldo-like	PAS-domain containing phosphoglycerate kinase, putative	32		57944	9.26	TcCLB.506945.20
	STE/STE11 serine/threonine-protein kinase, putative	33		112256	6.54	TcCLB.510565.70
	Serine/threonine-protein kinase, putative (fragment)	7		102877	5.78	TcCLB.505977.13
	STE group serine/threonine-protein kinase, putative	9		204034	6.17	TcCLB.508995.10
<i>T. rangeli</i>	PAS-domain containing phosphoglycerate kinase, putative	Not assigned		57843	9.52	TRSC58_04456
	Protein kinase, putative	Not assigned		149563	6.47	TRSC58_00261
	Protein kinase, putative	Not assigned		111547	6.96	TRSC58_05624

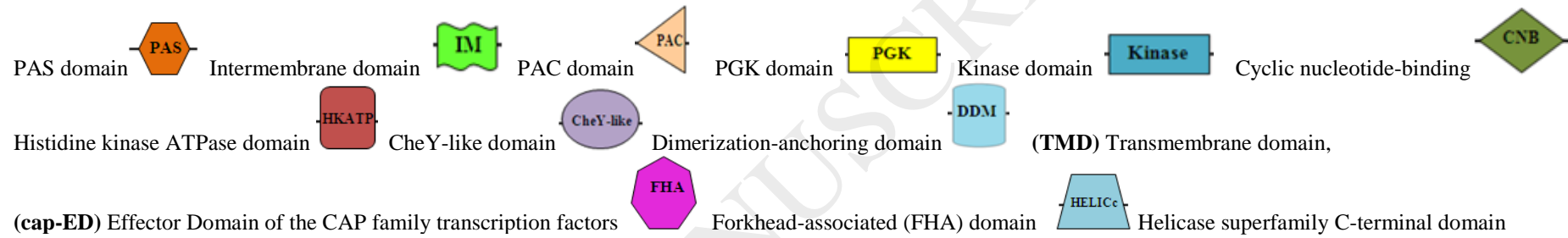


<i>T. grayi</i>	PAS-domain containing phosphoglycerate kinase, putative	Not assigned		57652	9.42	Tgr.6.1260
	Protein kinase, putative (cap-ED)	Not assigned		198289	5.32	Tgr.1039.1000
	Protein kinase, putative (TMD)	Not assigned		111990	7.11	Tgr.304.1050
	Mitogen activated kinase-like protein	Not assigned		110006	6.78	Tgr.238.1040
<i>Salivarian Trypanosomes</i>						
<i>T. brucei brucei</i> <b>TREU927</b>	STE Protein kinase, putative (cap-ED)	1		191477	6.11	Tb927.1.1530
	STE /STE 11 serine /threonine- protein kinase, putative	9		112317	5.33	Tb427.9.3 120
<i>T. congolense</i>	Protein kinase, putative	9		112039	6.91	TcIL3000_9_920
	Protein kinase, putative	Not assigned		194502	5.68	TcIL3000_0_12910
	Protein kinase, putative	Not assigned		200028	5.79	TcIL3000_0_00130
	Protein kinase, putative (fragment)	Not assigned		61609	7.54	TcIL3000_0_43080
<i>T. vivax</i>	Protein kinase, putative	1		197789	6.09	TvY486_0100670
	Protein kinase, putative	Not assigned		197792	6.25	TvY486_0027590
	Protein kinase, putative (fragment)	9		97849	6.29	TvY486_0901000
<i>Leishmania</i>						



<i>L. major</i>	PAS-domain containing phosphoglycerate kinase, putative	30		57570	8.75	LmjF.30.3380
	STE group serine/threonine-protein kinase, putative (cap-ED)	15		287443	6.58	LmjF.15.1200
	STE group serine/threonine-protein kinase, putative (cap-ED)	20		419207	6.57	LmjF.20.0770
	STE/STE 11 serine/ threonine- protein kinase, putative (TMD)	26		113736	6.56	LmjF.26.1730
	Mitogen activated kinase-like protein	36		108204	6.90	LmjF.36.3680
<b><i>Leptomonas</i></b>						
<i>L. seymouri</i>	PAS-domain containing phosphoglycerate kinase, putative	Not assigned		57348	9.10	Lsey_0231_0130
	Protein kinase, putative (cap-ED)	Not assigned		449325	7.00	Lsey_0040_0040
	Protein kinase, putative (TMD)	Not assigned		114214	6.89	Lsey_0045_0290
	Mitogen activated kinase-like protein	Not assigned		102095	5.56	Lsey_0172_0060
	Protein kinase, putative (cap-ED)	Not assigned		315051	7.32	Lsey_0311_0040
<b><i>Crithidia</i></b>						
<i>C. fasciculata</i>	PAS-domain containing phosphoglycerate kinase, putative	26		57208	9.60	CFAC1_260059100
	Protein kinase, putative (TMD)	16		114077	6.50	CFAC1_160018200
	Protein kinase, putative (cap-ED)	18		470028	6.40	CFAC1_180014500
	Protein kinase, putative	24		323033	6.93	CFAC1_240044200
	Mitogen activated kinase-like protein	28		110426	6.70	CFAC1_280040900

PAS domain-containing sequences were found in the TriTrypDB, GeneDB and NCBI databases using BLAST. Their domain composition was then assessed using the software tools from ExPaSy, SMART, ProteinBlast and InterPro EMBL-EBI.



----- : This sequence contains several consecutive undefined AA. Its pI and Mw cannot be computed

This table presents proteins containing PAS domains as detectable in representative species of kinetoplastid genera. Information about PAS proteins of additional species of the genera *Trypanosoma*, *Leishmania* and *Leptomonas* can be found in Table S1 of the Supplementary Information.