The structure of a Trypanosoma cruzi glucose-6-phosphate dehydrogenase reveals differences from the mammalian enzyme

Mercaldi, Gustavo F.; Dawson, Alice; Hunter, William N. ; Cordeiro, Artur T.

Published in:
FEBS Letters

DOI:
10.1002/1873-3468.12276

Publication date:
2016

Document Version
Peer reviewed version

The structure of a *Trypanosoma cruzi* glucose-6-phosphate dehydrogenase ternary complex with substrate and cofactor reveals differences from the mammalian enzyme

Gustavo F. Mercaldi[ab], Alice Dawson[c], Willian N. Hunter[c], Artur T. Cordeiro[a]

[a] *Brazilian Biosciences National Laboratory, Center of Research in Energy and Materials, Brazil*
[b] *Institute of Biology, University of Campinas, Brazil*
[c] *Division of Biological Chemistry and Drug Discovery, School of Life Sciences, University of Dundee, UK*

Correspondence
A. T. Cordeiro, Brazilian Biosciences National Laboratory, Center of Research in Energy and Materials, R. Giuseppe Máximo Scolfaro 10000, Campinas CEP 13083-970, Brazil
Fax: +55 19 3512.1006
Tel: +55 19 3512.1121
email: artur.cordeiro@lnbio.cnpem.br

ABSTRACT

The enzyme glucose-6-phosphate dehydrogenase from *Trypanosoma cruzi* (*Tc*G6PDH) catalyses the first step of the pentose phosphate pathway and is considered a promising target for the discovery of a new drug against Chagas Diseases. In the present work, we describe the crystal structure of *Tc*G6PDH obtained in a ternary complex with the substrate glucose-6-phosphate and the reduced 'catalytic' cofactor NADPH, which reveals the molecular basis of substrate and cofactor recognition. A comparison with the homologous human protein sheds light on differences in the cofactor-binding site that might be explored towards the design of new $NADP^+$ competitive inhibitors targeting the parasite enzyme.

Keywords: glucose-6-phosphate dehydrogenase; ternary-complex; Chagas disease; trypanosomatids; uncompetitive inhibitors; drug target.

1. Introduction

Glucose-6-phosphate dehydrogenase (G6PDH, EC 1.1.1.49) is a ubiquitous enzyme that catalyses the oxidation of β-D-glucose-6-phosphate (G6P) to 6-phosphoglucono-δ-lactone, with concomitant reduction of $NADP^+$ to NADPH, in the first and rate limiting step of the pentose phosphate pathway (PPP). NADPH generated in the oxidative branch of the PPP is used as a reducing agent essential for redox balance maintenance and for lipid biosynthesis [1,2]. Furthermore, the non-oxidative branch of the PPP produces a variety of metabolic intermediates that includes ribose-5-phosphate, used in nucleotide biogenesis, and erythrose 4-phosphate, a precursor of coenzymes and aromatic amino acids [3].

The enzyme G6PDH plays a key functional role [2,4] with implications in human diseases such as cancer [5], metabolic disorders [6,7], and cardiovascular diseases [8]. G6PDH is associated with the most common human enzymopathy, affecting more than 400 million people worldwide [9]. Individuals with G6PDH deficiency may develop hemolytic anemia under oxidative stress, which ~~could~~ can be triggered by factors like microbial infections, certain drugs or aspects of diet. In protozoan parasites, G6PDH has been shown to be essential for survival [10]. In bloodstream form *Trypanosoma brucei*, depletion of G6PDH levels by RNAi [11] or its inhibition by steroids, like dehydroepiandrosterone or epiandrosterone, kill parasites *in vitro* [12]. Additionally, steroids and quinazolinones were shown to inhibit the *T. cruzi* G6PDH and to kill the epimastigote forms of the parasite *in vitro* [13,14], suggesting that G6PDH is an attractive target for development of new trypanocidal drugs. Some efforts have already been made to develop G6PDH inhibitors with therapeutic utility against trypanosomiasis [11,13,14] and also cancer [15,16]. Steroids and quinazolinones represent the most potent G6PDH inhibitors known to date. These inhibitors act through an uncompetitive mechanism, but their binding site on G6PDH and the interactions relevant to inhibition are unknown.

At present, G6PDH structures of *Leuconostoc mesenteroides* (*Lm*G6PDH), *Mycobacterium avium* (*Ma*G6PDH), human (*Hs*G6PDH) and *Trypanosoma cruzi* (*Tc*G6PDH) are available in the Protein Data Bank. However, most of the structural knowledge about G6PDHs was established in the studies of *Lm*- and *Hs*G6PDH enzymes. Studies with recombinant *Lm*G6PDH revealed that the bacterial enzyme assemblies as a homodimer and that each subunit is composed of an NAD(P)-binding Rossmann-like domain and a β + α domain [17]. Site direct mutagenesis studies on *Lm*G6PDH revealed the substrate and cofactor binding sites and informed on the mechanism of catalysis [18-20]. The crystal structure of *Hs*G6PDH revealed an unprecedented tetrameric assembly and the binding of an additional $NADP^+$ to the β + α domain, which became known as the structural $NADP^+$ [21,22]. After a six-year gap, Ortíz and collaborators described the crystallization

of the short form of *Tc*G6PDH (*Tc*G6PDH-S; Met38 to Ala555) [23] and made available the coordinates of an apo and an enzyme-substrate complex (PDB entries: 4E9I and 4EM5, respectively). More recently the Seattle Structural Genomic Centre for Infectious Diseases deposited the structure of *Ma*G6PDH at 2.3 Å resolution [24]. To date, no detailed discussion of the *Ma*- and *Tc*-G6PDH structures have been reported.

In the present work, we describe the crystal structure of *Tc*G6PDH obtained in a ternary complex with G6P and NADPH and compare it to previous available G6PDH structures. A detailed analysis of the NADP-binding Rossmann-like domain revealed unique features that might be explored to the design of specific inhibitors against the parasite enzyme.

2. Methods

2.1. Truncated construct of G6PDH from *T. cruzi*

Based on the *T. cruzi* (CL Brener) G6PDH sequence available at the NCBI (entry: XP_820060.1), a nucleotide sequence, codon optimized for bacterial expression, was designed to produce a fragment of *Tc*G6PDH spanning from Asp58 to Thr545 (*Tc*ΔG6PDH). The synthetic gene was purchased (GenScript USA Inc., Piscataway, New Jersey, USA) and sub-cloned into a modified pET28 vector in which the sequence encoding a thrombin recognition site, to allow for removal of a histidine-tag, was changed to that for tobacco etch virus (TEV) protease (pET28-TEV). The *Tc*ΔG6PDH gene was inserted within BamH*I* and Xho*I* sites of the pET28-TEV and the integrity of the pET28-TEV_*Tc*ΔG6PDH construct was confirmed by ~~gene~~ sequencing.

2.2. Recombinant Protein Production and Purification

Competent *E. coli* BL21 (DE3) cells were transformed with pET28-TEV_*Tc*ΔG6PDH and grown on auto-induction media ZYM-5052 (Studier, 2005) containing kanamycin 50 $\mu g.mL^{-1}$, under 200 rpm agitation, at 37 ºC for 3h and then at 20 ºC for 21h. Cells were harvested by centrifugation at 3.500 g for 30 minutes at 4 ºC and resuspended in Buffer A (50 mM Tris-HCl pH 8.0, 0.5 M NaCl, 20 mM imidazole, 5% glycerol and 5 mM 2-mercaptoethanol). DNAse and EDTA-Free Protease Inhibitor Tablets (ThermoFisher) were added to the suspension and lysis performed with a cell disruptor (Pressure Cell Homogeneizer, Stansted) using a pressure of 20.000 psi. The samples were centrifuged at 4 ºC with a RCF equal to 40.000g for 30 min and the supernatant subjected to immobilized metal affinity chromatography (IMAC) using a 5 mL HisTrap HP (GE Healthcare). Buffer B (50 mM Tris-HCl pH 8.0, 0.5 M NaCl, 500 mM imidazole, 5%

glycerol and 5 mM 2-mercaptoethanol) was employed to generate an imidazole gradient and *Tc*ΔG6PDH eluted from the column with approximately 200 mM imidazole. Samples containing *Tc*ΔG6PDH were pooled and concentrated using an ultrafiltration unit (3.000 g at 4 ºC, Vivaspin 20 MWCO 30 kDa, Sartorius). As a final purification step, size exclusion chromatography was performed using a HiLoad Superdex 200 26/60 column (GE Healthcare) and GF buffer (20 mM Tris-HCl and 0.2 M NaCl). Then, *Tc*ΔG6PDH was concentrated again in GF buffer (20 mM Tris-HCl and 0.2 M NaCl) containing 5 mM of 2-mercaptoethanol.

The gene of *Tc*G6PDH long form (*Tc*G6PDH-L; Met1 to Ala555) was sub-cloned from the previous reported pET28_*Tc*G6PDH-L construct [25] to a pET28-TEV vector between *Nhe*I and *Xho*I ~~rectriction~~restriction sites. The integrity of the pET28-TEV_*Tc*G6PDH-L construct was confirmed by ~~gene~~ sequencing. *Tc*G6PDH-L harbouring an N-terminal His-tag was expressed in *E. coli* BL21 (DE3) cells transformed by pET28-TEV_*Tc*G6PDH-L construct. The recombinant protein was purified by IMAC following the same procedure applied to the purification of *Tc*ΔG6PDH.

2.4. Enzyme Kinetics

*Tc*ΔG6PDH and *Tc*G6PDH-L activities were measured following NADPH production in the forward reaction. NADPH fluorescence ($\lambda_{Exc/Em}$: 340 nm / 460 nm) was monitored using the FLUOstar OPTIMA plate reader (BMG LABTECH). Enzyme activities were measured in a buffer containing 50 mM Tris-HCl pH 7.6, 0.5 M NaCl, 5% glycerol and 2 mM 2-mercaptoethanol. The reactions were performed in triplicate at 25 ºC, using black 96 wells plates, with a final volume of 120 µL per well. Apparent Michaelis-Menten constants (K_m^{app}) and maximum velocity of reaction (V_{max}^{app}) values were calculated by non-linear regression of the data using the equation $v = (V_{max}^{app}.[S])/(K_m^{app} + [S])$ in the software GraphPad Prism. The catalytic constant (k_{cat}) were calculated using the equation $k_{cat} = V_{max}/E_t$, where E_t is the enzyme concentration used in the assay. For the *Tc*ΔG6PDH, the K_m^{app} of G6P was measured varying the substrate concentration from 2 mM to 15.6 µM and keeping $NADP^+$ at 1 mM. Likewise, K_m^{app} of $NADP^+$ was measured varying its concentration from 1 mM to 7.8 µM and keeping G6P at 2 mM. In both assays *Tc*ΔG6PDH concentration was 2 nM. Values of kinetic parameters for *Tc*G6PDH-L were obtained varying the G6P concentration between 2 mM to 15.6 µM and $NADP^+$ between 1 mM and 7.8 µM, while keeping the other substrate at saturating concentrations, 600 µM and 2 mM for $NADP^+$ and G6P, respectively. *Tc*G6PDH-L concentration was ke~~e~~pt at 2 nM.

2.5. Protein Crystallization and Data Collection

*Tc*ΔG6PDH crystallization screens used the sitting-drop vapor diffusion method. The protein samples had previously been filtered using a 0.1 µm Ultrafree-MC centrifugal filter (~~Milipore~~Millipore). Crystallization plates were prepared using a Phoenix/RE liquid handler (Art Robbins Instruments) and the JCSG-plus (Molecular Dimensions), Classics Suite (Qiagen), PEGs (Qiagen), and $AmSO_4$ (Qiagen) screens. The trial drops consisted of 100 nL of protein solution (10 $mg.mL^{-1}$ of His-tagged *Tc*ΔG6PDH, 2mM G6P, and 2 mM NADPH, 20 mM Tris-HCl pH 8.0, 0.2 M NaCl) and an equivalent volume of reservoir, equilibrated against 50 µL of reservoir at 20 ºC in a CrystalMation Plate Hotel system (Rigaku).

Reproducible hits were obtained with condition G1 of the JCSG-plus screen (30% jeffamine ED-2003, 0.1 M HEPES pH 7.0) and this was optimized using hanging-drop vapor diffusion ~~method~~ at 18 ºC. Suitable crystals of *Tc*ΔG6PDH were obtained when 2 µL of protein solution (10 $mg.mL^{-1}$ *Tc*ΔG6PDH, 5 mM G6P, 2 mM NADPH, 20 mM Tris-HCl pH 8.0, 0.2 M NaCl and 5 mM 2-mercaptoethanol) was mixed with 1 µL of reservoir solution (27-32 % jeffamine ED2003 and 0.1 M HEPES pH 6.8-7.2). Crystals were cryo-protected in 45% jeffamine ED2003 with 0.1 M HEPES pH 6.8, and then flash-cooled in liquid nitrogen before in-house screening using a Rigaku MicroMax-007 HF rotating-anode X-ray source equipped with a Saturn 944 HG+ CCD detector with sample maintained at about -170 ºC. Several crystals were sent for full data collection at the Diamond Light Source, beamline I04-1 and the best dataset, as judged by data processing statistics (see below) was identified.

2.6. X-ray Data Processing and *Tc*ΔG6PDH Structure Determination and Refinement

Data were indexed and integrated using XDS [26] and scaled using AIMLESS [27]. The structure was solved by molecular replacement with PHASER [28] using chain A of the PDB entry 4E9I as search model. The program COOT [29] was used for model manipulation and incorporation of solvent and ligands. REFMAC5 [30] was used to perform ~~restrained~~ refinements, using Translation/Libration/Screw [31] and non-crystallographic symmetry restraints. MOLPROBITY [32] was used to inspect model geometry in combination with the validation tools provided by COOT. Molecular figures were prepared in PyMOL v.1.8 (Schrödinger, LLC). Electrostatic potential surfaces were computed with PDB2PQR [33] and APBS [34]. Crystallographic statistics are presented in Table 1.

3. Results and Discussion

We have significantly advanced the characterization of *T. cruzi* G6PDH. The new structural data on this validated target has potential to inform the development of new drugs with therapeutic utility. The truncated construct *Tc*ΔG6PDH, lacking 57 N-term and 10 C-terminal residues, was designed to improve the crystallizability of this protein. This approach was chosen taking in account the success attained in the crystallization the *Tc*G6PDH-S, missing the first 37 residues [23]. His-tagged *Tc*ΔG6PDH was produced in *E. coli* and purified to homogeneity in two steps using affinity and size exclusion chromatography (Figures S1.A and S1.B, Suppl. Material). The size exclusion chromatography ~~result~~ data indicate~~s~~ that *Tc*ΔG6PDH forms a tetramer, in agreement ~~to~~ with previous ~~published data~~work [23]. When *Tc*G6PDH-L is expressed in *E. coli* ~~cells~~, almost all the protein ~~appeared in~~formed inclusion bodies [25]. ~~Nonetheless~~Nevertheless, the small amount of soluble *Tc*G6PDH-L recovered by IMAC was enough to perform ~~the intended~~ kinetic~~s studies~~analyses. ~~In measuring the *Tc*ΔG6PDH activity,~~ The NaCl and glycerol present in the reaction buffer ~~showed to be~~proved essential to keep the enzyme stable. The ~~C~~calculated K_m^{app} values of G6P and $NADP^+$ were 306.1 ± 20.3 and 80.1 ± 5.9 μM, respectively (Figure S2.A, Suppl. Material). The k_{cat} values obtained when varying the concentrations of G6P and $NADP^+$ were 53.6 ± 1.2 and 51.8 ± 1.1 s^{-1}, respectively. The K_m^{app} values obtained ~~to~~ for the *Tc*G6PDH-L, using the same assay condition~~s~~, were 210.3 ± 21.6 and 47.4 ± 4.3 μM for G6P and $NADP^+$, respectively (Figure S2.B, Suppl. Material). Additionally, k_{cat} values for the long form of the enzyme were 61.6 ± 1.9 for G6P and 52.8 ± 1.3 s^{-1} for $NADP^+$. Thus, ~~despite a reduced stability,~~ the truncated enzyme remains ~~functional~~ active and has an affinity for both substrate and cofactor that are only slightly higher (around 1.5 times) than those of the *Tc*G6PDH-L.

Commented [WH1]: There are no data relating to "stability". The lack of soluble material might be simply the folding process whihc is distinct from whether the protein is stabile or not.

3.1. *Tc*ΔG6PDH Overall Structure

Bipyramidal crystals of *Tc*ΔG6PDH obtained in the presence of both G6P and NADPH attained a maximum dimension of 0.3 mm within 15 days and diffracted to a resolution limit of 2.65 Å using synchrotron radiation (Figure S3, Suppl. Material). Molecular replacement calculations positioned three polypeptide chains, labelled A, B and C, in the asymmetric unit. In the refined structure (PDB ID 5AQ1), each polypeptide chain comprises residues P62 to T545, plus one molecule of G6P and NADPH. There are no outliers in the Ramachandran plot, which returned 98% and 2% of the residues in favoured and allowed regions, respectively. Two solvent exposed loops ~~found in the final model,~~ spanning ~~from~~ residues R129 to H136 and A290 to Y295, ~~showed to be~~are poorly ordered. Those loops do not participate in crystal contacts neither in ligand~~s~~ binding. The superposition of chains A, B and C did not reveal significant differences in the C^{α} trace (RMSD and

number of atoms used in the alignment: AB, 0.10 Å and 434 C^{α} atoms; AC, 0.23 Å and 445 C^{α} atoms; BC, 0.17 Å and 448 C^{α} atoms). ~~and no outliers were observed in Ramachandran analysis, which returned 98% and 2% of the residues in favoured and allowed regions, respectively.~~ The topology of ~~the~~ *Tc*ΔG6PDH is very similar to the human and bacterial homologous enzymes [17,21], being composed ~~by~~ of an NAD(P)-binding Rossmann-like domain (residues P62 to I245) and a β + α domain (residues D246 to T545) (Figure 1A). The β + α domain has a large β-sheet with nine antiparallel strands that in *Hs*G6PDH binds ~~the~~ a structural $NADP^+$ ~~in the *Hs*G6PDH~~. ~~In the present *Tc*ΔG6PDH structure t~~There is no electron density to support the presence of a structural $NADP^+$ in *Tc*ΔG6PDH. ~~Although t~~There are three polypeptide chains in the AU of *Tc*ΔG6PDH crystals and ~~,~~ the ~~expected~~ tetrameric biological unit ~~can~~ is ~~be~~ reconstructed by crystallographic symmetry operations. The *Tc*ΔG6PDH tetramer ~~can be~~is formed by the packing of two B:C dimers ~~or~~ and by four A chains (Figure 1B). No significant differences were observed between ~~those~~ these tetramers. In the tetrameric assembly all inter-chain contacts are established by residues from the β + α domains burying about 12 % of each subunit surface area (equivalent to an area of approximately 3250 $Å^2$).

3.2. G6P binding site

G6PDH catalyses the formation of a double bond between C1 and O1 of G6P to produce 6-phosphoglucono-δ-lactone with concomitant reduction of $NADP^+$, by the transfer of a hydride ion to C4 of the nicotinamide (Scheme 1). In the *Tc*ΔG6PDH structure, the β-anomer of G6P is found in a chair $^4C_1$ conformation establishing H-bonds with side-chains of K217, D246, H247, Y248, K251, E285, D304, H309, K403, R408 and Q437 (Figure 2A). These residues are highly conserved in the human and *L. mesenteroides* enzymes [19,22]. The conformation of the G6P observed in the *Tc*ΔG6PDH orients ~~its~~ the C1 hydroxyl group towards the H309 $N^{\varepsilon 2}$, and in turn the H309 $N^{\varepsilon 1}$ H-bonds to D246 $O^{\delta 1}$. These interactions are in agreement with the proposed mechanism of reaction for G6PDH [18,19], where a catalytic dyad, represented by H309 and D246 in ~~the~~ *Tc*ΔG6PDH, is responsible for proton abstraction of the substrate.

A comparison between the ternary complex of *Tc*ΔG6PDH-G6P-NADPH and the binary complex of *Tc*G6PDH-G6P (PDB ID 4EM5; Buschiazzo, A., Botti, H., Ortiz, C., Comini, M.A. unpublished) reveals a striking and unexpected difference in the conformation of G6P. In the binary complex there are four polypeptide chains in the AU and in three of them, the pyranose ring of G6P is inverted, with the C4 hydroxyl group participating in an H-bond to H309 $N^{\varepsilon 2}$ (Figure 2B). This ~~mis~~position of~~ed~~ G6P, which is not compatible with the proposed mechanism, is observed in chains A, B and C. In ~~, but not in~~ chain D, ~~which presents~~ the G6P ~~in~~ is in the expected orientation with the

C1 hydroxyl group pointing to ~~the~~ H309. The low resolution of the binary complex, 3.35 Å, ~~might have led to a biased interpretation of the electron density and misorientation of the substrate~~may have complicated the analysis. In this ~~way~~respect, we were fortunate to obtain much higher resolution diffraction data and therefore provide a more reliable model.~~our results represent a correction to the binary complex of the G6PDH from *T. cruzi*.~~

Commented [WH2]: We need to be circumspect here. We are not comparing like with like! Binary vs ternary – so I've tried to keep this very simple.

3.3. Catalytic NADP(H) binding site

At the catalytic site, NADPH binds to the Rossmann-like domain forming H-bonds with S77, D79, L80, R109, S110, Y151, L186, and K217; in addition to forming a cation-π interaction between R109 and the adenine moiety (Figure 3A). The nicotinamide moiety is oriented in the *syn* conformation and with C4 oriented towards G6P C1at a distance of 3.9 Å, compatible with hydride transfer (Figure S4, Suppl. Material). All of the residues involved in binding the catalytic NADPH are conserved in *Tc*ΔG6PDH and *Hs*G6PDH structures, with one notable exception. In the human enzyme Y147 O^{η} donates its hydrogen to E170 carbonyl group forming an H-bond (Figure 3B). In *Tc*ΔG6PDH, Y147 is replaced by F191 with the χ_1 angle rotated by almost 120°. This conformational difference, perhaps allowed by the loss of a restraining H-bond, leads to the formation of a cavity just below the nicotinamide riboside in *Tc*ΔG6PDH (Figures 3C). This cavity, with volume of about 220 $Å^3$ calculated using the software KVFinder [35], would be large enough to accommodate cyclic substituents linked to a nicotinamide riboside (Figure S5, Suppl. Material). It is noteworthy that this cavity is absent in *Hs*G6PDH (Figure 3D), and has not been found in other human proteins that, in accordance to CATH [36], also ~~have~~ possess an NAD(P)-binding Rossmann-like domain (Table S3, Suppl. Material). The cavity in the *T. cruzi* enzyme offers opportunities for a structure-based approach to develop novel G6PDH inhibitors and since this structural feature represents a difference with respect to the human enzyme then also might assist the discovery of inhibitors selective for the trypanosomal G6PDH over the human enzyme.

3.4. Structural $NADP^+$ binding site

A structural $NADP^+$ site, located on the β + α domain, has been reported for *Hs*G6PDH (PDBs entries 1QKI and 2BH9) and postulated as important for protein stability [21,22]. In ~~the~~ *Hs*G6PDH, the structural $NADP^+$ participates in H-bonds to K238, K366, R370, R393, Y401, K403, D421, T423, and R487 (Figure 4A). Additionally, W509 and Y503 establish π-stacking interactions with the nicotinamide and adenine rings, respectively. The comparison of the structural $NADP^+$ binding site between the *Hs*- and *Tc*ΔG6PDH shows that K366, R487, and Y503 (in

*Hs*G6PDH) are replaced by L409, C528, and T544 (in *Tc*ΔG6PDH), respectively (Figure 4B). L409 and C528 in the *Tc*ΔG6PDH would not be able to H-bond the 2'-phosphate of an NADP(H) ~~possible~~ that might occup~~yying~~ this site. Similarly, ~~the~~ T544 would not be able to make a π-stacking interaction with the adenine moiety. In addition to these natural occurring differences, the *Tc*ΔG6PDH C-terminus is artificially truncated. In ~~the~~ *Hs*G6PDH, the C-terminus caps the structural NADP(H) providing an aromatic residue (W509) that stacks ~~in~~ on the adenine moiety. These ~~changes~~ differences might compromises the binding of a structural NADP(H) to *Tc*ΔG6PDH and so explain why it was not observed in the structure. Interestingly, in the kinetic assays performed with *Tc*ΔG6PDH an active enzyme requires the presence of NaCl and glycerol, additives ~~that are~~ known to improve protein~~s~~ stability.

The *Tc*ΔG6PDH structure is the first enzyme-substrate-cofactor complex of a eukaryotic G6PDH. Comparisons with other G6PDH structures resulted in several observations. Firstly, interactions made by G6P in *Tc*ΔG6PDH are in agreement with the proposed mechanism of reaction [18,19] and the substrate conformation is similar to those observed in complexes of the bacterial and human enzymes. In the binary *Tc*G6PDH-G6P complex (PDB ID 4EM5), three (out of the four) chains of the asymmetric unit have the~~, the substrate is likely to be misoriented, with the~~ pyranose ring of the substrate in an ~~flipped~~ orientation incompatible with a mechanism that involves proton abstraction from the C1-hydroxyl group ~~by~~ of the catalytic H309. Secondly, in the cofactor~~-~~binding site, the presence of, and conformation adopted by a phenylalanine ~~residue~~ in the *Tc*ΔG6PDH (F191) instead of a tyrosine ~~residue~~ as in *Hs*G6PDH (Y147), results in the formation of a cavity unique to the parasite enzyme. We believe this cavity might be further explored for the development of NADP(H) competitive inhibitors with selectivity against *Tc*G6PDH enzyme.

Acknowledgments

We thank Paul Fyfe for support and advice with data collection, Diamond for provision of beamtime and staff at the I04 MX beamline in particular, FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) for research grant 2013/03983-5 (ATC) and BEPE fellowship 2014/07533-7 (GFM), and the Wellcome Trust for grant 094090 (WNH).

Authors Contributions

GFM conduced all experiments. AD assisted GFM with diffraction data collection, structure solution and refinement. WNH supervised GFM during his internship in Dundee. ATC was the project coordinator and supervisor of GFM Brazil. All authors reviewed and contributed to the

manuscript.

References

1 Pandolfi PP, Sonati F, Rivi R, Mason P, Grosveld F & Luzzatto L (1995) Targeted disruption of the housekeeping gene encoding glucose 6-phosphate dehydrogenase (G6PD): G6PD is dispensable for pentose synthesis but essential for defense against oxidative stress. *EMBO J.* **14**, 5209–15.

2 Ho H, Cheng M & Chiu DT (2007) Glucose-6-phosphate dehydrogenase--from oxidative stress to cellular functions and degenerative diseases. *Redox Rep.* **12**, 109–18.

3 Wood T (1986) Physiological functions of the pentose phosphate pathway. *Cell Biochem. Funct.* **4**, 241–247.

4 Stanton RC (2012) Glucose-6-phosphate dehydrogenase, NADPH, and cell survival. *IUBMB Life* **64**, 362–9.

5 Zhang C, Zhang Z, Zhu Y & Qin S (2014) Glucose-6-phosphate dehydrogenase: a biomarker and potential therapeutic target for cancer. *Anticancer. Agents Med. Chem.* **14**, 280–289.

6 Park J, Rho HK, Kim KH, Choe SS, Lee YS & Kim JB (2005) Overexpression of glucose-6-phosphate dehydrogenase is associated with lipid dysregulation and insulin resistance in obesity. *Mol. Cell. Biol.* **25**, 5146–57.

7 Park J, Choe SS, Choi a H, Kim KH, Yoon MJ, Suganami T, Ogawa Y & Kim JB (2006) Increase in glucose-6-phosphate dehydrogenase in adipocytes stimulates oxidative stress and inflammatory signals. *Diabetes* **55**, 2939–49.

8 Gupte SA (2008) Glucose-6-phosphate dehydrogenase: a novel therapeutic target in cardiovascular diseases. *Curr Opin Investig Drugs* **9**, 993–1000.

9 Cappellini MD & Fiorelli G (2008) Glucose-6-phosphate dehydrogenase deficiency. *Lancet* **371**, 64–74.

10 Gupta S, Igoillo-Esteve M, Michels PA & Cordeiro AT (2011) Glucose-6-Phosphate Dehydrogenase of Trypanosomatids: Characterization , Target Validation , and Drug Discovery. *Mol. Biol. Int.* **2011**, 1 – 10.

11 Cordeiro AT, Thiemann OH & Michels PA (2009) Inhibition of Trypanosoma brucei glucose-6-phosphate dehydrogenase by human steroids and their effects on the viability of cultured parasites. *Bioorg. Med. Chem.* **17**, 2483–9.

12 Gupta S, Cordeiro AT & Michels PA (2011) Glucose-6-phosphate dehydrogenase is the target for the trypanocidal action of human steroids. *Mol. Biochem. Parasitol.* **176**, 112–5.

13 Cordeiro AT & Thiemann OH (2010) 16-bromoepiandrosterone, an activator of the mammalian immune system, inhibits glucose 6-phosphate dehydrogenase from Trypanosoma cruzi and is toxic to these parasites grown in culture. *Bioorg. Med. Chem.* **18**, 4762–8.

14 Mercaldi GF, Ranzani AT & Cordeiro AT (2014) Discovery of new uncompetitive inhibitors of glucose-6-phosphate dehydrogenase. *J. Biomol. Screen.* **19**, 1362–1371.

15 Hamilton NM, Dawson M, Fairweather EE, Hamilton NS, Hitchin JR, James DI, Jones SD, Jordan AM, Lyons AJ, Small HF, Thomson GJ, Waddell ID & Ogilvie DJ (2012) Novel steroid inhibitors of glucose 6-phosphate dehydrogenase. *J. Med. Chem.* **55**, 4431–45.

16 Preuss J, Richardson AD, Pinkerton A, Hedrick M, Sergienko E, Rahlfs S, Becker K & Bode L (2013) Identification and characterization of novel human glucose-6-phosphate dehydrogenase inhibitors. *J. Biomol. Screen.* **18**, 286–97.

17 Rowland P, Basak a K, Gover S, Levy HR & Adams MJ (1994) The three-dimensional structure of glucose 6-phosphate dehydrogenase from Leuconostoc mesenteroides refined at 2.0 A resolution. *Structure* **2**, 1073–87.

18 Cosgrove MS, Naylor C, Paludan S, Adams MJ & Levy HR (1998) On the mechanism of the reaction catalyzed by glucose 6-phosphate dehydrogenase. *Biochemistry* **37**, 2759–2767.

19 Cosgrove MS, Gover S, Naylor CE, Vandeputte-Rutten L, Adams MJ & Levy HR (2000) An examination of the role of Asp-177 in the His-Asp catalytic dyad of Leuconostoc mesenteroids Glucose 6-phosphate Dehydrogenase: X-ray structure and pH dependence of kinetic parameters of D177N mutant enzyme. *Biochemistry* **39**, 15002–11.

20 Naylor CE, Gover S, Basak AK, Cosgrove MS, Levy HR & Adams MJ (2001) NADP + and NAD + binding to the dual coenzyme specific enzyme Leuconostoc mesenteroides glucose 6-phosphate dehydrogenase: different interdomain hinge angles are seen in different binary and ternary complexes. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **57**, 635–648.

21 Au SW, Gover S, Lam VM & Adams MJ (2000) Human glucose-6-phosphate dehydrogenase: the crystal structure reveals a structural NADP(+) molecule and provides insights into enzyme deficiency. *Structure* **8**, 293–303.

22 Kotaka M, Gover S, Vandeputte-Rutten L, Au SWN, Lam VMS & Adams MJ (2005) Structural studies of glucose-6-phosphate and NADP+ binding to human glucose-6-phosphate dehydrogenase. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **61**, 495–504.

23 Ortíz C, Larrieux N, Medeiros A, Botti H, Comini M & Buschiazzo A (2011) Expression, crystallization and preliminary X-ray crystallographic analysis of glucose-6-phosphate dehydrogenase from the human pathogen Trypanosoma cruzi in complex with substrate. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **67**, 1457–61.

24 Baugh L, Phan I, Begley DW, Clifton MC, Armour B, Dranow DM, Taylor BM, Muruthi MM, Abendroth J, Fairman JW, Fox III D, Dieterich SH, Staker BL, Gardberg AS, Choi R, Hewitt SN, Napuli AJ, Myers J, Barrett LK, Zhang Y, Ferrell M, Mundt E, Thompkins K, Tran N, Lyons-Abbott S, Abramov A, Sekar A, Serbzhinskiy D, Lorimer D, Buchko GW, Stacy R, Stewart LJ, Edwards TE, Van Voorhis WC & Myler PJ (2015) Increasing the structural coverage of tuberculosis drug targets. *Tuberculosis* **95**, 142–148.

25 Igoillo-Esteve M & Cazzulo JJ (2006) The glucose-6-phosphate dehydrogenase from Trypanosoma cruzi: Its role in the defense of the parasite against oxidative stress. *Mol. Biochem. Parasitol.* **149**, 170-81.

26 Kabsch W (2010) XDS. *Acta Crystallogr. Sect. D* **66**, 125–132.

27 Evans PR & Murshudov GN (2013) How good are my data and what is the resolution? *Acta*

Crystallogr. D. Biol. Crystallogr. **69**, 1204–14.

28 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674.

29 Emsley P, Lohkamp B, Scott WG & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501.

30 Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner R a., Nicholls RA, Winn MD, Long F & Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367.

31 Painter J, Merrit EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D. Biol. Crystallogr.* **62**, 439-50.

32 Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS & Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 12–21.

33 Dolinsky TJ, Nielsen JE, McCammon JA & Baker NA (2004) PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665–W667.

34 Baker NA, Sept D, Joseph S, Holst MJ & McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10037–41.

35 Oliveira SH, Ferraz FA, Honorato RV, Xavier-Neto J, Sobreira TJ, de Oliveira PS (2014) KVFinder: steered identification of protein cavities as a PyMOL plugin. *BMC Bioinformatics*. **15**. doi: 10.1186/1471-2105-15-197.

36 Sillitoe I, Lewis TE, Cuff A, Das S, Ashford P, Dawson NL, Furnham N, Laskowski RA, Lee D, Lees JG, Lehtinen S, Studer RA, Thornton J, Orengo CA (2015) CATH: comprehensive structural and functional annotations for genome sequences. *Nucl. Acids Res.* **43**, D376-D381.

Figure Legends

Figure 1: Carton representation of *Tc*ΔG6PDH. **A)** Structure of a subunit, showing the N-term NAD(P)-binding Rossmann-like domain (orange), C-term β + α domain (gray), NADPH (magenta spheres) and G6P (green spheres). Site for structural NADP is highlighted (red ellipse). **B)** *Tc*ΔG6PDH biological unit: interface residues involved in the tetrameric assembly are exclusively located in the β + α domains. Two unique tetramers could be generated from *Tc*ΔG6PDH subunits, one using chain A and its symmetry related neighbours (A', A" and A'") and the other using the dimer of chains B and C and a symmetry related dimer (B' and C').

Figure 2: Different orientations of G6P (green sticks) observed in the ternary complex (**A**) obtained in the present work (PDB ID 5AQ1) and in the binary complex (**B**) (PDB ID 4EM5). H309 and D246 (both yellow sticks) form the catalytic dyad responsible for proton abstraction from the C1-hydroxyl of the G6P pyranose ring. Residues in gray and yellow belong to the β + α domain, and K217 in orange is from the NAD(P)-binding Rossmann-like domain. *Fo-Fc* omit map for G6P contoured at 3.8 σ. Dashed lines represent H-bonds and distances observed in the *Tc*ΔG6PDH (present work) are reported in Table S1, Suppl. Material. Carbon atoms of the glucopyranose are numbered. Figure was prepared using chain A of both structures.

Figure 3: Comparison between the catalytic NADP(H) binding site of *Tc*ΔG6PDH (chain A; PDB ID 5AQ1) and *Hs*G6PDH (chain A; PDB ID 2BH9) [22]. **A)** *Tc*ΔG6PDH residues (orange) making H-bonds (dashed lines) to the catalytic NADPH (magenta sticks). NADPH *Fo-Fc* omit map contoured at 2.5 σ. H-bond distances are reported in Table S2, Suppl. Material. **B)** *Hs*G6PDH residues (cyan) making H-bonds (dashed lines) to the catalytic NADPH (yellow sticks). **C)** Electrostatic surface of *Tc*ΔG6PDH showing a cavity below the nicotinamide riboside. **D)** Electrostatic surface of the *Hs*G6PDH displayed in the vicinity of $NADP^+$.

Figure 4: Structural $NADP^+$ site from *Hs*G6PDH (**A**) (PDB ID 2BH9) and comparison with *TcΔG6PDH* (**B**). Structural $NADP^+$ (yellow sticks) binds to residues of the β + α domain from the human enzyme through H-bonds (dashed lines) or π-stacking interactions. Residues K366, R487 and Y503 in *Hs*G6PDH correspond to L409, C528 and T544 in *Tc*ΔG6PDH (labelled in red), respectively, and may compromises the binding of the structural $NADP^+$. Figure prepared using chain A of both the human and *T. cruzi* enzyme structures.

Figure 1

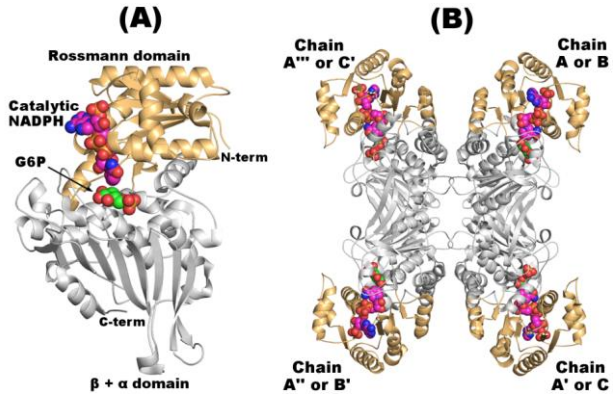

(A)
Rossmann domain
Catalytic
NADPH
G6P
N-term
C-term
β + α domain
(B)
Chain
A''' or C'
Chain
A or B
Chain
A'' or B'
Chain
A' or C

Figure 2

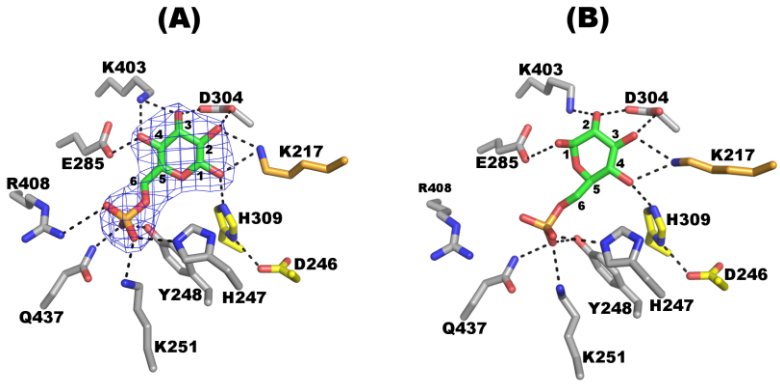

(A)
K403
D304
E285
K217
R408
H309
D246
Y248
H247
Q437
K251
(B)
K403
D304
E285
K217
R408
H309
D246
Y248
H247
Q437
K251

Figure 3

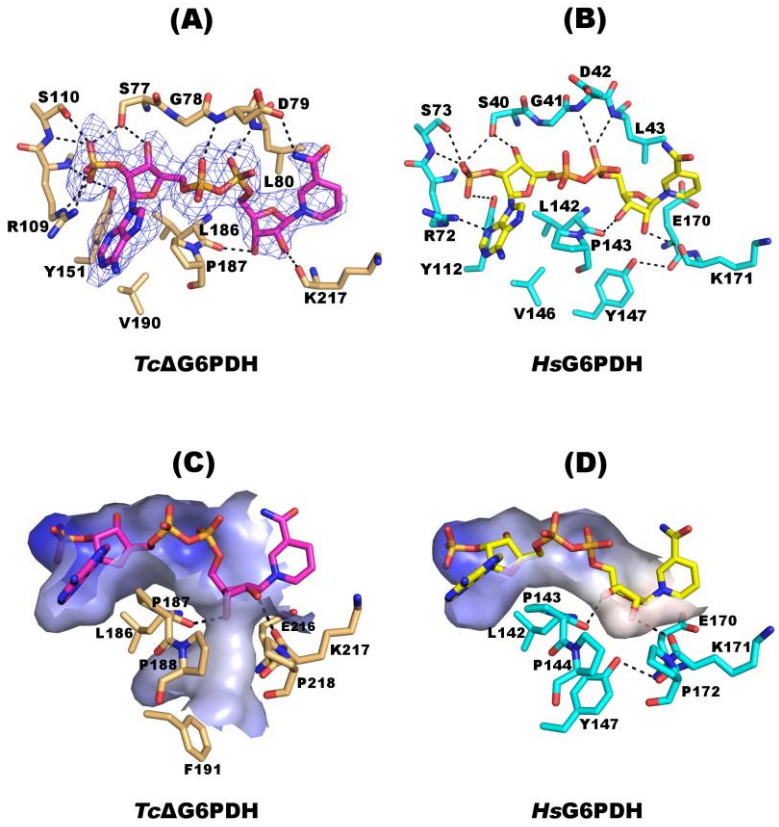
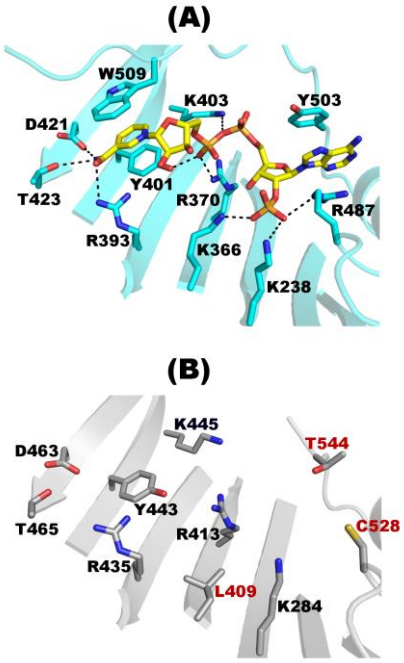
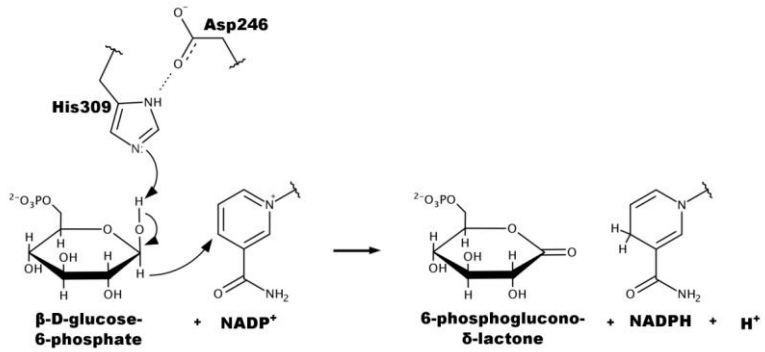

(A)
(B)
S110
S77
G78
D79
L80
R109
L186
Y151
P187
K217
V190
D42
S73
S40
G41
L43
L142
R72
E170
P143
Y112
K171
V146
Y147
TcΔG6PDH
HsG6PDH
(C)
(D)
P187
L186
E216
K217
P188
P218
F191
P143
L142
E170
K171
P144
P172
Y147
TcΔG6PDH
HsG6PDH

Figure 4


(A)
W509
K403
Y503
D421
Y401
T423
R370
R487
R393
K366
K238
(B)
K445
D463
T544
Y443
T465
R413
C528
R435
L409
K284

Scheme 1 – Reaction catalysed by G6PDH.*



* – H309 $N^{\delta 1}$ makes a hydrogen bond with D246 $O^{\delta 1}$, forming the catalytic dyad of the G6PDH. H309 $N^{\delta 2}$ is the general base that abstracts ~~the~~ a proton from the C1-OH of G6P, inducing the transfer of C1-hydride to C4 of the nicotinamide moiety of $NADP^+$. 6-phosphoglucono-δ-lactone and NADPH are the reaction products. Only the nicotinamide moiety of NADP(H) is represented.

Table 1 – Crystallographic statistics for *Tc*ΔG6PDH-G6P-NADPH ternary complex.

Data Collection	
Resolution Range (Å)	49.13-2.65 (2.72-2.65)
Space group	$I4_122$
Unit cell axes and angles (Å, °)	a = b = 154.95, c = 348.36 α = β = γ = 90
Wavelength (Å)	0.91741
No. of Reflections	508422 (39534)
No. of Unique Reflections	58510 (4584)
R_{merge}† (%)	12.9 (83.5)
Completeness (%)	95.4 (96.9)
$\langle I/\sigma(I)\rangle$	11.1 (2.5)
Multiplicity	8.7 (8.6)
Solvent Content (%)	60
Subunits per asymmetric unit	3
Wilson *B* factor (A^2)	39.6
Refinement	
R_{work}§ (%)	19.99
R_{free}¶ (%)	22.55
R.m.s.d., bonds (Å)	0.0048
R.m.s.d., angles (°)	1.0476
Total protein residues	1458
Total protein atoms	11571
No. of solvent atoms	253
Average *B* factors ($Å^2$)	
Protein – Chains A / B / C	43.0 / 46.9 / 54.2
BG6 – Chains A / B / C	29.2 / 32.7 / 37.1
NDP – Chains A / B / C	51.8 / 53.5 / 67.4
Waters	34.1

[a] - Values in parentheses refer to the highest resolution bin of 2.72-2.65 Å.

† $R_{merge} = \Sigma_{hkl} \Sigma_i | I_i(hkl) - \langle I(hkl)\rangle | / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the *i*th measurement of reflection *hkl* and $\langle I(hkl)\rangle$ is the mean value of $I_i(hkl)$ for all *i* measurements.

§ $R_{work} = \Sigma_{hkl} | |F_{obs}| - |F_{calc}| | / \Sigma_{hkl} |F_{obs}|$, where F_{obs} is the observed structure-factor amplitude and the F_{calc} is the structure-factor amplitude calculated from the model.

¶ R_{free} is the same as R_{work} except calculated with a subset (5%) of data that were excluded from refinement calculations.