Supporting information

Supplementary Text

T1: Supplementary materials and methods

Expression TcPYK

cDNA corresponding to the *Trypanosoma cruzi* PYK (*Tc*PYK) gene was generated by Genscript DNA synthesis service. The *Tc*PYK codons were optimised for expression in *Escherichia coli,* which increased expression and final yields of soluble protein. An *NdeI* restriction site was added to the 5' end of the gene, which also incorporated the ATG start (initiator) codon. An *XhoI* restriction site was added downstream of the stop codon. The gene was cloned by GENSCRIPT into a pUC57 vector. Inserts were digested from the pUC57 using *NdeI* and *XhoI* and then purified using the Qiagen gel purification kit. The digested inserts were ligated into either *NdeI-XhoI*-digested pET30a vector (untagged construct) or *NdeI-XhoI*-digested pET28a [N-terminal hexahistidine (His_6) tagged construct] and verified by DNA sequencing.

Chemically competent BL21(DE3) *E. coli* cells (Novagen) were transformed with pET28a_*Tc*PYK or pET30a_*Tc*PYK. Single colonies of transformed *E. coli* harboring the desired *Tc*PYK gene construct were picked from Lysogeny-Broth (LB) kanamycin plates (50 µg/ml) and used to inoculate 50 ml of 2xTY medium (containing 50 µg/ml of kanamycin). Cultures were grown overnight at 37^{o}C with agitation. 20 ml of the overnight culture was used to inoculate one-liter of 2xTY medium containing kanamycin (50 µg/ml). One-liter cultures were grown to an OD_{600} of 0.8-1.0, and expression was induced at 18°C by adding IPTG (isopropyl-β-D-thiogalactopyranoside) to a final concentration of 1 mM. The incubation was then continued for 18 h and the cells were harvested at 8,000 rpm in a JLA-9.1000 rotor for 12 min at 10°C. Cell pellets from 1 liter cultures were then frozen in liquid nitrogen and stored at -80°C. All purification steps were performed at 6°C unless stated otherwise.

Purification of His_6-tagged and untagged TcPYKs

His_6*Tc*PYK - cell pellets were resuspended in lysis buffer [50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0, 10% glycerol (GOL) and 1 tablet EDTA-free protease inhibitors (Roche)] to a final volume of 30 ml/5 g (~1 liter of cell culture) of cells and lysed using a constant cell disruption system (pressure set to 25 MPa). The lysate was centrifuged at 23000 rpm at 10°C for 45 min and the supernatant was filtered through a 0.2 µm syringe filter. The supernatant was loaded onto a 5 ml IMAC Hitrap HP Sepharose column (precharged with cobalt) at 2 ml min^{-1}. The column was maintained at a constant flow rate of 2 ml min^{-1} throughout the purification process. The column was washed with twenty column volumes (CV) of buffer A (50 mM NaH_2PO_4, 300 mM NaCl, 10% GOL and 20 mM imidazole, pH 8.0). An additional wash step of 20% buffer B (50 mM NaH_2PO_4, 300 mM NaCl, 10% GOL and 250 mM imidazole, pH 8.0) was also included, which minimised contaminating proteins. The His_6*Tc*PYK protein was eluted over a ten CV step gradient using 100% buffer B. Eluted fractions were pooled and concentrated using a Vivaspin column (molecular mass cut off = 100 kDa). 5-8 ml samples were loaded onto a Hiprep 26/10 desalting column, pre-equilibrated in buffer C [20 mM triethanolamine-HCl (TEA) buffer (pH 7.2) and 20% GOL]. Samples containing His_6*Tc*PYK were pooled, concentrated (10-20 mg/ml) and loaded onto a Superdex 200 10/300 GL column, pre-equilibrated in buffer C and eluted with the same buffer.

Untagged *Tc*PYK - Cell pellets were resuspended in buffer A (50mM TEA, 20mM KCl, 20% glycerol, pH7.2 and 1 tablet EDTA-free protease inhibitors (Roche)) to a final volume of 30 ml/5 g (~1 liter of cell culture) of cells and lysed using a constant cell disruption system (pressure set to 25 MPa). The lysate was centrifuged at 23000 rpm at 10°C for 45 minutes and the supernatant was filtered through a 0.2 µm syringe filter. The supernatant was applied (2 ml min^{-1}) to two ion exchange columns (Hiprep DEAE FF 16/10 and Hiprep SP FF 16/10 - fitted sequentially) pre-equilibrated in buffer A. The experiment was maintained at a constant flow rate of 2 ml min^{-1} throughout the purification process. The flow-through was fractionated and collected over four CVs (80ml). *Tc*PYK was found in the flow-through (the pH is similar to the theoretical pI (7.13) of *Tc*PYK resulting in a minimal net charge) while many of the

contaminating *E. coli* proteins are charged at pH 7.2 and therefore bind the ion exchange columns. Samples were analysed by SDS-PAGE and fractions containing *Tc*PYK were pooled and concentrated using a Vivaspin column (MW cutoff =100kDa) to a final volume of 10 ml (~10mg/ml). Samples were loaded onto a Superdex 200 prepgrade 26/60 XK column, pre-equilibrated in buffer C.

Eluted fractions were analysed by SDS-PAGE, and pure *Tc*PYK samples (≥95%) were pooled, concentrated to ~30 mg/ml, screened for activity, divided into small-volume samples, and stored at -20^{o}C. *Tc*PYK concentration was determined using the theoretical extinction coefficient, ε280nm = 19200 $M^{-1}.cm^{-1}$.

T2: ***Crystallisation of sulfonic acid complexes***

*Lm*PYK-PTS/suramin data were indexed in the orthorhombic space group I222, and contained two protein chains per asymmetric unit with a crystallographic 2-fold axis generating the biologically relevant tetramer (**Figure 3b**). Clearly interpretable electron density was observed for only one of the two B-domains, which is not uncommon in PYK structures (1,2). The missing B-domain of Chain B was modelled in a similar position to the B-domain of Chain A. Both chains were very similar, with a root mean square deviation of 0.28 Å for C^{α} atoms of all residues 1-498.

Crystals of *Lm*PYK were grown in the presence of 250 µM acid blue 80 [*Lm*PYK-AB80 – **Figure 1e(i)**]. The structure was determined at 2.85 Å resolution (see **Table 1** for data collection and refinement statistics) and contained two protein chains per asymmetric unit with a crystallographic 2-fold axis generating the biologically relevant tetramer. Clearly interpretable electron density was observed for only one of the two B-domains (the missing B-domain was not included in the final refined model).

Crystals of *Lm*PYK were also grown in the presence of 0.5 mM benzothiazole-2,5-disulfonic acid (*Lm*PYK-BDS). The refined 2.45 Å resolution (see **Table 1** for data collection and refinement statistics) structure contained two protein chains per asymmetric unit with a crystallographic 2-fold axis generating the biologically relevant tetramer. Clearly

interpretable electron density was observed for only one of the two B-domains. For both monomers BDS was observed clearly bound to the active sites in identical positions.

Crystals of *Lm*PYK were also grown in the presence of 0.5 mM ponceau S and 0.25 mM acid blue 25 (*Lm*PYK-ponceau S). The refined 2.7 Å resolution (see **Table 1** for data collection and refinement statistics) structure contained four protein chains per asymmetric unit. The ponceau S and acid blue 25 molecules form a stacking interaction (**Figure S4a**), with one ponceau S molecule binding to the active site in a position similar to that of the other sulfonic acid containing molecule described here.

Crystals of *Tc*PYK were also grown in the presence of 0.5 mM ponceau S [*Tc*PYK-AB80 – **Figure 1d(i)**]. The structure was determined at 2.1 Å resolution (see **Table 1** for data collection and refinement statistics) and contained two protein chains per asymmetric unit with a crystallographic 2-fold axis generating the biologically relevant tetramer (**Figure S4c**). Clearly interpretable electron density was observed for only one of the two B-domains (the missing B-domain was not included in the final refined model).

Refinement strategies for LmPYK and TcPYK structures

***Lm*PYK-PTS/suramin complex.** Water molecules were added to the model using COOT, and after several rounds of restrained refinement (using loose geometrical restraints) the final model yielded R/R_{free} values of 21.4 and 26.0% respectively. The B-domain of Chain B, for which no clear electron density was observed, was modelled in an identical position to that of the well defined B-domain of Chain A, and after a final round of refinement yielded R/R_{free} values of 20.9 and 25.9%. A further round of Translation/Libration/Screw (TLS) restrained refinement to account for vibrational motion was performed (five optimal TLS groups were determined using the TLSMD procedure (3)), and yielded final R/R_{free} values of 19.3/23.1%. TLS groups were as follows: (1) residues 1-86; (2) residues 87-187; (3) residues 188-298; (4) 299-424; (5) 425-498. All atoms were modeled with full occupancy and with individual atomic B-factors.

***Lm*PYK-AB80 complex.** The B-domain for which there was clear electron density had to be built manually using COOT, and build progression was monitored by decreases in R/R_{free} values. Metal ions and AB80 were added to the model, for which there was clear $F_o–F_c$ electron density. Only one active site had unambiguous $F_o–F_c$ electron density for AB80 (**see Figure S4 for details of asymmetric unit**). The model was then subjected to several rounds of restrained refinement, and after manual adjustments and the addition of water molecules, the final R/R_{free} values were 23.5 and 26.7%.

***Lm*PYK-BDS complex**. The model was refined in a manner similar to that described for the *Lm*PYK-AB80 complex, although both active sites had unambiguous $F_o–F_c$ electron density for BDS.

***Lm*PYK-Ponceau S complex.** The model was refined in a manner similar to that described for the *Lm*PYK-AB80 complex. Ponceau S and acid blue 25 (**see Figure S4 for details of asymmetric unit**) molecules were added, and clear $F_o–F_c$ electron density was observed.

***Tc*PYK-PonceauS complex.** The model was refined in a manner similar to that described for the *Lm*PYK-AB80 complex. Ponceau S molecules (**see Figure S4 for details of asymmetric unit**) were added to the model using $F_o–F_c$ electron density as a guide.

Supplementary Tables

Table S1. Suramin-complex structures

PDB ID	Protein	Res. Å	REF
3GAN	Uncharacterized protein At3g22680	2.00	Unpublished
3BF6	Thrombin	2.50	(4)
2H9T	Thrombin	2.40	(4)
3BJW	Phospholipase A2	2.30	(5)
1Y4L	Phospholipase A2 homolog 2	1.70	(6)
2NYR	NAD-dependent deacetylase sirtuin-5	2.06	(7)
1Y8E	Complement control protein	2.20	(8)

Taken from the PDB database on the 22th July 2010

Supplementary Figures

Figure S1

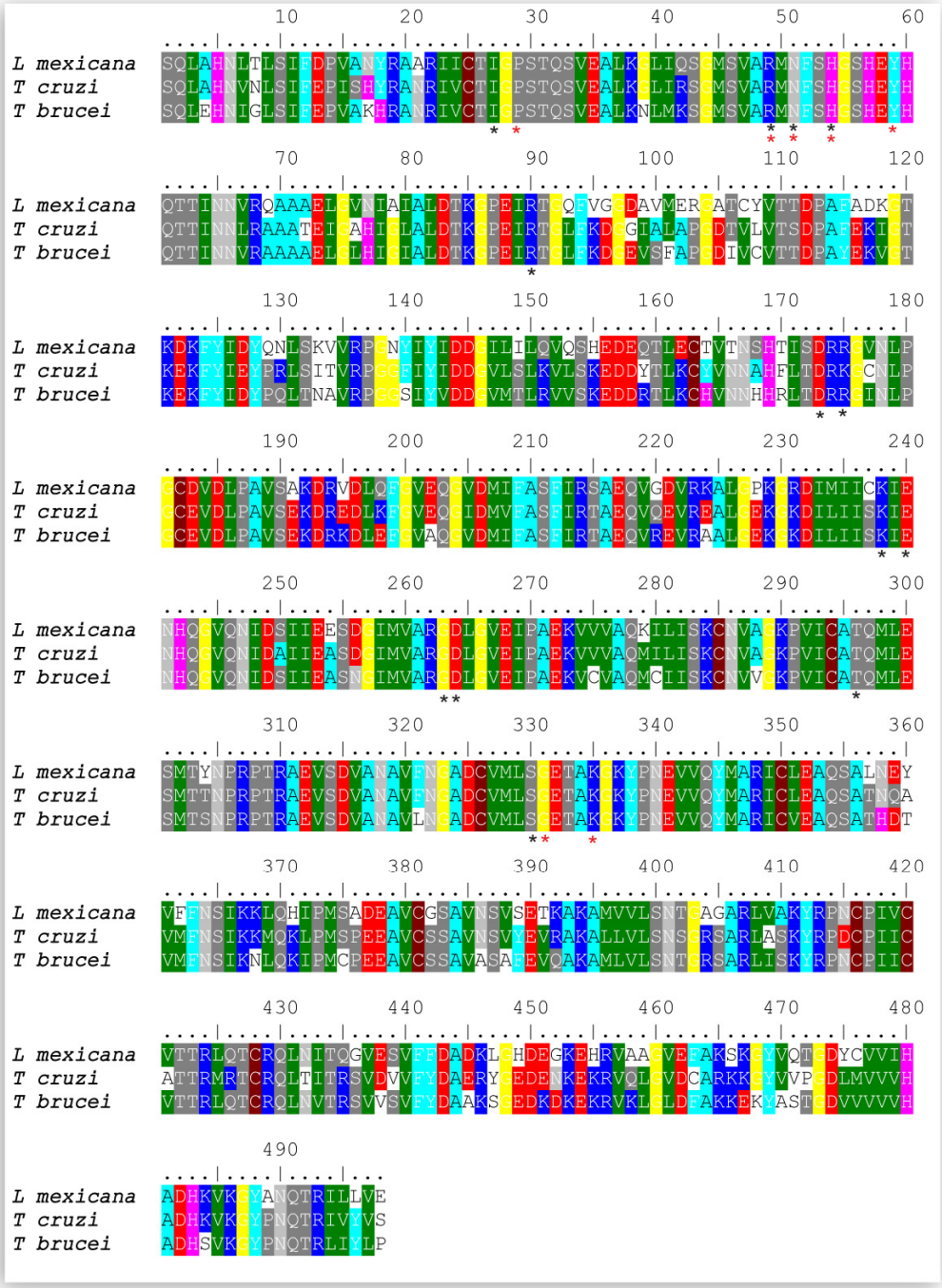

Figure S1 Sequence alignment of three trypanosomatid PYKs.

All amino acid residues are shown, and conserved residues are indicated as colored blocks. Amino acids involved in ATP binding (*) and suramin binding (*) are indicated by asterisks.

Figure S2

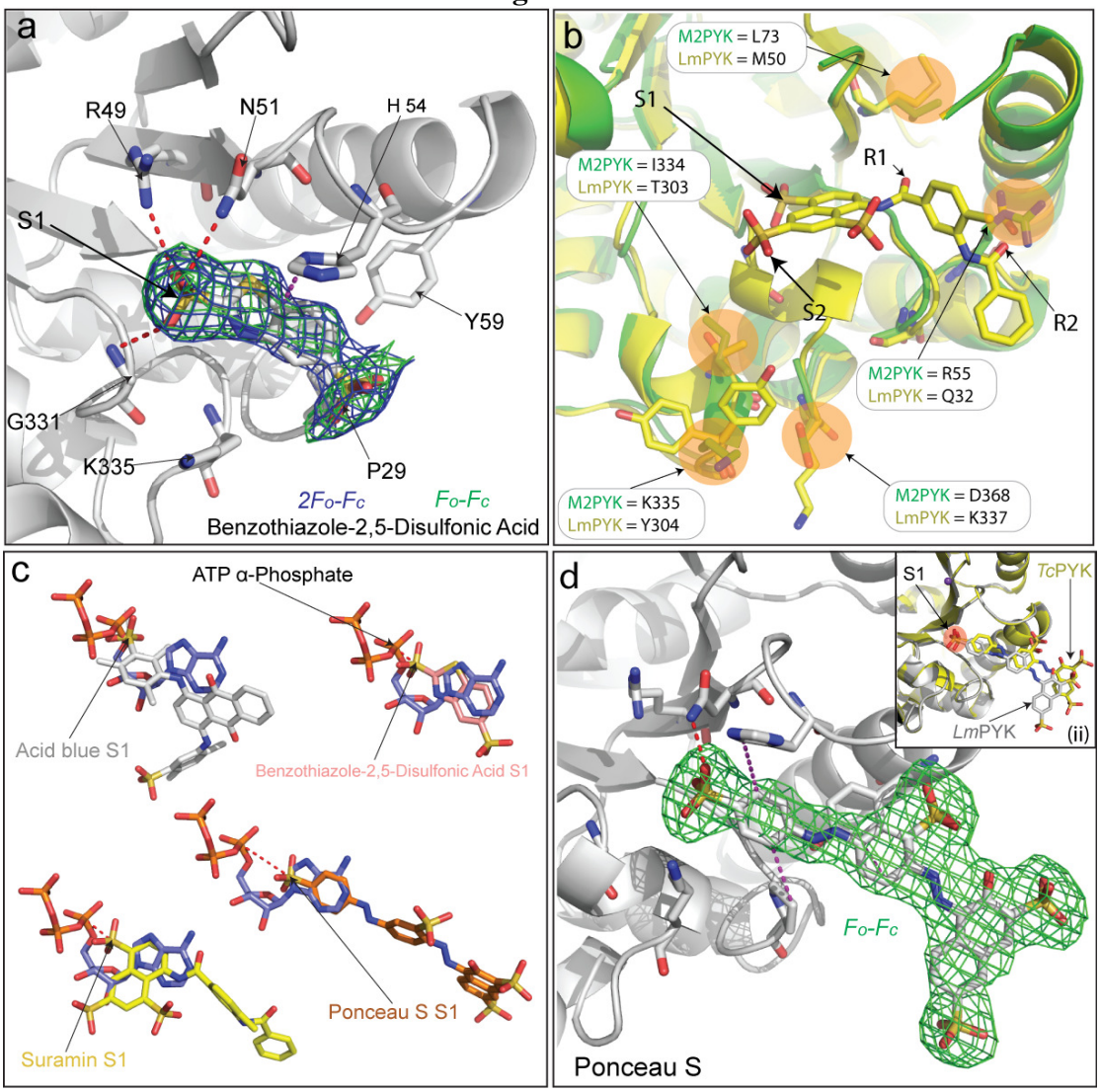

Figure S2. Binding positions of sulfonic acid groups in relationship to the γ-phospho position of the ATP molecule.

(**a**) Close up of the active site of the *Lm*PYK-BDS structure. The BDS molecule is shown with an unbiased F_o-F_c electron density map contoured at 2.5 σ (green). The 2.45 Å resolution electron density ($2F_o$-F_c map) map is also shown (blue), contoured at 1σ. (**b**) The human M2PYK monomer (colored green – PDBID = 3BJF) superposed onto the *Lm*PYK-suramin structure (colored yellow). The A- and C-domains of both structures were superposed (RMS fit of the C-α atoms of domains A and C is 1.0 Å). Differing amino acids in and around the suramin binding site have been highlighted (orange circles). Amino acid residues involved in binding positions S2, R1 and R2 of suramin could be modified to provide inhibitor selectivity

for *Lm*PYK over M2PYK and vice-versa. (**c**) The sulfur belonging to sulfonic acid group 1 (S1) of acid blue 80 is 1.3 Å away from the phosphorous of the α-phospho group of ATP, compared to 2.6 Å for suramin, 2.0 Å for BDS and 4.3 Å for Ponceau S. (**d**) Close up of the active site of the *Tc*PYK-Ponceau S structure. The Ponceau S molecule is shown with an unbiased F_o-F_c electron density map contoured at 2.5 σ (green). The 2.45 Å resolution electron density ($2F_o$-F_c map) map is also shown (blue), contoured at 1σ. **Inset ii.** The *Tc*PYK-Ponceau S structure superimposed onto the *Lm*PYK-Ponceau S structure, with an RMS fit of 0.41 Å. The sulfonic acid group (S1) binds to the active site of both *Lm* and *Tc*PYK in a near identical position.

Figure S3

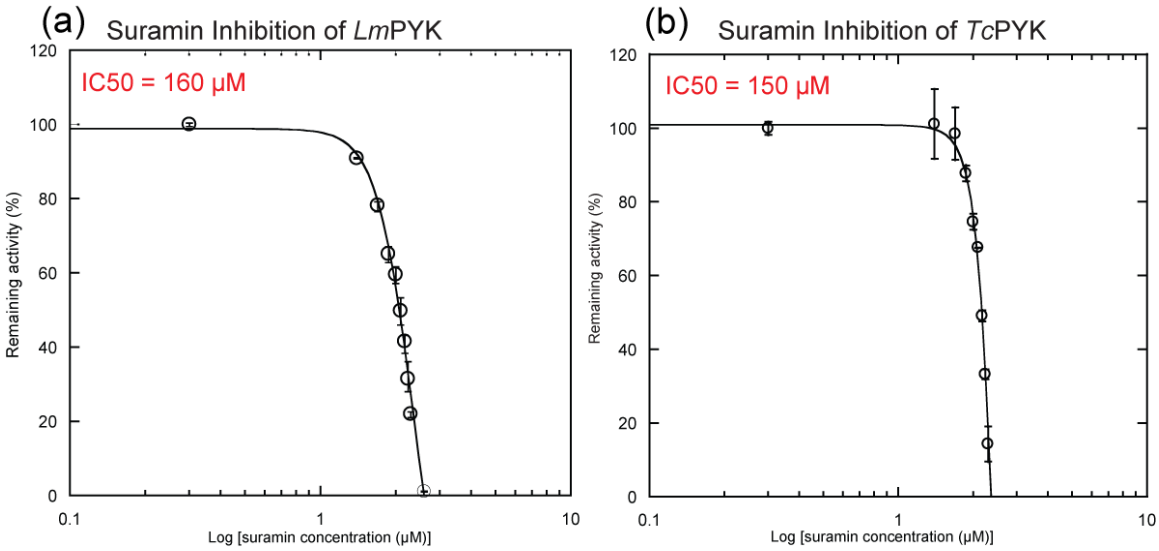

Figure S3. Suramin inhibition of trypanosomatid PYKs.

Concentration-response curves observed for the titration of suramin against *Lm*PYK (**a**) and *Tc*PYK (**b**) activity, with IC_{50} values of 160 µM and 150 µM, respectively. The LDH-linked assays were performed in triplicate.

Figure S4

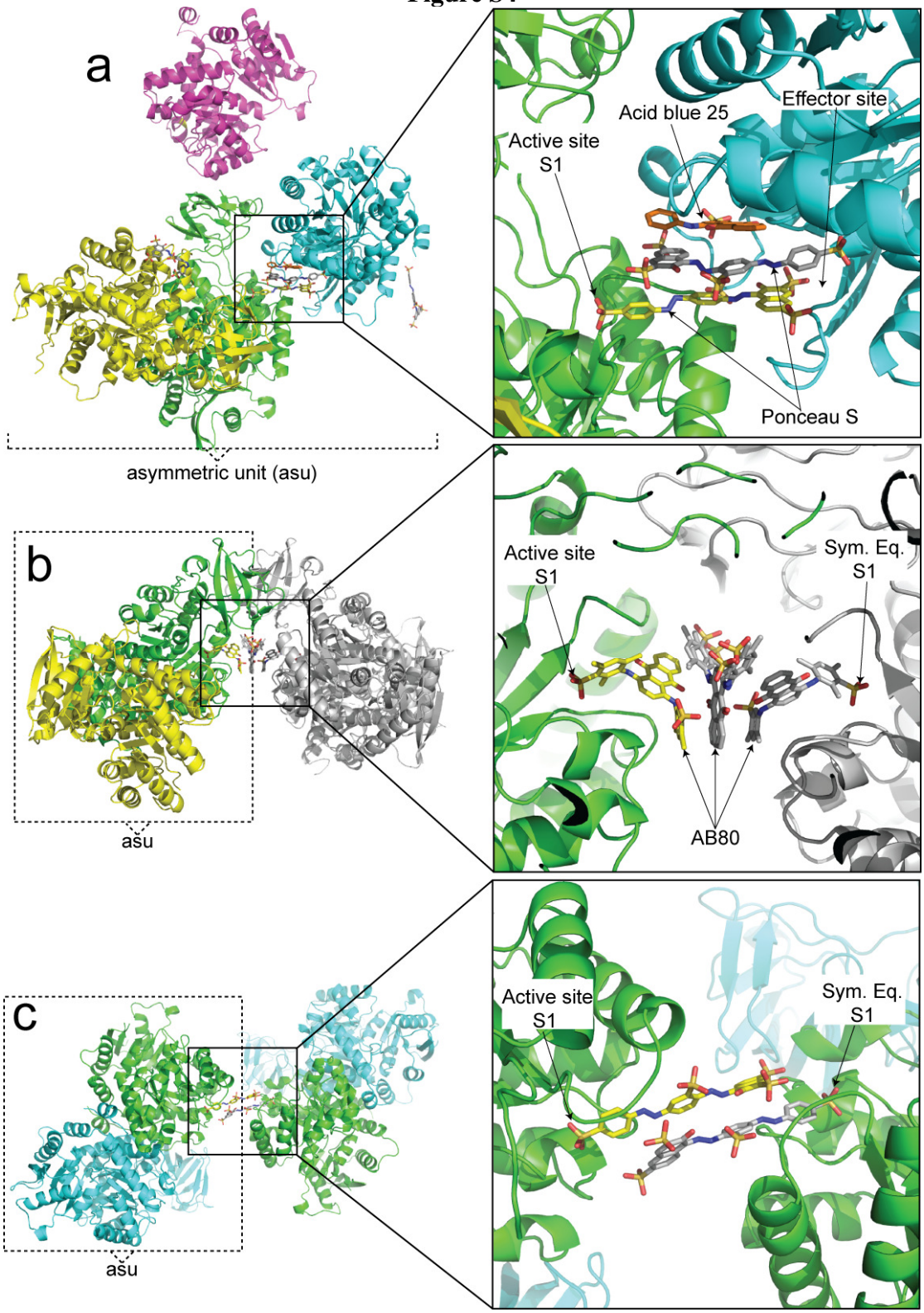

Figure S4. Asymmetric unit of trypanosomatid PYK crystal structures.

The arrangement of PYK monomers making up an asymmetric unit (asu) of: (**a**) *Lm*PYK co-crystallised in the presence of Ponceau S and Acid blue 25; (**b**) *Lm*PYK co-crystallised in the presence of acid blue 80 (AB80); (**c**) *Tc*PYK co-crystallised in the presence of Ponceau S.

Although all ligands bind to the active site in the manner discussed previously (**Figure 1**) they also provide an intermolecular bridge between macromolecules in the crystal.

REFERENCES

1. Valentini, G., *et al.* (2002) *J. Biol. Chem.* **277**, 23807-23814
2. Wernimont, A., *et al.* Crystal Structure of Pyruvate Kinase from toxoplasma gondii. *(PDB ID = 3EOE).*
3. Painter, J., and Merritt, E. A. (2006) *Acta. Cryst. D* **62**, 439-450
4. Lima, L., *et al.* (2009) *Biochim. Biophys. Acta.* **1794**, 873-881
5. Zhou, X., *et al.* (2008) *Biophys. J.* **95**, 3366-3380
6. Murakami, M. (2005) *J. Mol. Biol.* **350**, 416-426
7. Schuetz, A., *et al.* (2007) *Structure* **15**, 377-389
8. Ganesh, V. K., Muthuvel, S. K., Smith, S. A., Kotwal, G. J., and Murthy, K. H. M. (2005) *Biochem.* **44**, 10757-10765