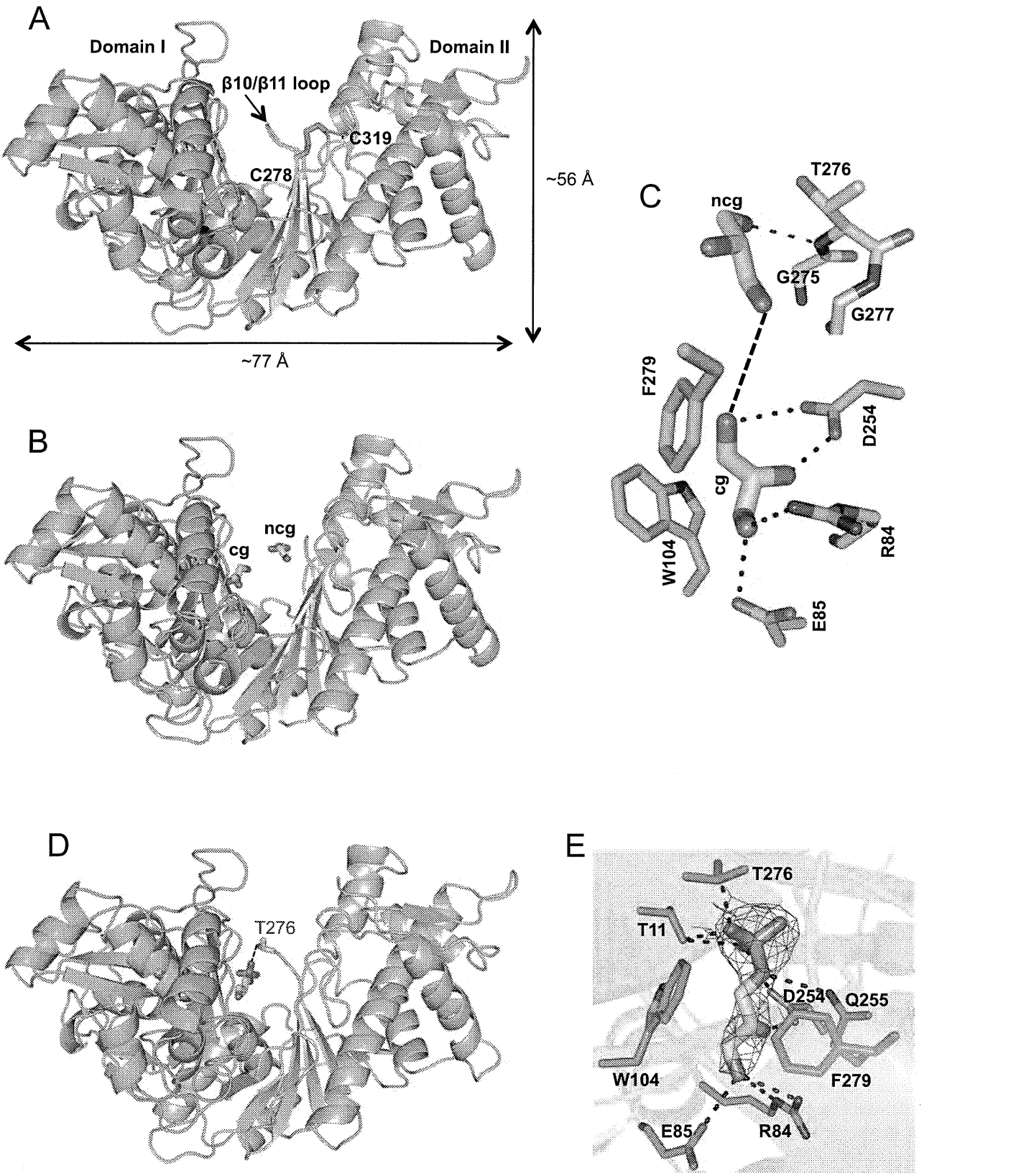

(Koga *et al.*, 2008) and *Cellulomonas* sp. (Imada *et al.*, 2006). The domains are linked by an α11/β10 loop (residues 262–267) and form a deep V-shaped active site cleft between them (Fig. 2A). The lower portion of the cleft is constructed by α-helices and β-sheets from domains I (α11 and β9–β4–β1–β2–β3) and II (α16 and β12–β13–β11–β10–β16–β17–β18), whereas the upper part is mainly assembled by β1/β2, β4/β5, β5/α2, α3/α4 and β10/β11 loops (Fig. 2A, Fig. S2A–C). In addition, non-conserved Cys278 and Cys319 located on the β10/β11 and β13/α12 loops, respectively, form a disulfide bond (Fig. 2A), which makes the β10/β11 loop to assume a closed conformation in contrast to an open conformation observed in the structure of the glycerol-bound form as described below. The closed conformation of the β10/β11 loop explains why the preparation of ligated crystals by soaking or cocrystallization was unsuccessful when ligand-free TbgGK was used.

In contrast to the GKs of *E. coli* (Hurley *et al.*, 1993), *E. casseliflavus* (Yeh *et al.*, 2004) and *S. aureus* (Minasov *et al.*, 2009), which are able to physiologically exist as both homodimer and homotetramer, TbgGK is exclusively a homodimer (Balogun *et al.*, 2013) like the GKs from *P. falciparum* (Schnick *et al.*, 2009), *T. kodakaraensis* (Koga *et al.*, 2008) and *Cellulomonas* sp. (Imada *et al.*, 2006). Since the tetramerization is a prerequisite for the inhibitory regulation of GK by fructose 1,6-bisphosphate (Yu and Pettigrew, 2003), the homodimer structure of TbgGK may explain its insensitivity to fructose 1,6-bisphosphate regulation, another attribute contrasting GKs of African trypanosomes (Kralova *et al.*, 2000; Haanstra *et al.*, 2008) from those of other organisms (Zwaig and Lin, 1966).

### *Glycerol and glycerol 3-phosphate binding site*

The overall structures of TbgGK in complexes with glycerol and G3P are essentially identical to that of the ligand-free form (inferred from the respective average rmsd values of 0.94 and 1.52 Å calculated for superposed Cα positions) (Fig. 2A, B and D), except that they possess no disulfide bond between Cys278 and Cys319 as indicated by Sγ–Sγ distances larger than 8.2 Å.

In the TbgGK-glycerol complex, two glycerol molecules (catalytic glycerol and non-catalytic glycerol) are bound to the active site cleft of each TbgGK monomer (Fig. 2B). One is located at the canonical site of the catalytic glycerol and forms hydrogen bonds with conserved residues of domain I (Arg84, Glu85, and Asp254; Fig. 2C and Fig. S3) as observed in all GKs with known structures [*E. casseliflavus* (Yeh *et al.*, 2004), *S. aureus* (Minasov *et al.*, 2009), *E. coli* (Hurley *et al.*, 1993) and *P. falciparum* (Schnick *et al.*, 2009)], whereas the other glycerol molecule (non-catalytic glycerol) that is not found in the active sites of other GKs binds at the non-canonical site near the β10/β11 loop and forms a hydrogen bond with Thr276 (Fig. 2B and C). It appears that the non-catalytic glycerol prevents Cys278 and Cys319 from forming the disulfide bond observed in the ligand-free TbgGK (Fig. 2A) and keeps the β10/β11 loop in the open state, which explains why glycerol 3-phosphate (G3P) and ADP can be bound to the TbgGK-glycerol complex but not to the ligand-free TbgGK.

A single G3P molecule is bound to each TbgGK monomer with its glycerol moiety in the catalytic glycerol site and the phospho group near the non-catalytic glycerol site (Fig. 2D and E). In addition to hydrogen bonds that are involved in the binding of the catalytic and non-catalytic glycerol, the phospho group accepts additional hydrogen bonds from the conserved Thr11 and Gln255 (Fig. 2E and Fig. S3, Table S2). The β10/β11 loop of the G3P-bound form is in the closed conformation, maintained by a hydrogen bond between the G3P phospho group and Thr276 in the loop. In this form, the loop acts as a lid veiling the bound G3P (Fig. 2D).

### *Novel binding site of ADP in TbgGK*

The ADP-bound structure of the TbgGK homodimer reveals an interesting peculiarity of the enzyme *i.e.* that chains A and B differ remarkably to each other in their ADP binding site and binding mode. One ADP molecule (active site ADP) is bound to the active site of chain A, whereas in chain B, two ADP molecules (non-active site ADP) are found at a so far not yet described ligand binding site. This site is on the molecular surface, at a distance far away (~ 30 Å) from the active site (Fig. 3A).

The active site ADP (ADP_as) accepts hydrogen bonds from Thr12 and Gln255 and takes a *Syn* configuration with its ribose group pointing to the bulk solvent, which is in contrast to the *Anti* configuration observed in structures of other GKs [*S. aureus* (Minasov *et al.*, 2009), *P. falciparum* (Schnick *et al.*, 2009) and *E. coli* (Feese *et al.*, 1994)]. There are seven completely conserved residues (Thr11, Ser13, Gln255, Thr273, Gly275, Thr276 and Phe279) and two highly conserved residues (Gly10 and Thr12) within the distance of 4 Å from the active site ADP (Fig. 3B). It is worth noticing that the active site ADP is surrounded by an 'army' of four threonine residues (Thr11, Thr12, Thr273, and Thr276). As described above, Thr11 and Thr276 form hydrogen bonds with the phospho group of G3P, whereas the phospho group of ADP accepts hydrogen bonds from Thr12. In order to assess the importance of these threonine residues, the enzymatic activity of the reverse reaction was analyzed for mutated versions of the TbgGK proteins (T11V, T12V, T273V, and T276V mutants).

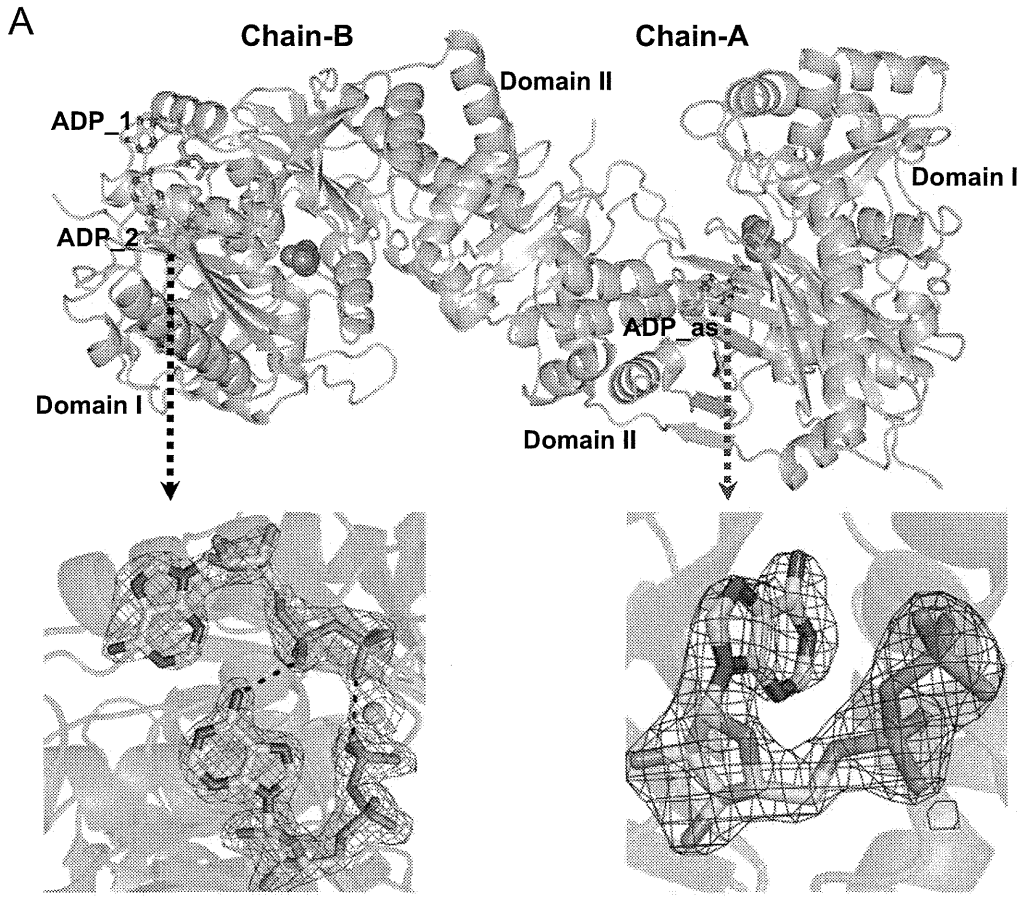


**Fig. 3.** Structure of TbgGK in complex with ADP (yellow sticks), glycerol (red and purple spheres) and $Mg^{2+}$ (yellow sphere).
A. Dimer structure. In chain-A, ADP and glycerol molecules are proximally bound in the active site cleft to domains II and I respectively. ADP is not bound in the active site cleft of chain-B, but the adduct of two ADPs intercalated by a magnesium ion (yellow sphere) is bound far (~ 30 Å) from the active site. The black dashed arrow points to a zoom of the 2ADP·$Mg^{2+}$ adduct with the $2F_o$–$F_c$ map (blue mesh) contoured at 2.0 σ showing the density for the complex of two ADP molecules with a magnesium ion bound to chain-B. The blue dashed arrow points to the $2F_o$–$F_c$ map contoured at 2.0 σ showing the density for the ADP molecule bound in the active site of Chain-A.
B. Residues around ADP. The binding of ADP in the active site of TbgGK accomplished by an 'army' of four threonine residues. Dots represent hydrogen bonds.
C. Enzymatic activities of mutants for the threonine residues. The upper panel presents an SDS-PAGE electrophoretogram showing the purity of each purified mutant protein used for activity measurements (lower panel) alongside the wild-type (wt) enzyme as a control.
D. Hydrogen bonds depicted as dots between the Mg·$ADP_2$ adduct and TbgGK. Yellow sticks and the green sphere represent ADP and the magnesium ion respectively.
E. SDS-PAGE electrophoretogram (upper panel) and enzymatic activities (histogram) of the purified mutants for residues interacting with the Mg·$ADP_2$ adduct.
F. Alignment of the sequence containing the residues (highlighted with asterisks) involved in binding the Mg·$ADP_2$ adduct; the residues are not conserved.

While T273V mutant retains an activity comparable with that of wild-type TbgGK, the other mutants show almost complete loss of activity (Fig. 3C).

In chain B, an adduct formed by two ADP molecules and a $Mg^{2+}$ ion (AMP–$PO_4^{2-}$···$Mg^{2+}$···$^{2-}PO_4$–AMP) is bound to the novel site (Fig. 3A) through hydrogen bonds with six residues (Tyr3, Asp20, Arg22, Arg24, Arg477 and Glu478; Fig. 3D and Table S3). It is apparent from Fig. 3E that among the mutations of these residues to alanine, R24A, R477A and E478A caused considerable (70–90%) impairment of the enzymatic activity. Although the roles of R24, Arg477, and E478 and the 2ADP·$Mg^{2+}$ adduct in the reverse reaction of TbgGK are currently unclear, it is important to point out that these residues are not conserved across organisms (Fig. 3F) but conserved only within species of the genus *Trypanosoma*.

### *Proposed mechanism of the reverse reaction*

Although structures of TbgGK in complexes with G3P and ADP reveal that these ligands are bound to different sites in the active site of TbgGK, all attempts to prepare crystals

 

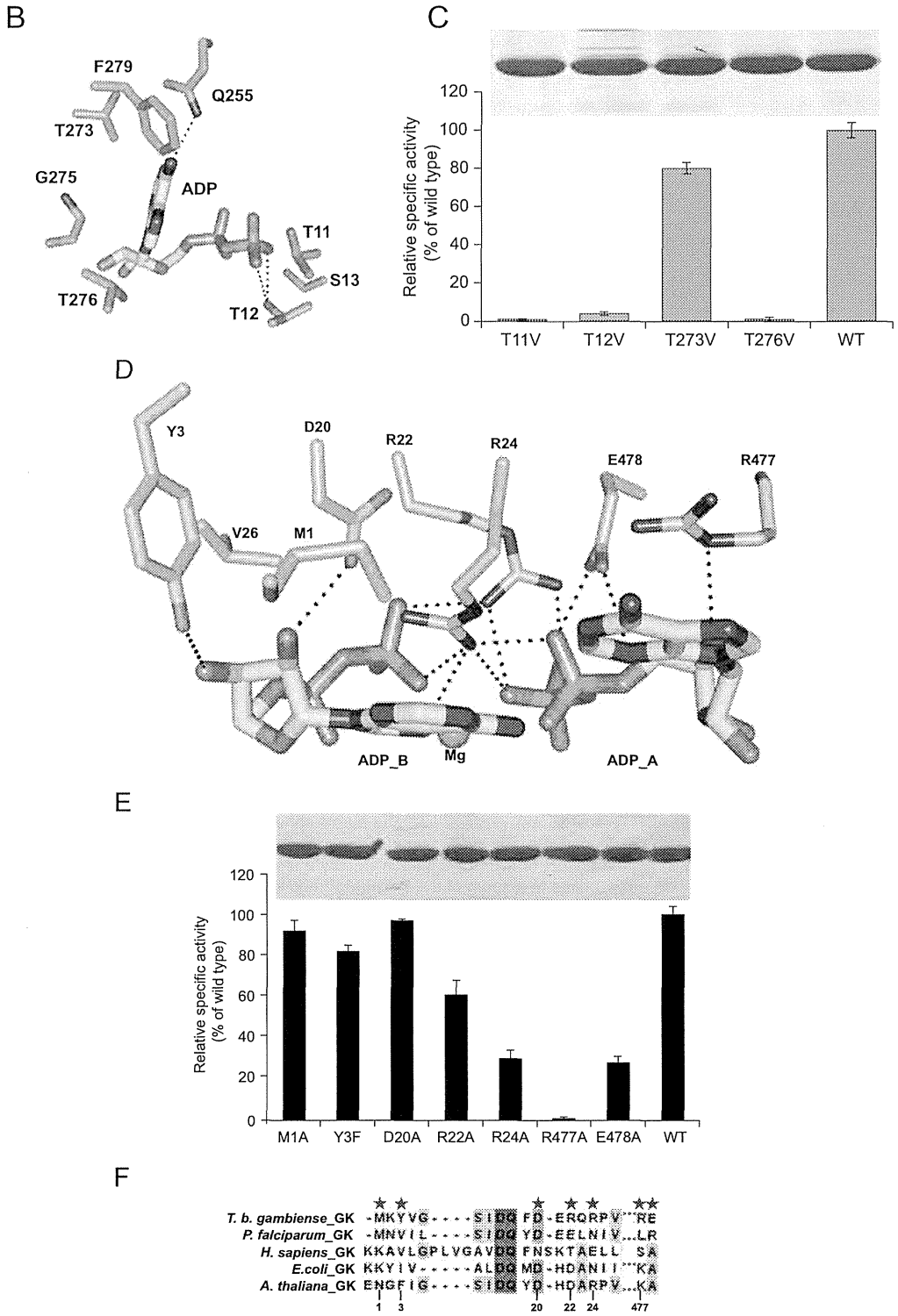


**Fig. 3.** *cont.*

of a ternary TbgGK–G3P–ADP complex were unsuccessful. While crystal damaging possibly due to drastic conformational changes was observed after soaking TbgGK–ADP crystals with G3P, intact data-quality crystals were obtainable when TbgGK–G3P crystals were soaked with ADP, although ADP could not bind even after an extended soaking time. The latter case is probably because the closed conformation of the β10/β11 loop observed in the

TbgGK–G3P complex structure prevents ADP from binding to the active site. This is in contrast to *E. coli* GK whose ternary structure in complex with both the G3P analog glyceraldehyde-3-phosphate (G3H) and ADP has been determined (Feese *et al.*, 1994). The region of Tyr265–Gly268 in the ternary structure of *E. coli* GK, which corresponds to the β10/β11 loop of TbgGK, is in a relatively open conformation. This explains why both substrates were able to bind to the *E. coli* enzyme, and corroborate the random mechanism reported for this GK (Pettigrew *et al.*, 1990). Formation of the closed conformation in TbgGK is brought about by hydrogen bonds between Thr276 of the β10/β11 loop and the phospho group of G3P. Since the β10/β11 loop must open up in order to accommodate ADP, we therefore rationalize the pertinent facts described above, and propose a mechanism that we refer to as a 'local signal transduction' for the reverse reaction of TbgGK as summarized in Fig. 4. First, G3P binds to the active site with its phospho group oriented in proximity to Thr11 Oγ, which becomes transiently autophosphorylated through a nucleophilic attack of its hydroxyl group on the phosphorous atom of the phospho group, producing free glycerol. Although both Thr11 and Thr276 form H-bonds with the phospho group of G3P, the distance (> 4 Å) between Thr276 Oγ and P is relatively too long for a nucleophilic attack on P, hence, Thr11 is located in a favorable position (~ 2 Å). In addition, the nucleophilicity of Thr11 Oγ seems to be enhanced by the hydrogen bond between Thr11 Oγ and Asp254 mediated by an oxygen atom of the G3P's phospho group (Fig. 2E). Second, the transformation of G3P into the first product, glycerol, accompanied by the phosphorylation of Thr11, would results in breaking the hydrogen bond between Thr276 and G3P, which causes opening up of the β10/β11 loop from the active site cleft towards domain II. Finally, ADP is bound to the active site and the reversed autophosphorylation reaction occurs through an inline phospho transfer from the transition state phosphothreonine to ADP, and the second product, ATP, is produced. Release of the ATP and binding of the next G3P will again trigger the closure of the β10/β11 loop to initiate the next cycle of catalysis (Fig. 4). In this model, mobility of the β10/β11 loop through ordered substrate binding (G3P is bound and dephosphorylated before ADP binding) is key to enabling the reverse reaction as compared to the simultaneous substrate binding reported for other GKs where the reaction is intrinsically less favored.

The proposed signal transduction-like mechanism for the reverse reaction catalyzed by TbgGK currently seems to be the most logical based on our structural and mutagenic analyses. As reviewed by Johnson and Barford (1993), a number of kinases display such a mechanism, including the recent example of autophosphorylation in the proto-oncogene Akt (Li *et al.*, 2006).

## Discussion

This study provided the first structure of the glycerol kinase from a global health-threatening pathogen, the African human trypanosome, *T. b. gambiense*. It also represents the first structure for any potentially essential eukaryotic GK. The structure was determined in the both ligand-free form and forms complexed with glycerol, G3P and ADP. The enzyme, similar to earlier reported GKs (of *E. coli*, *E. casseliflavus* and *P. falciparum*), is a homodimer with each monomer composed of two functional domains and containing the central four-stranded antiparallel β sheet (β10, β11, β13, and β12) (Fig. 1A, Figs S2C and S3), a secondary structural signature that is lacking in other members of the sugar kinases/hsp70/actin superfamily (Hurley *et al.*, 1993; Yeh *et al.*, 2004; Schnick *et al.*, 2009). Despite the high sequence identities (Fig. S3), which culminated in the overall similar structural architecture of TbgGK and GKs from the above organisms (as judged from the percentage of identity and rmsd values, *P. falciparum* GK is the closest), the present crystal structure reveals a number of intriguing differences such as: structural re-arrangements in response to substrates binding; substrate binding/catalytic sites; and importantly, features that seem to promote the rate of performing the reverse catalysis. Interestingly, some of these differences stem from the structurally non-conserved portion of the protein and manifests in unique biological functions and properties.

Contrary to the usual closure of the active site cleft occurring in members of the kinases/hsp70/actin superfamily upon substrate binding, it is herein shown that substrate binding can induce a dramatic conformational change and lead to the creation of an additional cleft (Fig. S4) in a member of the superfamily upon the binding of a substrate. This attribute, which helps us to explain the catalytic mechanism of TbgGK, may be peculiar to bisubstrate-biproduct enzymes. In the context of TbgGK's reverse reaction, the sequence of substrate binding and effects is summarized as follow: G3P is the first substrate; when bound, the catalytic cleft is in a closed conformation, aiding its de-phosphorylation. This is corroborated by the fact that some G3P coordinating residues (Thr11 and Thr276) in the enzyme active site play prominent roles in maintaining the closed conformation (Fig. 2D and E), and the enzyme lost activity when these residues were mutated to alanine (Fig. 3C). The closed conformation is kept by a H-bond between the phospho group of G3P and the hydroxyl group of Thr276 in the β10/β11 loop (Fig. 2D). The products of G3P dephosphorylation are glycerol and, most likely, a transiently phosphorylated enzyme; such autophosphorylated protein intermediates have been reported as mechanistic attributes of some kinases and phosphatases

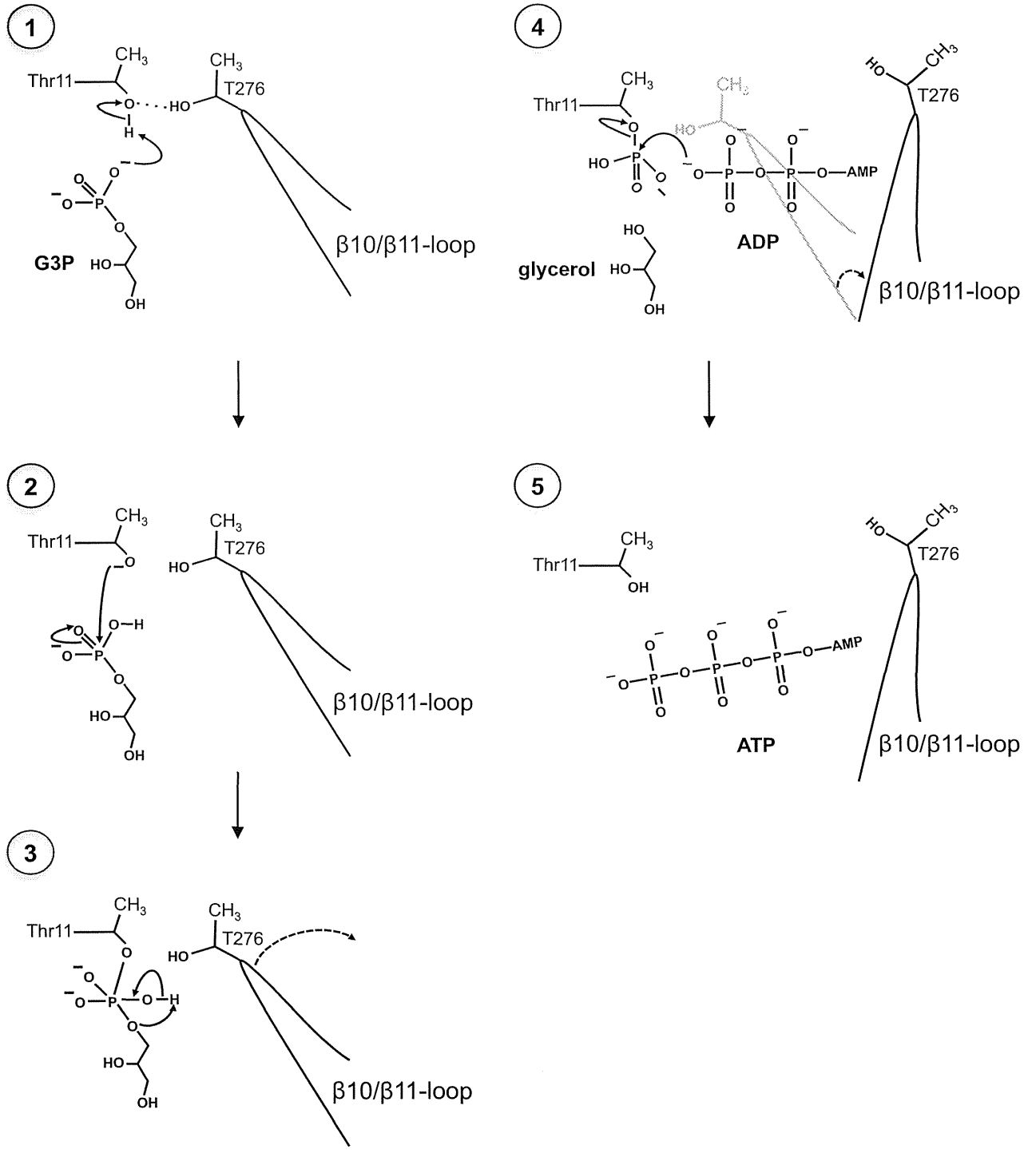


**Fig. 4.** The proposed reverse reaction mechanism of TbgGK. (1) Activation of Thr11. A nucleophilic attack by O2P of G3P on the OH group of Thr11 converts it to an oxyanion-containing residue. (2) Reversed nucleophilic attack. The activated Thr11 becomes reactive and nucleophilically attacks the P-atom of G3P, yielding a transition state pentacoordinated P atom. (3) Autophosphorylation and loop movement. Thr11 becomes transiently phosphorylated, and the β10/β11 loop that was partly maintained by H-bond (dashed line) between Thr276 (in the loop) and Thr11 shifts (dashed arrows) out of the active site. (4)–(5) Binding of ADP and formation of ATP. The movement of the loop creates the ADP binding site, which binds with its β-phosphate proximally oriented towards the transition state phosphothreonine and accepts the phosphate group through another nucleophile-mediated associative phospho transfer to yield ATP. Release of the ATP and binding of the next G3P will again trigger the lid's structural re-adjustment into the active site cleft to initiate the next cycle of catalysis.

(Llinas *et al.*, 2005; Zorba *et al.*, 2014). With reference to the G3P-bound enzyme, the presence of glycerol produced marked conformational changes in the protein (rmsd for 510 Cα atoms = 1.1 Å) causing a combined 10° outward rotation of both domains and importantly, an average outward rotation of 19° by β10, β11, and β13 that resulted in the creation of a secondary catalytic cleft (Fig. S4C). Interestingly, this newly created cleft at the catalytic site forms part of the ADP binding site of TbgGK. Without the creation of the new groove, ADP was unable to bind in the catalytic cleft. It therefore can be interpreted that dephosphorylation of G3P to produce glycerol precedes the ADP binding required for the last part of the catalytic cycle. The dephosphorylation breaks the H-bond between G3P and Thr276, which causes the β10/β11 loop to open, creating the secondary catalytic cleft (Fig. S4C). ADP is bound in the catalytic site with its ribose ring and adenine head portion at the formed secondary cleft and its phospho group proximally oriented towards the G3P-dephosphorylating cleft. The β-phosphate is only 2.1 Å away from the predicted transiently phosphorylated Thr11 residue (Fig. 3B). This distance is optimal for transphosphorylation between an acceptor and a donor via the associative mechanism (Schlichting and Reinstein, 1997; Lahiri *et al.*, 2003).

Another striking revelation in the present work is the manner in which the substrates were coordinated to the enzyme: First, there is an additional glycerol (ncg) binding site, which is exclusively present in domain II (Fig. 2B). The glycerol at this site cannot be phosphorylated, the ADP-bound structure of TbgGK revealed that it is actually the ADP binding site (Fig. S4D), and the glycerol could be loosely bound via the hydroxyl groups. Their differential affinities explain why ADP easily displaced the ncg when crystals of glycerol-bound TbgGK were soaked in a solution containing ADP. Second, ADP binding in the TbgGK active site is accomplished by a set of residues (Thr11, Thr12, T273, G275 and T276) that are completely different in *P. falciparum* GK (Thr268, G313, V316, P328, S332, G414 and N418). In the GK of other organisms, the ADP is bound by residues, which line their active sites, and are well conserved among them (Hurley *et al.*, 1993; Yeh *et al.*, 2004; Schnick *et al.*, 2009). However, these residues are not structurally conserved in TbgGK and are, on average, about 16 Å away from its active site. The third difference in terms of substrate binding is the coordination of additional ADPs as the $Mg^{2+}$·$ADP_2$ adduct on the enzyme surface far away from the catalytic site, by residues that are conserved only in different specie of African trypanosomes (Fig. 3A, D and F). It is well established that GK of African trypanosomes lack the usual regulatory mechanism of other GKs (Haanstra *et al.*, 2008), but our mutagenesis experiments point to the essentiality of the novel Mg·$ADP_2$ binding site to TbgGK activity (Fig. 3E). Therefore, the finding that the $Mg^{2+}$·$ADP_2$ adduct binds to TbgGK at this unique site raises a question about a novel regulatory mechanism of African trypanosomal GK. The participation of $Mg^{2+}$·$UTP_2$ has recently been described as 'a new paradigm' for allostery in aspartate transcarbamoylase of *E. coli* (Cockrell *et al.*, 2013). Therefore, we hypothesize that the simultaneous binding of ADP and G3P in the active site of both monomers of TbgGK may limit its catalytic rate relative to the situation when only a single monomer is occupied. Binding of Mg·$ADP_2$ to one monomer, prevented ADP from binding in the active site of same monomer (Fig. 3A and Fig. S5), hence, the Mg·$ADP_2$-free monomer converts the ADP and G3P at a higher rate. Interestingly, this unique site corresponds to the segment of the *E. coli* protein that makes it susceptible to allosteric inhibition by the $IIA^{Glc}$ (Yu *et al.*, 2007). Binding of $IIA^{Glc}$ to *E. coli* GK induces unwinding of a coil and α-helix structures at position 472–479, to a $3_{10}$-helix (Holtman *et al.*, 2001).

In conclusion, the structure of glycerol kinase from African human trypanosomes provides the first comprehensive picture of a GK that seem to have evolved to optimize its rate for reverse catalysis. The structure revealed a number of unique features of African trypanosomes GK, which include: (i) formation of the catalytic site by structurally non-conserved threonine residues (Thr11, Thr12, Thr273, and Thr276) that participate in the catalysis, (ii) structural re-arrangements that possibly resulted from a short-lived autophosphorylation of threonine residues, leading to the creation of a secondary cleft for the accommodation and proximal orientation the phospho group-acceptor substrate, ADP, and (iii) the presence of an additional ADP binding site away from the active site, and which may play regulatory roles in this enzyme. These structural features have not been reported in homologous proteins, and provide an additional explanation for the feasibility of the reverse reaction that has so far been detected only for the trypanosomal GK. Importantly, these peculiarities have led us to conclude that the structure of African trypanosomes GK, in addition to the reason of compartmentalization of glycolysis that has been proposed previously, contributes to enabling the reverse catalysis. Further, the highlighted uniqueness of this enzyme has provided a potential molecular framework for the development of highly specific inhibitors of this parasite's protein that are expected to spare the human homolog. Indeed, we have utilized the structural information to identify an encouraging number of inhibitors with novel scaffolds (data not shown). The combination of the present data with those reported on the recent advances on TAO by Shiba *et al.* (2013) may be useful in the development of a new generation of drugs against sleeping sickness.

## Experimental procedures

### Cloning, expression, and purification

Details on production of bacterially expressed TbgGK are described in Balogun *et al.* (2013). Briefly, a blunt ended GK encoding gene (*gk*) was amplified from cDNA of Tbg (IL2343), gel-purified, and ligated into the pET151/D-TOPO vector (Invitrogen) to create a recombinant plasmid construct. The construct was amplified in One Shot TOP10 *E. coli* cells (Invitrogen) and then used to transform the *E. coli* JM109 (DE3 + pRARE2) expressor strain (Novagen). Cells of a selected transformant were grown as a 10 l culture at 37°C under 600 r.p.m. agitation and maximum aeration in a BIOFLO 2000 fermenter to $OD_{600}$ = 0.6, IPTG was added to a final concentration of 25 μM, and the culturing was continued at 20°C until late log-phase. Cells were harvested by centrifugation, resuspended in lysis buffer [100 mM phosphate buffer, pH 6.8, 50 mM imidazole, 300 mM NaCl, 10 mM $MgSO_4$, 0.1 mM PMSF, 1 mg $ml^{-1}$ lysozyme and 10% (v/v) glycerol], and disrupted in a French pressure cell operated at 140 MPa. The soluble fraction obtained by centrifugation was then applied to a Ni-NTA Agarose column (Qiagen; 1.5 × 15 cm) of 15 ml bed volume that was pre-equilibrated with buffer A [100 mM phosphate buffer, pH 6.8, containing 50 mM imidazole, 300 mM NaCl, 10 mM $MgSO_4$ and 1% (v/v) glycerol]. After washing the column with 100 ml of buffer A, TbgGK was eluted with 500 ml of buffer A containing a linear gradient of 50–500 mM imidazole. Fractions containing active TbgGK with higher purity as assessed by SDS-PAGE (Laemmli, 1970) were pooled, concentrated to approximately 40 mg TbgGK $ml^{-1}$ using a centrifugal ultrafiltration tube (Amicon Ultra-15, 30 000 cut-off; Millipore), and stored at –30°C in the presence of 50% (v/v) glycerol until the next purification step. About 5 mg of the affinity-purified protein was further purified by gel-filtration chromatography using Superdex 200 (1 × 30 cm; GE Healthcare Bio-sciences) equilibrated with 100 mM phosphate buffer, pH 6.8, containing 0.3 M NaCl and 1% (v/v) glycerol. The elution was carried out at a flow rate of 0.5 ml $min^{-1}$ by high-performance liquid chromatography (HPLC). Each fraction (0.5 ml) was analyzed on SDS-PAGE, and fractions containing highly pure TbgGK were pooled. After exchange of the buffer for 10 mM MOPS buffer, pH 6.8, containing 10 mM $MgSO_4$, 20 mM NaCl, and 0.001% (v/v) glycerol, the purified TbgGK was concentrated to about 10 mg $ml^{-1}$ for crystallization experiments. The presence of glycerol during the purification was necessary to keep the enzymatic activity of TbgGK.

### TbgGK assay

The TbgGK activity was assayed using the reverse reaction (glycerol 3-phosphate + ADP → glycerol + ATP). To 1.0 ml of the reaction mixture (1 mM EDTA, 5 mM $MgSO_4$, 0.5 mM $NADP^+$, 50 mM glucose, 2 mM ADP, 10 mM glycerol 3-phosphate and 1 unit of each hexokinase and glucose 6-phosphate dehydrogenase), TbgGK was added at 37°C. Using ATP produced by TbgGK, hexokinase converts glucose to glucose 6-phosphate, and finally glucose-6-phosphate dehydrogenase produces NADPH from glucose 6-phosphate and $NADP^+$. The rate of NADPH accumulation was spectrophotometrically monitored at 340 nm using the JASCO V-660 spectrophotometer. TbgGK was added after measurement of background activity for 1 min.

### Crystallization

Crystallization trials were performed using the sitting drop vapor diffusion method (Balogun *et al.*, 2010). The best crystals of the TbgGK-glycerol complex were prepared using a TbgGK solution (5 mg $ml^{-1}$) dissolved in 10 mM MOPS buffer, pH 6.8, containing 10 mM $MgSO_4$, 20 mM NaCl, 0.001% (v/v) glycerol and a reservoir solution composed of 0.1 M HEPES pH 7.5, 11% hexane-1,6-diol, 25% PEG400. Glycerol-free TbgGK was prepared by repeated dilution with buffer (10 mM $MgSO_4$, 20 mM NaCl and 10 mM MOPS, pH 6.8) and concentration with Amicon Ultra-15, and crystallized using 10% sorbitol, 12% isopropanol, and 0.1 M HEPES, pH 7.0, as a reservoir solution. All crystallization experiments were carried out at 20°C. Crystals of the TbgGK–ADP and –G3P complexes were prepared by soaking crystals of the glycerol bound form for 60–120 min in reservoir solution supplemented with 5 mM of either ADP or G3P.

### Data collection and structure determination

X-ray diffraction experiments were performed under cryo-cooled conditions (–176°C) at BL41XU (λ = 1.000 Å; a Rayonix CCD detector MX225HE) and BL44XU (λ = 0.900 Å; a Bruker DIP-6040 detector) at SPring-8 (Harima, Japan) and BL17A (λ = 0.98 Å; an ADSC Quantum 270 detector) at Photon Factory (Tsukuba, Japan). A crystal mounted in a nylon loop was transferred and soaked for 1 min in the reservoir solution containing 30% (w/v) PEG 400 and then flash-frozen at –176°C in a stream of nitrogen gas. A total of 180 images were recorded with an oscillation angle of 1.0°, an exposure time of 1 s per image. The diffraction data were processed and scaled with the *HKL-2000* software package (Otwinowski and Minor, 1997). The glycerol-bound structure was determined by the molecular replacement method with *CCP4*, using the *P. falciparum* GK structure (PDB code 2W40; Schnick *et al.*, 2009) as a search model. Structural adjustments were made by iterative cycles of manual adjustments in *COOT* (Emsley *et al.*, 2010) and refinements by REFMAC5 (Murshudov *et al.*, 1997). Water molecules were automatically placed in $F_o$–$F_c$ Fourier difference maps at a 3σ cut-off level and validated to ensure correct coordination geometries using *Coot*. Data collection and refinement statistics are presented in Table 1. The final model quality was confirmed using *PROCHECK* (Laskowski *et al.*, 1993). Structural figures were prepared with *PyMol* program (DeLano, 2002).

### Mutagenesis and mutant protein purification

Site-directed mutagenesis was performed using the QuickChange method (Stratagene) with slight modifications. Briefly, the mutagenic primers (listed in Table S4) were designed using the QuickChange Primer Design Program on http://www.genomics.agilent.com/primerDesignProgram.jsp. Mutations resulting in the amino acid substitutions Y3F, T11V, T12V, D20A, R22A, R24A, T273V, T276V, R477A and E478A in TbgGK were individually introduced into a copy of Tbg

gk-pET151/D-TOPO (wild-type plasmid construct) to obtain the corresponding mutated plasmids, using the respective forward and reverse mutagenic primers. The mutant plasmids were introduced into TOP10 chemically competent *E. coli* cells and amplified. Mutations were confirmed by gene sequencing using the BigDye-terminator method with an ABI Prism310 genetic analyzer (Applied Biosystems). Mutated plasmids were used to transform *E. coli* JM109 (DE3 + pRARE2) for the mutant protein expression, using the same protocol as described above for the expression of the rTgbGK (wild type).

Mutants of TbgGK were purified on a small scale from a 5 ml culture of the respective expression hosts. After induction with 25 μM IPTG, cells were grown at 20°C for 8 h and then harvested by centrifugation. Cell pellets were disrupted (5 × 10 s) on ice in 100 μl of lysis buffer (100 mM phosphate buffer, pH 6.8, 50 mM imidazole, 300 mM NaCl, 10 mM $MgSO_4$, 0.1 mM PMSF, and 1 mg $ml^{-1}$ lysozyme) with a Handy Sonicator (TOMY SEIKO). About 100 μl of supernatant obtained after centrifugation at 20 000 *g* was transferred to 50 μl of Ni-NTA resin pre-equilibrated with the lysis buffer, and continuously mixed with a rotator (RT-50; TAITEC) for 30 min at 4°C. The resin samples were recovered as pellet by centrifugation for 3 min at 4000 *g* and washed three times for 60 s with 1.0 ml washing buffer (100 mM phosphate buffer, pH 6.8, 100 mM imidazole, 100 mM NaCl, 10 mM $MgSO_4$, 0.1 mM PMSF). Washing buffer (50 μl) supplemented with 600 mM imidazole was transferred to the pellet, incubated for 5 min with continuous mixing at 4°C, and centrifuged. The supernatants, which represent solution of each TbgGK mutant, were recovered and analyzed by SDS-PAGE and activity measurements.

### *Accession codes*

The atomic coordinates and structure factors of structures described in this project have been deposited into PDB with the following codes 3WXI, 3WXJ, 3WXK, and 3WXL for apo-, G3P-, glycerol-, and ADP-bound TbgGK respectively.

## Acknowledgments

We thank all staff members of beam lines BL32XU, BL41XU and BL44XU at SPring-8 and BL17A at Photon Factory for their help with X-ray diffraction data collection. Experiments using BL44XU were performed under the Cooperative Research Program of Institute for Protein Research, Osaka University. This work was supported by a grant from the Targeted Proteins Research Program (TPRP to KK and SH), and in part by a Creative Scientific Research Grant 18GS0314 (to KK), Grant-in-aid for Scientific Research on Priority Areas 18073004 (to KK) and 19036010 (to SH) from the Japanese Society for the Promotion of Science, and from the Japanese Ministry of Education, Science, Culture, Sports and Technology (MEXT) respectively. We also acknowledge support from the Science and Technology Research Promotion Program for Agriculture, Forestry, Fisheries and Food Industry to KK and SH, and JST/JICA, SATREPS (Science and Technology Research Partnership for Sustainable Development) (10000284) to KK. EOB is a Fellow of the Japanese Society for the Promotion of Science (JSPS).

## Conflict of interest

The authors declare that they have no conflicting interest.

## References

Balogun, E.O., Inaoka, D.K., Kido, Y., Shiba, T., Nara, T., Aoki, T., *et al.* (2010) Overproduction, purification, crystallization and preliminary X-ray diffraction analysis of *Trypanosoma brucei gambiense* glycerol kinase. *Acta Crystallograph Sect F Struct Biol Cryst Commun* **66:** 304–308.

Balogun, E.O., Inaoka, D.K., Kido, Y., Shiba, T., Nara, T., Aoki, T., *et al.* (2013) Biochemical characterization of highly active *Trypanosoma brucei gambiense* glycerol kinase, a promising drug target. *J Biochem* **154:** 77–84.

Bringaud, F., Riviere, L., and Coustou, V. (2006) Energy metabolism of trypanosomatids: adaptation to available carbon sources. *Mol Biochem Parasitol* **149:** 1–9.

Cockrell, G.M., Zheng, Y., Guo, W., Peterson, A.W., Truong, J.K., and Kantrowitz, E.R. (2013) New paradigm for allosteric regulation of *Escherichia coli* aspartate transcarbamoylase. *Biochemistry* **52:** 8036–8047.

Colasante, C., Ellis, M., Ruppert, T., and Voncken, F. (2006) Comparative proteomics of glycosomes from bloodstream form and procyclic culture form *Trypanosoma brucei brucei*. *Proteomics* **6:** 3275–3293.

DeLano, W.L. (2002) Unraveling hot spots in binding interfaces: progress and challenges. *Curr Opin Struct Biol* **12:** 14–20.

Dipple, K.M., Zhang, Y.H., Huang, B.L., McCabe, L.L., Dallongeville, J., Inokuchi, T., *et al.* (2001) Glycerol kinase deficiency: evidence for complexity in a single gene disorder. *Hum Genet* **109:** 55–62.

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

Feese, M., Pettigrew, D.W., Meadow, N.D., Roseman, S., and Remington, S.J. (1994) Cation-promoted association of a regulatory and target protein is controlled by protein phosphorylation. *Proc Natl Acad Sci USA* **91:** 3544–3548.

Gualdron-Lopez, M., Brennand, A., Avilan, L., and Michels, P.A. (2013) Translocation of solutes and proteins across the glycosomal membrane of trypanosomes; possibilities and limitations for targeting with trypanocidal drugs. *Parasitology* **140:** 1–20.

Guerra-Giraldez, C., Quijada, L., and Clayton, C.E. (2002) Compartmentation of enzymes in a microbody, the glycosome, is essential in *Trypanosoma brucei*. *J Cell Sci* **115:** 2651–2658.

Haanstra, J.R., van Tuijl, A., Kessler, P., Reijnders, W., Michels, P.A., Westerhoff, H.V., *et al.* (2008) Compartmentation prevents a lethal turbo-explosion of glycolysis in trypanosomes. *Proc Natl Acad Sci USA* **105:** 17718–17723.

Hannaert, V., Bringaud, F., Opperdoes, F.R., and Michels, P.A. (2003) Evolution of energy metabolism and its compartmentation in Kinetoplastida. *Kinetoplastid Biol Dis* **2:** 11.

Holtman, C.K., Pawlyk, A.C., Meadow, N., Roseman, S., and Pettigrew, D.W. (2001) IIA(Glc) allosteric control of *Escherichia coli* glycerol kinase: binding site cooperative

transitions and cation-promoted association by Zinc(II). *Biochemistry* **40:** 14302–14308.

Hurley, J.H., Faber, H.R., Worthylake, D., Meadow, N.D., Roseman, S., Pettigrew, D.W., *et al.* (1993) Structure of the regulatory complex of *Escherichia coli* IIIGlc with glycerol kinase. *Science* **259:** 673–677.

Imada, K., Tamura, T., and Inagaki, K. (2006) Structure of glycerol kinase from *Cellulomonas* sp. NT3060 [WWW document]. URL http://www.rcsb.org/pdb/explore/explore.do?structureId=2d4w: Protein Data Base.

Johnson, L.N., and Barford, D. (1993) The effects of phosphorylation on the structure and function of proteins. *Annu Rev Biophys Biomol Struct* **22:** 199–232.

Koga, Y., Katsumi, R., You, D.J., Matsumura, H., Takano, K., and Kanaya, S. (2008) Crystal structure of highly thermostable glycerol kinase from a hyperthermophilic archaeon in a dimeric form. *FEBS J* **275:** 2632–2643.

Kralova, I., Rigden, D.J., Opperdoes, F.R., and Michels, P.A. (2000) Glycerol kinase of *Trypanosoma brucei*: cloning, molecular characterization and mutagenesis. *Eur J Biochem* **267:** 2323–2333.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227:** 680–685.

Lahiri, S.D., Zhang, G., Dunaway-Mariano, D., and Allen, K.N. (2003) The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction. *Science* **299:** 2067–2071.

Laskowski, R.A., Moss, D.S., and Thornton, J.M. (1993) Main-chain bond lengths and bond angles in protein structures. *J Mol Biol* **231:** 1049–1067.

Le Febvre, R.B., and Hill, G.C. (1986) *Trypanosoma rhodesiense*: mitochondrial proteins of bloodstream and procyclic trypomastigotes. *Exp Parasitol* **62:** 85–91.

Li, X., Lu, Y., Jin, W., Liang, K., Mills, G.B., and Fan, Z. (2006) Autophosphorylation of Akt at threonine 72 and serine 246: a potential mechanism of regulation of Akt kinase activity. *J Biol Chem* **281:** 13837–13843.

Llinas, P., Stura, E.A., Menez, A., Kiss, Z., Stigbrand, T., Millan, J.L., and Le Du, M.H. (2005) Structural studies of human placental alkaline phosphatase in complex with functional ligands. *J Mol Biol* **350:** 441–451.

Marshall, P.B. (1948) The glucose metabolism of *Trypanosoma evansi* and the action of trypanoscides. *Br J Pharmacol* **3:** 8–14.

Migchelsen, S.J., Buscher, P., Hoepelman, A.I., Schallig, H.D., and Adams, E.R. (2011) Human African trypanosomiasis: a review of non-endemic cases in the past 20 years. *Int J Infect Dis* **15:** e517–524.

Minagawa, N., Yabu, Y., Kita, K., Nagai, K., Ohta, N., Meguro, K., *et al.* (1997) An antibiotic, ascofuranone, specifically inhibits respiration and *in vitro* growth of long slender bloodstream forms of *Trypanosoma brucei brucei*. *Mol Biochem Parasitol* **84:** 271–280.

Minasov, G., Brunzelle, J., Skarina, T., Onopriyenko, O., Peterson, S.N., Savchenko, A., *et al.* (2009) 2.7 Angstrom Crystal Structure of Glycerol Kinase (glpK) from *Staphylococcus aureus* in Complex with ADP and Glycerol [WWW document]. URL http://www.rcsb.org/pdb/explore/explore.do?pdbId=3GE1: Protein Data Bank.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53:** 240–255.

Nolan, D.P., and Voorheis, H.P. (2000) Hydrogen ion gradients across the mitochondrial, endosomal and plasma membranes in bloodstream forms of *Trypanosoma brucei* solving the three-compartment problem. *Eur J Biochem* **267:** 4601–4614.

Ohashi-Suzuki, M., Yabu, Y., Ohshima, S., Nakamura, K., Kido, Y., Sakamoto, K., *et al.* (2011) Differential kinetic activities of glycerol kinase among African trypanosome species: phylogenetic and therapeutic implications. *J Vet Med Sci* **73:** 615–621.

Opperdoes, F.R., and Borst, P. (1977) Localization of nine glycolytic enzymes in a microbody-like organelle in *Trypanosoma brucei*: the glycosome. *FEBS Lett* **80:** 360–364.

Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276:** 307–326.

Pettigrew, D.W., Yu, G.J., and Liu, Y. (1990) Nucleotide regulation of *Escherichia coli* glycerol kinase: initial-velocity and substrate binding studies. *Biochemistry* **29:** 8620–8627.

Rahib, L., MacLennan, N.K., Horvath, S., Liao, J.C., and Dipple, K.M. (2007) Glycerol kinase deficiency alters expression of genes involved in lipid metabolism, carbohydrate metabolism, and insulin signaling. *Eur J Hum Genet* **15:** 646–657.

Saimoto, H., Kido, Y., Haga, Y., Sakamoto, K., and Kita, K. (2013) Pharmacophore identification of ascofuranone, potent inhibitor of cyanide-insensitive alternative oxidase of *Trypanosoma brucei*. *J Biochem* **153:** 267–273.

Schlichting, I., and Reinstein, J. (1997) Structures of active conformations of UMP kinase from *Dictyostelium discoideum* suggest phosphoryl transfer is associative. *Biochemistry* **36:** 9290–9296.

Schnick, C., Polley, S.D., Fivelman, Q.L., Ranford-Cartwright, L.C., Wilkinson, S.R., Brannigan, J.A., *et al.* (2009) Structure and non-essential function of glycerol kinase in *Plasmodium falciparum* blood stages. *Mol Microbiol* **71:** 533–545.

Shiba, T., Kido, Y., Sakamoto, K., Inaoka, D.K., Tsuge, C., Tatsumi, R., *et al.* (2013) Structure of the trypanosome cyanide-insensitive alternative oxidase. *Proc Natl Acad Sci USA* **110:** 4580–4585.

Simpson, A.G., Stevens, J.R., and Lukes, J. (2006) The evolution and diversity of kinetoplastid flagellates. *Trends Parasitol* **22:** 168–174.

Stijlemans, B., Caljon, G., Natesan, S.K., Saerens, D., Conrath, K., Perez-Morga, D., *et al.* (2011) High affinity nanobodies against the *Trypanosome brucei* VSG are potent trypanolytic agents that block endocytosis. *PLoS Pathog* **7:** e1002072.

Verlinde, C.L., Hannaert, V., Blonski, C., Willson, M., Perie, J.J., Fothergill-Gilmore, L.A., *et al.* (2001) Glycolysis as a target for the design of new anti-trypanosome drugs. *Drug Resist Updat* **4:** 50–65.

Yeh, J.I., Charrier, V., Paulo, J., Hou, L., Darbon, E., Claiborne, A., *et al.* (2004) Structures of enterococcal glycerol kinase in the absence and presence of glycerol:

correlation of conformation to substrate binding and a mechanism of activation by phosphorylation. *Biochemistry* **43:** 362–373.

Yu, P., and Pettigrew, D.W. (2003) Linkage between fructose 1,6-bisphosphate binding and the dimer-tetramer equilibrium of *Escherichia coli* glycerol kinase: critical behavior arising from change of ligand stoichiometry. *Biochemistry* **42:** 4243–4252.

Yu, P., Lasagna, M., Pawlyk, A.C., Reinhart, G.D., and Pettigrew, D.W. (2007) IIAGlc inhibition of glycerol kinase: a communications network tunes protein motions at the allosteric site. *Biochemistry* **46:** 12355–12365.

Zorba, A., Buosi, V., Kutter, S., Kern, N., Pontiggia, F., Cho, Y.J., and Kern, D. (2014) Molecular mechanism of Aurora A kinase autophosphorylation and its allosteric activation by TPX2. *eLife* **3:** e02667.

Zwaig, N., and Lin, E.C. (1966) Feedback inhibition of glycerol kinase, a catabolic enzyme in *Escherichia coli*. *Science* **153:** 755–757.

## Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site.

**Supporting Information**

# Molecular basis for the reverse reaction of African human trypanosomes glycerol kinase

Emmanuel Oluwadare Balogun, Daniel Ken Inaoka, Tomoo Shiba, Yasutoshi Kido, Chiaki Tsuge, Takeshi Nara, Takashi Aoki, Teruki Honma, Akiko Tanaka, Masayuki Inoue, Shigeru Matsuoka, Paul A.M. Michels, Kiyoshi Kita, Shigeharu Harada

## Supplemental Figure Legends

**Figure S1** Glycolysis in BSFs of African human trypanosome. Red labels indicate various cell compartments involved in energy metabolism. 1, hexokinase; 2, glucose-6-phosphate isomerase; 3, phosphofructokinase; 4, aldolase; 5, triose-phosphate isomerase; 6, glyceraldehyde-3-phosphate dehydrogenase; 7, phosphoglycerate kinase; 8, NAD-dependent glycerol-3-phosphate dehydrogenase; **9, glycerol kinase;** 10, phosphoglycerate mutase; 11, enolase; 12, pyruvate kinase; step 13, glycerol 3-phosphate oxidation (involving FAD-dependent glycerol-3-phosphate dehydrogenase, ubiquinone (Q), and the trypanosome alternative oxidase, TAO). Gal 3-P, glyceraldehyde 3-phosphate; 1,3-BPGA, 1,3-bisphosphoglycerate; 3-PGA, 3-phosphoglycerate; DHAP, dihydroxyacetone phosphate; G 3-P, glycerol 3-phosphate; PEP, phosphoenolpyruvate. Since there is no Krebs cycle, and the BSF trypanosomes also lack the gene of lactate dehydrogenase (Berriman *et al.*, 2005; van Weelden *et al.*, 2005), the accumulated pyruvate molecules are excreted from the parasites into the host circulatory system (Ryley, 1956).

**Figure S2** Detailed secondary structure arrangements in domains I and II. (A) Monomer colored by domains as shown. (B) and (C) are the respective domains with all the secondary structure elements labeled. N' and C' represent interdomain connection points, while N and C represent the amino- and carboxy-terminus, respectively. For the purpose of clarity, both domains were manually adjusted (a 180° horizontal flip followed by a 90° vertical rotation) to have the best view of all secondary structures.

**Figure S3** Protein sequence alignment. Sequence similarities between GKs from *T. b. gambiense*, *P. falciparum*, human, *E. coli* and *A. thaliana*. This figure was created with the Jalview Java alignment editor (Clamp *et al.*, 2004). The secondary structure elements in the TbgGK structure are shown by orange bars (α-helices) and green arrows (β-strands).

Blue-shaded amino-acid residues indicate conserved residues. Substrates-binding residues are indicated as dots (green, ADP; cyan, catalytic glycerol; black, ADP and catalytic glycerol; red, non-catalytic glycerol; blue, ADP and non-catalytic glycerol).

**Figure S4** Various conformational changes at the active site of TbgGK (surface representation) induced by substrate binding. (A) Unligated form. (B) G3P-TbgGK. (C) Glycerol-TbgGK. (D) ADP-TbgGK. Creation of the secondary catalytic site (SCS) after dephosphorylation of G3P is necessary for ADP binding. The surface model is colored by atom (magenta, blue, red, and yellow represent carbon, nitrogen, oxygen, and sulfur atoms, respectively.

**Figure S5** Conformational change in the active cleft of chain-B induced by $Mg{\cdot}ADP_2$ binding. The surface model of the structure reveals the closure of the active cleft (indicated by the arrow) as a result of the $Mg{\cdot}ADP_2$ binding to the enzyme surface rendering the active site inaccessible to the substrate. The colored sticks represent ADP molecules, while the green sphere represents the magnesium ion.

**Figure S6** Surface model for the structure of the active cleft in the presence of glycerol (colored sticks). Two glycerol molecules are bound; one in the catalytic site, labeled as cg and the other in the non-catalytic site, as ncg. The ncg weakly occupies a position that accommodates the adenosine head of the ADP substrate; it is displaced in the presence of ADP.

Supplemental references

Berriman, M., Ghedin, E., Hertz-Fowler, C.,Blandin, G.,Renauld, H., Bartholomeu, D.C., *et al.*, (2005) The genome of the African trypanosome *Trypanosoma brucei*. *Science*

**309**: 416-422.

Clamp, M., Cuff, J., Searle S.M. and Barton, G.J. (2004) The Jalview Java alignment editor. *Bioinformatics* **20**: 426-427.

Ryley, J.F., (1956) Studies on the metabolism of the Protozoa. 7. Comparative carbohydrate metabolism of eleven species of trypanosome. *Biochem J* **62**: 215-222.

van Weelden, S.W., van Hellemond, J.J., Opperdoes, F.R. and Tielens, A.G. (2005) New functions for parts of the Krebs cycle in procyclic *Trypanosoma brucei*, a cycle not operating as a cycle. *J Biol Chem* **280**: 12451-12460.

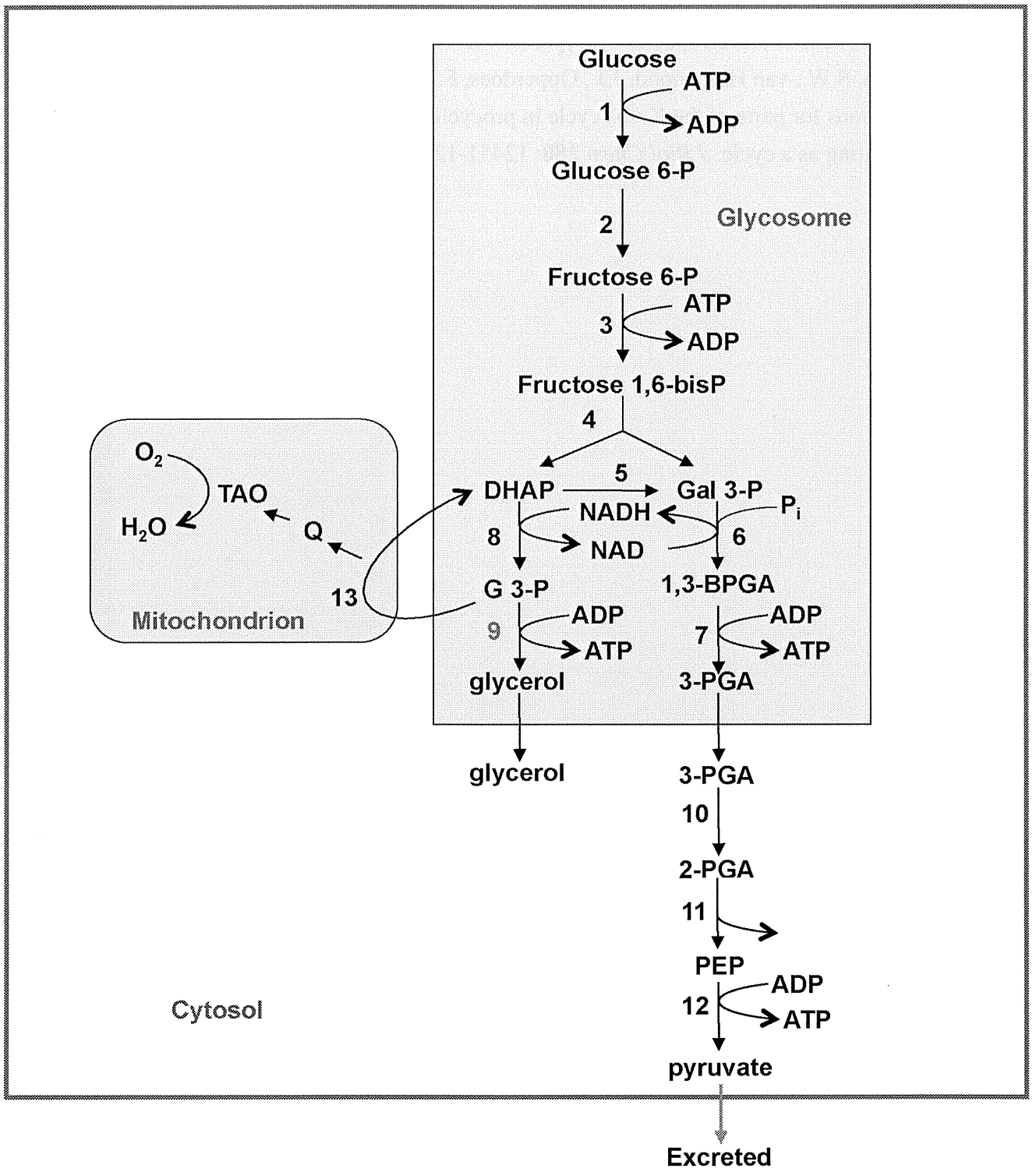
Glucose
ATP
1
ADP
Glucose 6-P
2
Glycosome
Fructose 6-P
ATP
3
ADP
Fructose 1,6-bisP
4
5
DHAP
Gal 3-P
NADH
$P_i$
8
NAD
6
G 3-P
1,3-BPGA
ADP
9
ATP
ADP
7
ATP
glycerol
3-PGA
$O_2$
TAO
$H_2O$
Q
13
Mitochondrion
glycerol
3-PGA
10
2-PGA
11
PEP
ADP
12
ATP
pyruvate
Cytosol
Excreted

Figure S1

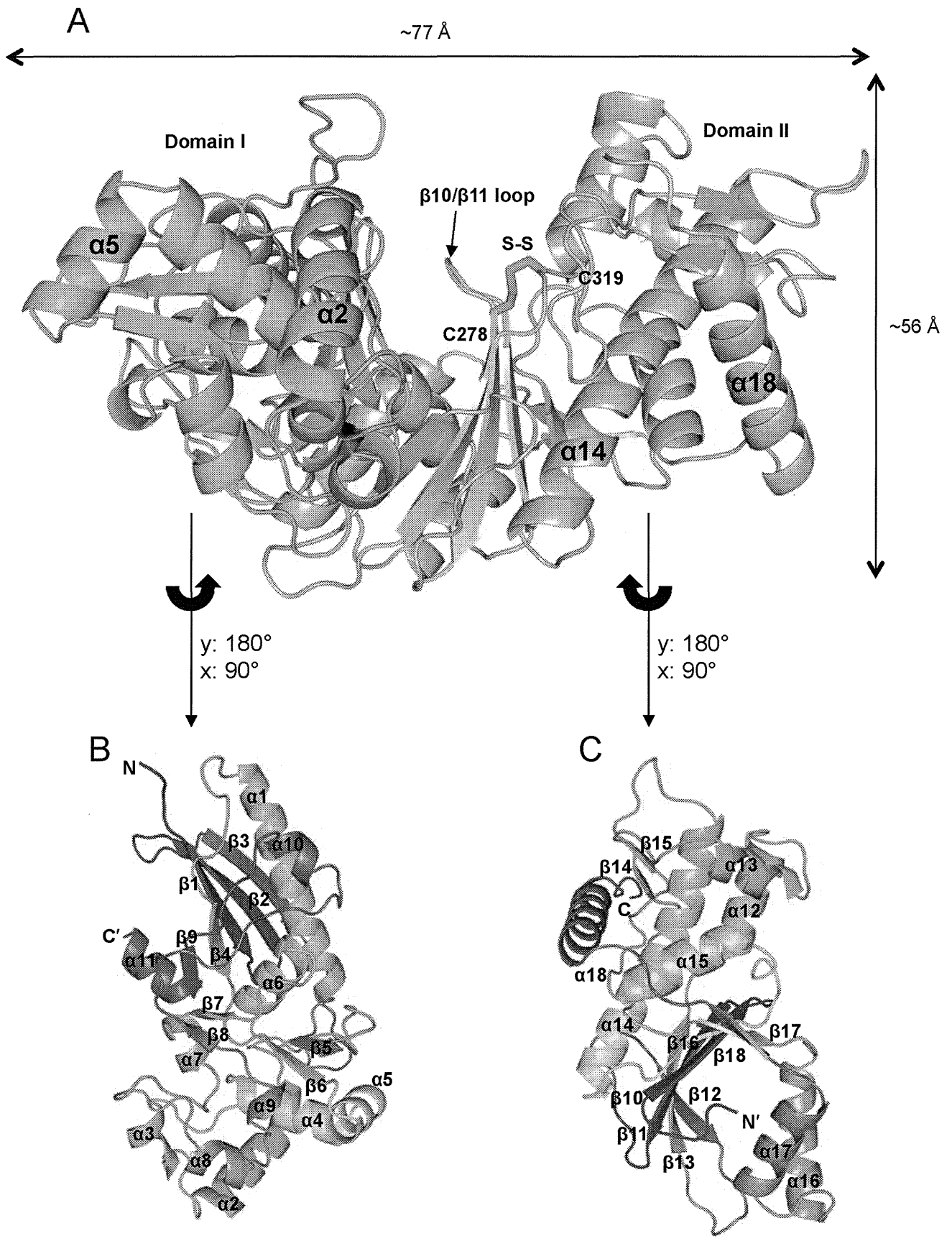
A
~77 Å
Domain I
Domain II
β10/β11 loop
S-S
C319
C278
α5
α2
α18
α14
~56 Å
y: 180°
x: 90°
y: 180°
x: 90°
B
N
α1
β3
α10
β1
β2
C′
β9
α11
β4
α6
β7
β8
β5
α7
β6
α5
α9
α4
α3
α8
α2
C
β15
β14
α13
C
α12
α15
α18
α14
β16
β17
β18
β10
β12
N′
β11
α17
β13
α16

Figure S2

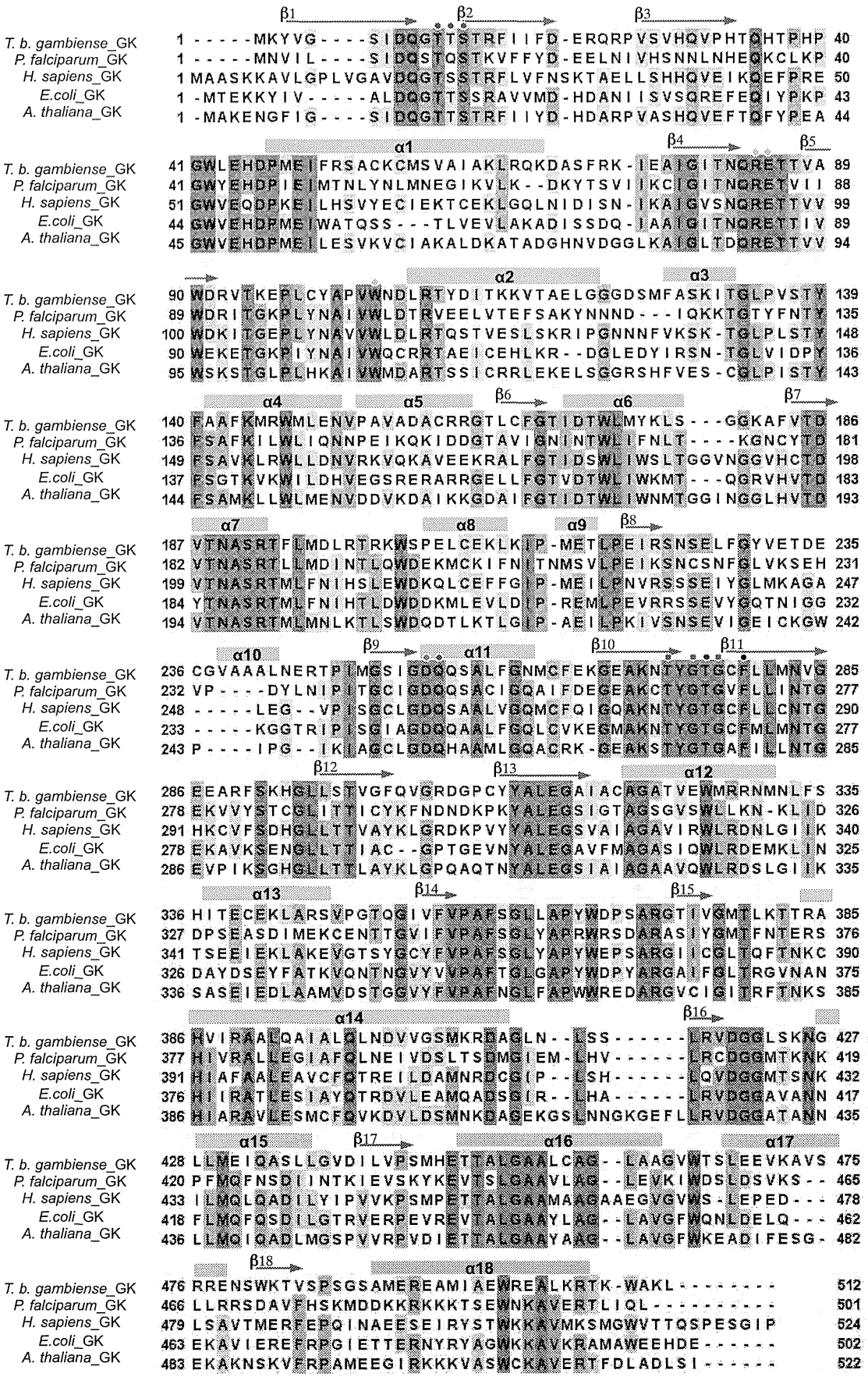
T. b. gambiense_GK
P. falciparum_GK
H. sapiens_GK
E.coli_GK
A. thaliana_GK
β1 β2 β3
1 -----MKYVG----SIDQGTTSTRFIIFD-ERQRPVSVHQVPHTQHTPHP 40
1 -----MNVIL----SIDQSTQSTKVFFYD-EELNIVHSNNLNHEQKCLKP 40
1 MAASKKAVLGPLVGAVDQGTSSTRFLVFNSKTAELLSHHQVEIKQEFPRE 50
1 -MTEKKYIV-----ALDQGTTSSRAVVMD-HDANIISVSQREFEQIYPKP 43
1 -MAKENGFIG----SIDQGTTSTRFIIYD-HDARPVASHQVEFTQFYPEA 44
α1 β4 β5
41 GWLEHDPMEIFRSACKCMSVAIAKLRQKDASFRK-IEAIGITNQRETTVA 89
41 GWYEHDPIEIMTNLYNLMNEGIKVLK--DKYTSVIIKCIGITNQRETVII 88
51 GWVEQDPKEILHSVYECIEKTCEKLGQLNIDISN-IKAIGVSNQRETTVV 99
44 GWVEHDPMEIWATQSS---TLVEVLAKADISSDQ-IAAIGITNQRETTIV 89
45 GWVEHDPMEILESVKVCIAKALDKATADGHNVDGGLKAIGLTDQRETTVV 94
α2 α3
90 WDRVTKEPLCYAPVWNDLRTYDITKKVTAELGGGDSMFASKITGLPVSTY 139
89 WDRITGKPLYNAIVWLDTRVEELVTEFSAKYNNND---IQKKTGTYFNTY 135
100 WDKITGEPLYNAVVWLDLRTQSTVESLSKRIPGNNNFVKSK-TGLPLSTY 148
90 WEKETGKPIYNAIVWQCRRTAEICEHLKR--DGLEDYIRSN-TGLVIDPY 136
95 WSKSTGLPLHKAIVWMDARTSSICRRLEKELSGGRSHFVES-CGLPISTY 143
α4 α5 β6 α6 β7
140 FAAFKMRWMLENVPAVADACRRGTLCFGTIDTWLMYKLS---GGKAFVTD 186
136 FSAFKILWLIQNNPEIKQKIDDGTAVIGNINTWLIFNLT----KGNCYTD 181
149 FSAVKLRWLLDNVRKVQKAVEEKRALFGTIDSWLIWSLTGGVNGGVHCTD 198
137 FSGTKVKWILDHVEGSRERARRGELLFGTVDTWLIWKMT---QGRVHVTD 183
144 FSAMKLLWLMENVDDVKDAIKKGDAIFGTIDTWLIWNMTGGINGGLHVTD 193
α7 α8 α9 β8
187 VTNASRTFLMDLRTRKWSPELCEKLKIP-METLPEIRSNSELFGYVETDE 235
182 VTNASRTLLMDINTLQWDEKMCKIFNITNMSVLPEIKSNCSNFGLVKSEH 231
199 VTNASRTMLFNIHSLEWDKQLCEFFGIP-MEILPNVRSSSEIYGLMKAGA 247
184 YTNASRTMLFNIHTLDWDDKMLEVLDIP-REMLPEVRRSSEVYGQTNIGG 232
194 VTNASRTMLMNLKTLSWDQDTLKTLGIP-AEILPKIVSNSEVIGEICKGW 242
α10 β9 α11 β10 β11
236 CGVAAALNERTPIMGSIGDQQSALFGNMCFEKGEAKNTYGTGCFLLMNVG 285
232 VP----DYLNIPITGCIGDQQSACIGQAIFDEGEAKCTYGTGVFLLINTG 277
248 -----LEG--VPISGCLGDQSAALVGQMCFQIGQAKNTYGTGCFLLCNTG 290
233 -----KGGTRIPISGIAGDQQAALFGQLCVKEGMAKNTYGTGCFMLMNTG 277
243 P----IPG--IKIAGCLGDQHAAMLGQACRK-GEAKSTYGTGAFILLNTG 285
β12 β13 α12
286 EEARFSKHGLLSTVGFQVGRDGPCYYALEGAIACAGATVEWMRRNMNLFS 335
278 EKVVYSTCGLITTICYKFNDNDKPKYALEGSIGTAGSGVSWLLKN-KLID 326
291 HKCVFSDHGLLTTVAYKLGRDKPVYYALEGSVAIAGAVIRWLRDNLGIIK 340
278 EKAVKSENGLLTTIAC--GPTGEVNYALEGAVFMAGASIQWLRDEMKLIN 325
286 EVPIKSGHGLLTTLAYKLGPQAQTNYALEGSIAIAGAAVQWLRDSLGIIK 335
α13 β14 β15
336 HITECEKLARSVPGTQGIVFVPAFSGLLAPYWDPSARGTIVGMTLKTTRA 385
327 DPSEASDIMEKCENTTGVIFVPAFSGLYAPRWRSDARASIYGMTFNTERS 376
341 TSEEIEKLAKEVGTSYGCYFVPAFSGLYAPYWEPSARGIICGLTQFTNKC 390
326 DAYDSEYFATKVQNTNGVYVVPAFTGLGAPYWDPYARGAIFGLTRGVNAN 375
336 SASEIEDLAAMVDSTGGVYFVPAFNGLFAPWWREDARGVCIGITRFTNKS 385
α14 β16
386 HVIRAALQAIALQLNDVVGSMKRDAGLN--LSS------LRVDGGLSKNG 427
377 HIVRALLEGIAFQLNEIVDSLTSDMGIEM-LHV------LRCDGGMTKNK 419
391 HIAFAALEAVCFQTREILDAMNRDCGIP--LSH------LQVDGGMTSNK 432
376 HIIRATLESIAYQTRDVLEAMQADSGIR--LHA------LRVDGGAVANN 417
386 HIARAVLESMCFQVKDVLDSMNKDAGEKGS-LNNGKGEFLLRVDGGATANN 435
α15 β17 α16 α17
428 LLMEIQASLLGVDILVPSMHETTALGAALCAG--LAAGVWTSLEEVKAVS 475
420 PFMQFNSDIINTKIEVSKYKEVTSLGAAVLAG--LEVKIWDSLDSVKS-- 465
433 ILMQLQADILYIPVVKPSMPETTALGAAMAAGAAEGVGVWS-LEPED--- 478
418 FLMQFQSDILGTRVERPEVREVTALGAAYLAG--LAVGFWQNLDELQ--- 462
436 LLMQIQADLMGSPVVRPVDIETTALGAAYAAG--LAVGFWKEADIFESG- 482
β18 α18
476 RRENSWKTVSPSGSAMEREAMIAEWREALKRTK-WAKL-------- 512
466 LLRRSDAVFHSKMDDKKRKKKTSEWNKAVERTLIQL---------- 501
479 LSAVTMERFEPQINAEESEIRYSTWKKAVMKSMGWVTTQSPESGIP 524
463 EKAVIEREFRPGIETTERNYRYAGWKKAVKRAMAWEEHDE------ 502
483 EKAKNSKVFRPAMEEGIRKKKVASWCKAVERTFDLADLSI------ 522

Figure S3

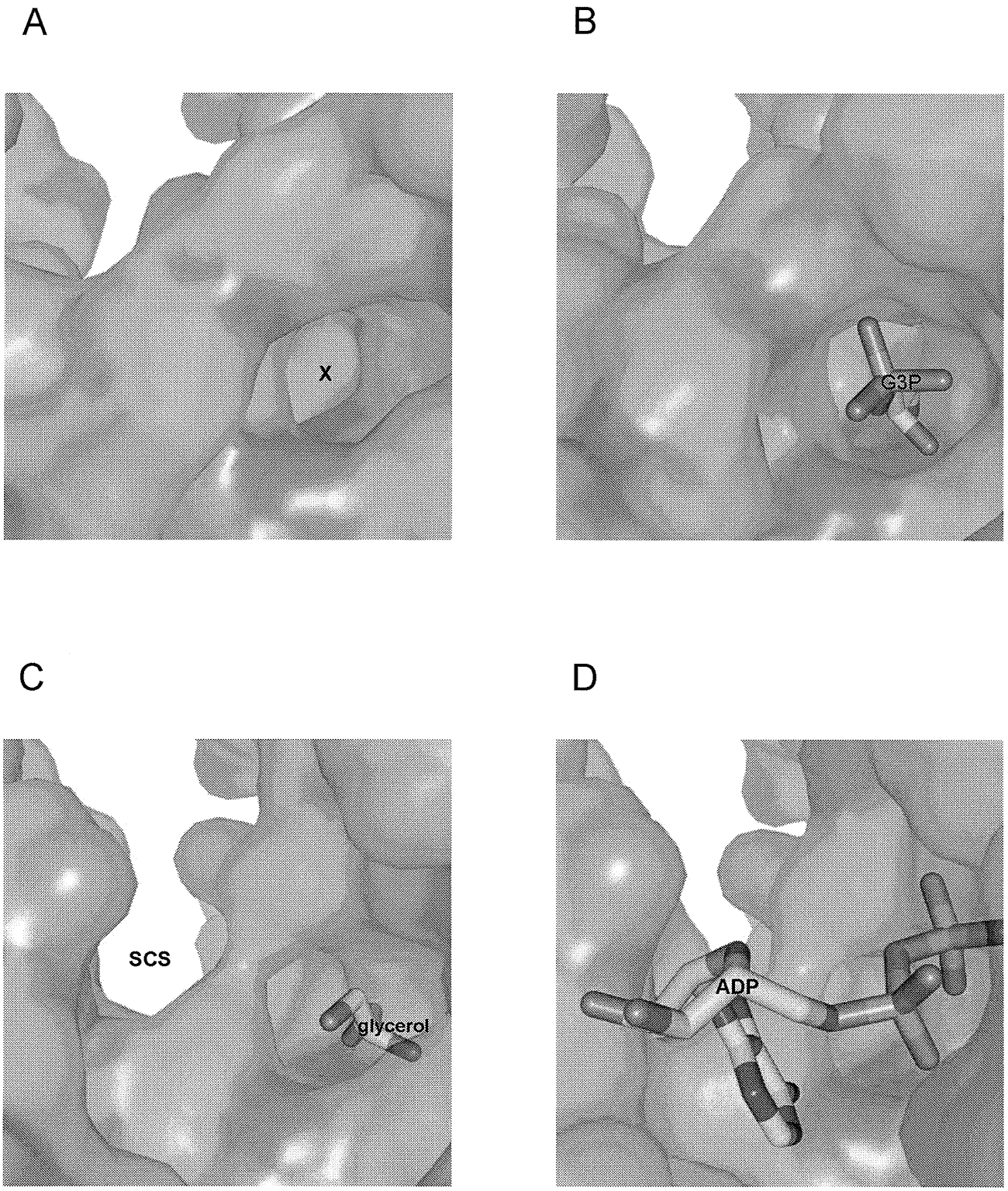


Figure S4

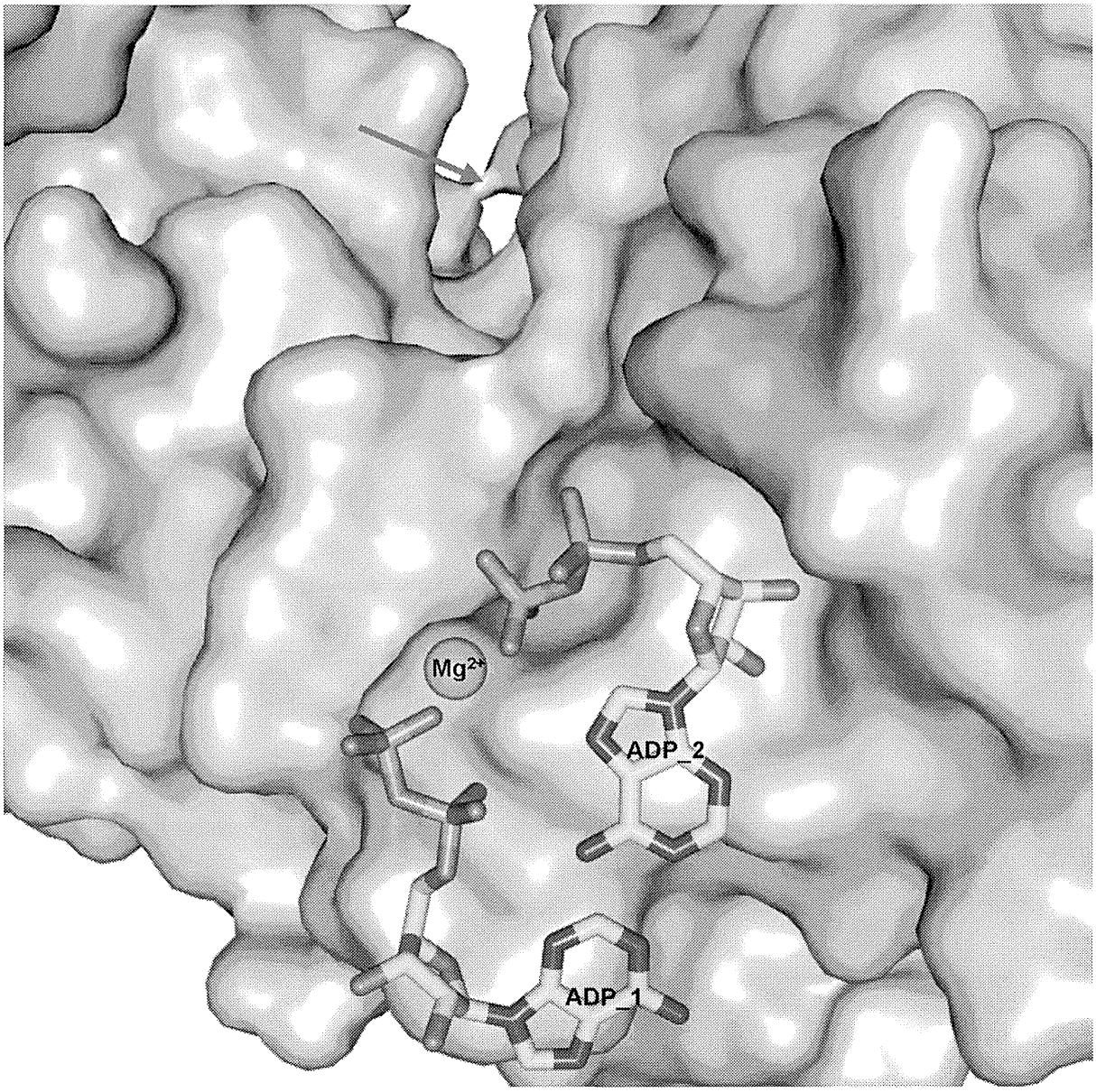


Figure S5