

The Molecular Dynamics of *Trypanosoma brucei* UDP-Galactose 4'-Epimerase: A Drug Target for African Sleeping Sickness

Aaron J. Friedman^{1*}, Jacob D. Durrant²,
Levi C. T. Pierce², Thomas J. McCorvie³,
David J. Timson³ and J. Andrew
McCammon^{4,5,6}

¹Biomedical Sciences Graduate Program, University of California San Diego, La Jolla, CA 92093-0365, USA

²Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093-0365, USA

³School of Biological Sciences, Queen's University Belfast, Medical Biology Centre, Belfast BT9 7BL, UK

⁴Department of Chemistry and Biochemistry, NSF Center for Theoretical Biological Physics, National Biomedical Computation Resource, University of California San Diego, La Jolla, CA 92093, USA

⁵Department of Pharmacology, University of California San Diego, La Jolla, CA 92093, USA

⁶Howard Hughes Medical Institute, University of California San Diego, La Jolla, CA 92093, USA

*Corresponding author: Aaron Friedman, a1friedm@ucsd.edu

During the past century, several epidemics of human African trypanosomiasis, a deadly disease caused by the protist *Trypanosoma brucei*, have afflicted sub-Saharan Africa. Over 10 000 new victims are reported each year, with hundreds of thousands more at risk. As current drug treatments are either highly toxic or ineffective, novel trypanocides are urgently needed. The *T. brucei* galactose synthesis pathway is one potential therapeutic target. Although galactose is essential for *T. brucei* survival, the parasite lacks the transporters required to intake galactose from the environment. UDP-galactose 4'-epimerase (*TbGalE*) is responsible for the epimerization of UDP-glucose to UDP-galactose and is therefore of great interest to medicinal chemists. Using molecular dynamics simulations, we investigate the atomistic motions of *TbGalE* in both the *apo* and *holo* states. The sampled conformations and protein dynamics depend not only on the presence of a UDP-sugar ligand, but also on the chirality of the UDP-sugar C4 atom. This dependence provides important insights into *TbGalE* function and may help guide future computer-aided drug discovery efforts targeting this protein.

Key words: African sleeping sickness, molecular dynamics, protein structure, *TbGalE*, *Trypanosoma brucei*, UDP-Galactose-4'-Epimerase

Abbreviations: *TbGalE*, *Trypanosoma brucei* UDP-galactose-4'-epimerase; HAT, Human African trypanosomiasis; MD, molecular dynamics; RMSD, root-mean-square deviation; PCA, principal component analysis; *HsJGDH*, human UDP-glucose dehydrogenase.

Received 24 January 2012, revised 7 March 2012 and accepted for publication 1 April 2012

Human African trypanosomiasis (HAT), a disease caused by the protist *Trypanosoma brucei* (*T. brucei*), directly affects thousands of sub-Saharan Africans and indirectly places hundreds of thousands more at risk. Current treatments are often ineffective or dangerous, necessitating the development of a new generation of HAT therapeutics.

Disease pathology occurs in two stages. The early hemolymphatic stage is characterized by fever, cephalgia, arthralgia, and pruritus. Current treatments for early-stage HAT include pentamine and suramin. However, pentamine, often used to treat the *T. brucei gambiense* strain, is associated with hypoglycemia and hypotension, and suramin, effective against *T. brucei rhodesiense*, is associated with severe cutaneous reactions, renal failure, anaphylactic shock, and neurotoxicity (1–4).

Once the parasite crosses the blood–brain barrier, the more serious and often irreversible symptoms of the neurological phase are manifest, including disrupted cognition, coordination, and sleep (5). Since 1949, melarsoprol has been the standard treatment for late-stage HAT. Despite its efficacy, a significant number of patients relapse; additionally, 5–10% of those who receive treatment develop severe encephalopathy, often leading to death (6). In the late 20th century, Aventis developed eflornithine, a drug that targets *T. brucei* ornithine decarboxylase. While eflornithine is significantly less dangerous than melarsoprol (7), it is ineffective against the *T. brucei rhodesiense* subspecies (8).

As current therapeutics are problematic, medicinal chemists are actively seeking to identify novel *T. brucei* drug targets. The proteins of the biochemical pathway involved in galactose synthesis are excellent candidates. Although *T. brucei* requires galactose for the synthesis of vital glycoproteins (9), it is unable to intake galactose from the environment. Instead, glucose is acquired via hexose transporters (10) and is subsequently converted to galactose. One of the proteins in the pathway that effectuates this conversion, UDP-galactose 4'-epimerase (*TbGalE*), inverts the stereochemistry of

the UDP-glucose C4 carbon atom to produce UDP-galactose. The two-step reaction proceeds via a transient ketose intermediate and requires rotation of the sugar portion of the ligand for epimerization (Figure 1) (11). Initial drug discovery efforts targeting this protein have identified several promising inhibitors (12). Unexpectedly, virtual screening and subsequent experimental validation identified several agonists as well.

This study provides insight into the general mechanisms of *TbGalE* ligand binding. The static lock-and-key model of binding, first proposed by Emil Fischer in 1894 (13), has been largely abandoned in favor of theories that account for receptor flexibility. Specifically, the induced-fit and population-shift models have been much debated over the course of the past 50 years (14). Induced fit suggests that *apo* and *holo* receptors sample distinct regions of conformational space; ligand binding itself induces conformational changes in the receptor (15). Population shift, on the other hand, suggests that an *apo* protein samples many conformations in solution, a subset of which are amenable to ligand binding. Ligands bind to rare but amenable receptor conformations and energetically stabilize them, causing the population of all conformations to shift toward those that can accommodate the ligand (16–19).

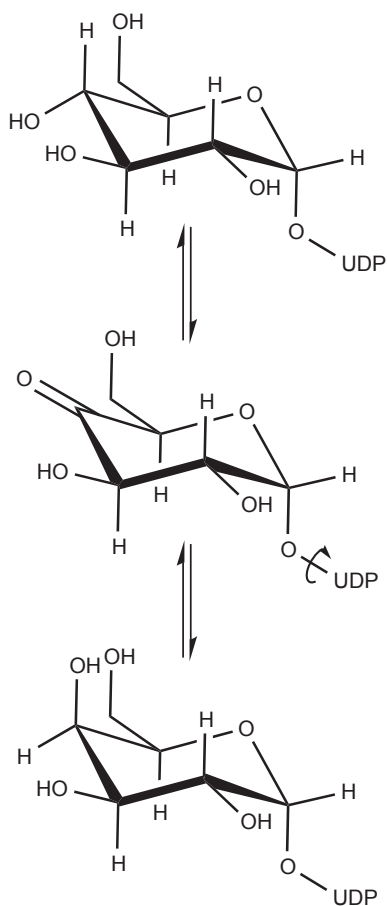


Figure 1: UDP-sugar epimerization. The epimerization reaction proceeds via a UDP-ketose intermediate. Epimerization requires a 180° rotation ('flip') of the sugar portion of the ligand.

Here, we explore the atomistic dynamics of the *TbGalE* homodimer by investigating the major steps along its reaction coordinate. Using molecular dynamics (MD) simulations, we study the motions of *TbGalE* homodimer in both the *apo* and *holo* forms, bound to UDP-galactose, UDP-glucose, and the UDP-ketose intermediate. The conformations sampled by the protein depend not only on the presence of a UDP-sugar ligand, but also on the chirality of the UDP-sugar C4 atom. This dependence provides important insight into *TbGalE* function and may help guide future computer-aided drug discovery efforts targeting this protein.

Methods and Materials

System preparation

A crystal structure of *TbGalE* homodimer [PDB ID: 2CNB (20)] was obtained from the Protein Data Bank (21). To generate missing loops, the structure was aligned to a model of *TbGalE* that had been created previously (12) based on the 1GY8 *TbGalE* structure (11). Following alignment, the coordinates of the missing loop atoms were copied from the 1GY8 model. All cocrystallized waters were retained; PDB2PQR (22) and PROPKA (23,24) were used to assign residue protonation states at pH 7.0. Histidine protonation states were visually inspected in VMD (25) to ensure optimal hydrogen bonding. The NAD(H) and Y173 protonation states were manually assigned to properly mimic the active site configuration needed for catalysis (11).

NAD⁺/NADH parameters were obtained from Walker *et al.* (26,27). Accelerlys Discovery Studio 2.5 was used to model the structure of UDP-galactose by changing the fluorine atom of the 2CNB UDP-4-deoxy-4-fluoro- α -D-galactose ligand to a hydroxyl group. UDP-glucose and the UDP-ketose intermediate were built by altering the stereochemistry and hybridization of the UDP-galactose C4 carbon atom. Hydrogen atoms were added to the three UDP-sugars using Discovery Studio. All ligand partial charges were generated using GAUSSIAN03,^a and ligand atoms were parameterized according to the GAFF force field (28).

Receptor atomic parameters and partial charges were assigned according to the Amber ff99SB force field (29) using the AMBER 10 *xleap* module.^b Sodium ions were subsequently added to bring the system to electric neutrality. The protein was then solvated in a TIP3P (30) water box that extended 10 Å beyond the protein in each direction, and additional sodium and chloride ions were added to bring the total salt concentration to 20 mM.

Molecular dynamics simulations

NAMD 2.6 (31) was used for all minimizations, equilibrations, and free-dynamics runs. Minimization and equilibration steps were performed as described previously by Wang *et al.* (28). In brief, each system was minimized in four phases totaling 45 000 minimization steps. Hydrogen atoms were relaxed in the first 5000 steps; hydrogen atoms and water molecules were relaxed in the next 5000 minimization steps; hydrogen atoms, water molecules, and the atoms of the protein backbone were relaxed in the next 10 000 minimization steps; and all atoms were relaxed for the last 25 000 minimization steps.

For equilibration, 1-ns simulations were performed at 310K using the final minimized structures as the initial coordinates. Harmonic constraints were placed on the atoms of the protein backbones and relaxed in a series of four 250-ps steps. The harmonic restraining force was weakened from 4.0 kcal/mol/Å² in the first 250-ps segment to 3.0, 2.0, and 1.0 kcal/mol/Å² in the following steps, respectively. Before beginning the productive dynamics simulations, each system was checked to ensure that the root-mean-square deviation (RMSD) between the equilibrated and preminimization structures was <1 Å.

For each of the four systems, a 59-ns MD simulation was then performed with a 2-fs time step. Bonds with hydrogen atoms were constrained using the RATTLE algorithm (32), and water geometries were maintained using SETTLE (33), with a bond length error of 0.0005 Å. The temperature bath was kept at 310K with Langevin dynamics. The pressure was maintained at 1 atm using the Nose-Hoover-Langevin piston method (34) with period and decay times set at 100 and 50 femoseconds, respectively. Long-range electrostatics were calculated using Particle mesh Ewald (35). The free-dynamics runs were performed on the TACC Ranger supercomputer. A typical benchmark on the 102 911 and 102 884 atom systems was 4.35 nseconds/day of simulation on 96 processors. The system was sampled every 1 pseconds, generating a total of 59 000 coordinate snapshots. For analysis, every 5th frame was used. Each frame was aligned to the first frame of the trajectory by minimizing the alpha carbon root-mean-square (RMS) deviation using the RMSD Trajectory Tool in VMD (25).

Trajectory clustering

The monomers of each homodimer simulation were isolated, and the two resulting trajectories were concatenated to form a single monomeric trajectory. These monomeric trajectories were subsequently clustered using the gromos algorithm as implemented in the GROMOS++ analysis software (36). The alpha carbon atoms in the active site, defined as all alpha carbon atoms belonging to a residue that was within 5 Å of the NAD or UDP-sugar in the first frame of the trajectory, were used for the mass-weighted RMSD clustering. The RMSD cutoff was increased by 0.05 Å until the trajectory clustering yielded fewer than 30 clusters, with over 90% of all frames contained in the seven largest clusters.

Hydrogen bond analysis

Frames extracted from the simulation every 50 pseconds were used for hydrogen bond analysis. The program HBONANZA (37) was set to identify all persistent hydrogen bonds present in at least 75% of the frames. Angle and distance cutoffs of 30° and 3.5 Å were used, respectively.

Principal component analysis

Principal component analysis (PCA) was performed using the *ptraj* module in AMBER 10⁹. Normalizing each eigenvalue of the covariance matrix to its total sum yields the percent of all *TbGalE* movements attributable to the corresponding eigenvector. That is, the largest eigenvalues correspond to the PC modes that best explain

the molecular motions sampled by the system trajectories. The principal component projections were visualized using a modified version of the Interactive Essential Dynamics module in VMD (38).

Identifying highly conserved active site residues

UniProt (39) was used to identify 7142 members of the sugar epimerase family. Only reviewed sequences were considered for subsequent analysis. These sequences were loaded into the MultiSeq (40) module of the VMD molecular graphics computer program (25). MultiSeq was used to generate a non-redundant set of four representative sequences from this input. CLUSTALW (41) was then used to align these four sequences to the sequence of *TbGalE* obtained from the PDB structure 2CNB (20). A set of active site residues was obtained by identifying all receptor residues from the 2CNB structure (chain A) that came within 3.5 Å of the cocrystallized NAD⁺ cofactor and the UDP-sugar substrate.

HsGalE inhibition and thermal scanning fluorimetry assay

Recombinant wild-type *HsGalE* and human UDP-glucose dehydrogenase (*HsUGDH*) proteins were expressed in *E. coli* and purified as described previously (42,43). DTP compounds were obtained from the NCI/DTP Open Chemical Repository (<http://dtp.cancer.gov>).

The rate of the *HsGalE*-catalyzed reaction was determined by coupling it to the oxidation of UDP-glucose by the action of *HsUGDH*, essentially as previously described (42). Reaction mixtures (20 nM *HsGalE*, 1.2 μM *HsUGDH*, 100 μM UDP-Gal, 10 mM NAD⁺, 10 mM HEPES-NaOH, pH 8.8, 1% (v/v) DMSO) were set up in triplicate with and without DTP compounds (100 μM) in a total volume of 150 μL. The rate of NADH formation was measured at 340 nm for 20 min at 37 °C using a Multiskan Spectrum plate-reader spectrophotometer (Thermo Scientific, Hemel Hempstead, UK) and was used to calculate the rate of production of UDP-glucose by *HsGalE*. Initial rates were calculated from the linear section of each progress curve and the activity expressed as a percentage of the activity in the absence of the compound.

The binding of the DTP compounds was measured using a thermal scanning fluorimetry assay (44). This assay has been used previously to identify small molecular chaperones for the treatment of phenylketonuria (45), to identify stabilizing additives that facilitate crystallization (46), and to measure the binding affinities of carbonic-anhydrase inhibitors (47). Sypro orange (Sigma, Poole, UK), a fluorescent dye, was diluted from a 5000× solution (manufacturer's concentration definition) into a 50 × solution with 10 mM HEPES, pH 8.8. This stock solution was mixed well prior to each use. Reactions were carried out in a total volume of 20 μL and contained 5 μM *HsGalE*, 100 μM DTP compound, 10 mM HEPES, pH 8.8, 1% (v/v) DMSO, 5× Sypro orange. Melting curves were determined using a Rotor-Gene Q cycler (Qiagen, Crawley, UK) with the following protocol: high-resolution melt run (460-nm source, 510-nm detector), 25–95 °C ramp with a 1 °C rise for each step, and no gain optimization. The melting temperatures (T_m) were calculated using the inbuilt analysis software.

Results and Discussion

Evidence of simulation equilibration

Alpha carbon RMSD plots of each homodimer trajectory are shown in Figure 2. As the initial protein conformation of each system was that of the crystallographic UDP-galactose-bound state, the first 9 nseconds of each homodimer simulation were discarded to account for system equilibration. The remaining 50 nseconds of the dimeric simulation were used for subsequent analysis. In total, 400 nseconds of productive *TbGalE* monomer simulation were generated.

A gating mechanism may mediate ligand binding

To better understand how receptor dynamics might impact ligand binding, PCA was used to identify the most important molecular motions of each of the four simulations. Principal components, or eigenvectors, were first calculated for the backbone atoms of each trajectory, and each trajectory was subsequently projected onto the *apo* eigenvectors for reference. A majority of the variance in the molecular motions could be explained by the first two principal components, with almost 40% of the variance explained by the first eigenvector alone (Figure 3A). Consequently, only the first two principal components were used for subsequent analysis (Figure 3B).

The Interactive Essential Dynamics computer package (38) was used to identify the extreme structures of each simulation (Figure 3C,D), as judged by projection onto the first *apo* eigenvector.

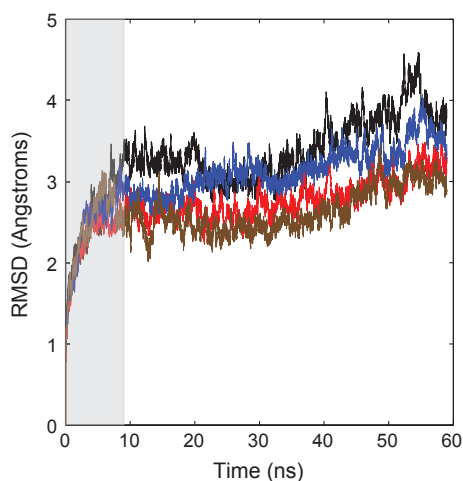


Figure 2: Trajectory RMSD. Each dimer simulation was aligned to the first frame by minimizing the root-mean-square deviation (RMSD) of the C α 's. RMSD was calculated using the first frame as a reference. For this plot, and in subsequent figures, *apo* is depicted in black, UDP-galactose in blue, UDP-glucose in red, and the UDP-ketose intermediate in brown. As each substrate was derived from UDP-galactose, the systems needed 9 nseconds to equilibrate (gray box). Trajectory analysis was performed on the subsequent 50 nseconds of each dimer simulation.

The largest differences in the motions defined by the first principal component were present in the loop region containing residues 230-255. This loop was already thought to be highly dynamic, given that residues 236-248 were unresolved in the 2CNB crystal structure (20).

Much of the difference in the minimum extreme structures can be explained by a kink that forms at the C-terminal end of the α -helix preceding the loop containing residues 230-255 in the *apo* simulation (Figure 3E). This helical bend causes the flexible loop to move toward the active site, as evidenced by the change in the angle

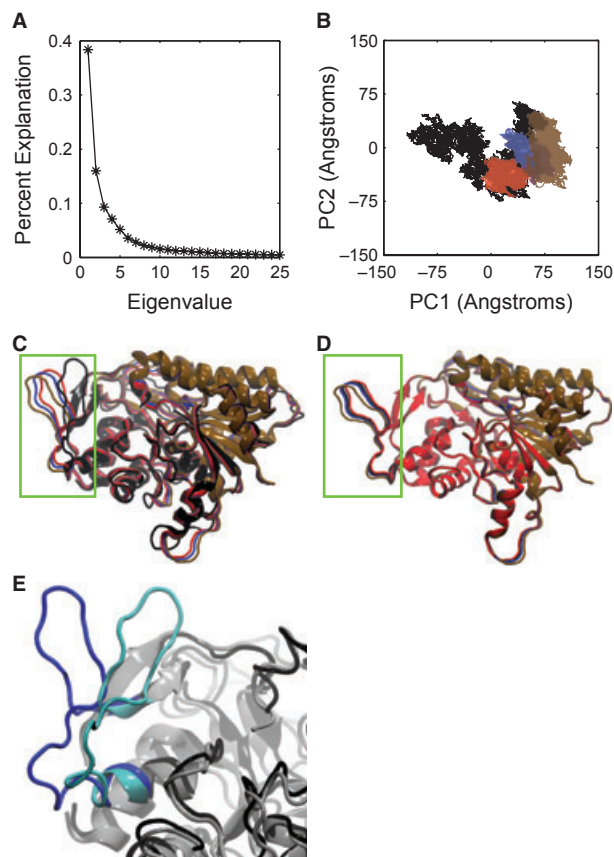


Figure 3: Principal component analysis (PCA). To facilitate comparison, all trajectories were projected onto the principal components of the *apo* simulation. *Apo* is shown in black, UDP-galactose in blue, UDP-glucose in red, and the UDP-ketose intermediate in brown. (A) Percent of the trajectory dynamics attributable to each of the top 25 *apo* eigenvectors. A majority of these motions can be explained by the first two principal components (PC). (B) Projection of the trajectories onto the first two *apo* eigenvectors. Each *holo* simulation explores its own unique set of motions. (C) The PC1 minimum extreme structures for each simulation, color-coded as above. The location of gate is boxed in green. (D) The PC1 maximum extreme structures, color-coded as above. The location of gate is boxed in green. (E) The closed conformation evident in the PC1 minimum *apo* structure results from a kink in the highlighted α -helix. The PC1 minimum structure is shown in cyan/gray, and the maximum is shown in blue/black for reference.

defined by the alpha carbon atoms of D245, G228, and R109, which ranges from 67.69° to 36.01° as PC1 decreases, compacting the protein. Additionally, the corresponding distance between D245 and R109 similarly changes from 29.81 to 16.55 Å. The loop may therefore serve as a gating mechanism likely relevant to ligand binding (Figure 3). As the gating loop containing residues 230–255 is near the active site, it is possible that the movement of this loop helps control the UDP-sugar rate of entry.

UDP-sugar binding: induced fit vs. population shift

The principal components suggest that both induced fit and conformational shift play a role in the binding of *TbGalE* to its natural substrates. The projection of the *apo* simulation onto the first two principal components demonstrates that the *apo* protein explores a large region of conformational space (Figure 3B). Similar projections of the remaining simulations onto the same *apo* principal components reveal that the UDP-glucose and UDP-galactose simulations sample distinct regions within this larger, *apo* conformational space, suggesting a population shift mechanism of binding. On the other hand, the UDP-ketose intermediate samples a region of conformational space largely unexplored by the *apo* protein; the intermediate itself may therefore provoke receptor conformational changes via an induced-fit mechanism. Limited conformational overlap was expected as the ligands only differ at the UDP-sugar C4 atom; these overlaps likely represent the transitions between states during the actual reaction.

These results suggest a general binding mechanism that may be relevant to the study of other receptor–ligand systems as well. The *apo* protein likely samples a large region of conformational space, occasionally assuming conformations in which the gating loop is open, permitting access to the main, UDP-sugar binding site. When UDP-glucose binds, it stabilizes certain conformations, causing the region of the sampled conformational space to constrict via a population shift mechanism. Next, the bound ligand induces changes in the receptor conformation uncharacteristic of the *apo* protein via an induced-fit mechanism. These conformational changes are likely caused by the transformation of the sugar and are required to accommodate the UDP-ketose intermediate. As the epimerization progresses, the conformational space sampled by the protein returns to that sampled in the *apo* state, again indicative of a population shift mechanism of binding.

Remarkably, the regions of the *apo* conformational space sampled by the UDP-galactose- and UDP-glucose-bound proteins are distinct, despite the fact that these two ligands differ by only a chiral inversion at a single sugar carbon atom, demonstrating that even small differences in bound ligands can drastically change the region of conformational space sampled. This finding has significant implications for computer-aided drug design, as it suggests that the crystallographic and MD-derived structures used in virtual screening can vary remarkably depending on the bound ligand. As previous computer-aided drug discovery efforts directed toward *TbGalE* only used *holo* simulations to generate receptor structures for virtual screening (12), additional screening against *apo* conformations may reveal new, effective *TbGalE* inhibitors.

Hydrogen bonding of the UDP-sugar

To examine how the hydrogen bond network of the ligand changes over the course of the reaction, we first identified persistent hydrogen bonds present in over 75% of the trajectory frames analyzed. As expected, the hydrogen bond network that mediates the binding of the UDP-ketose intermediate differs from the networks associated with UDP-galactose and UDP-glucose (Figure 4A–C). Two of the identified hydrogen bonds mediate interactions with Y173 and S142, conserved residues of the catalytic triad that are critical for the two-step GalE epimerization reaction (Figure 1) (11,48).

The hydrogen bond with S142 is particularly interesting, as earlier studies of *EcGalE* suggested that S142 shuttles a proton to Y173 (49). However, crystallographic studies of several GalE proteins have called into question this theory, as they reveal that the distance between the S142 and Y173 hydroxyl groups is generally large. Our results suggest that the specific role of S142 is to stabilize the UDP-ketose intermediate via hydrogen bonding with the C4 hydroxyl group during the conformational flip required for epimerization (Figure 4B).

The UDP-sugar hydrogen bond networks were generally similar regardless of which monomer was used for analysis, with the exception of UDP-galactose. When acceptor–donor heavy-atom distances were calculated over the course of the trajectories (Figure 4D–F), it was noted that the chain B active site of UDP-galactose underwent a conformational shift approximately 5 nseconds into its trajectory, causing a break in the hydrogen bonds between the UDP-galactose and S142 and Y173, respectively. However, this shift was not observed in chain A, which had a UDP-sugar hydrogen bond network more consistent with the *TbGalE* crystal structure, 2CNB. Additionally, despite the conformational shift evident in chain B, the key sugar contacts with S142 and Y173 continued to transiently reform during the course of the simulation, confirming that these residues play an important role in stabilizing the UDP-sugar. Other residues, such as N202 and H221, also formed hydrogen bonds with the UDP-sugar, but more transiently (data not shown).

Like S142 and Y173, L102 may also be important for catalysis, warranting further pharmacological study. Our simulations suggest that L102 may be fundamental to the conformational flip required for epimerization (Figure 1). The carbonyl oxygen atom of L102 is predicted to form a persistent hydrogen bond with the C6 hydroxyl group of the UDP-ketose intermediate. Previous crystallographic studies of a UDP-galactose-like ligand suggested that L102 might mediate ligand binding through interactions with the UDP-sugar C6 hydroxyl group (20). Our simulations suggest that an additional function of the L102 backbone carbonyl group is to maintain the UDP-ketose-intermediate in a conformation approximately halfway between that of UDP-galactose and UDP-glucose via a persistent hydrogen bond. The average UDP-sugar [C1-O-P-O] dihedral angle of the UDP-ketose intermediate is approximately halfway between that of UDP-galactose and UDP-glucose (Figure 4G), supporting this notion.

The same hydrogen bond networks that mediate UDP-sugar binding may be germane to the design and optimization of potential small-molecule therapeutics. For example, a common practice in designing inhibitors is to create transition-state analogs. Our simulation of the UDP-ketose intermediate bound to *TbGalE* has eluci-

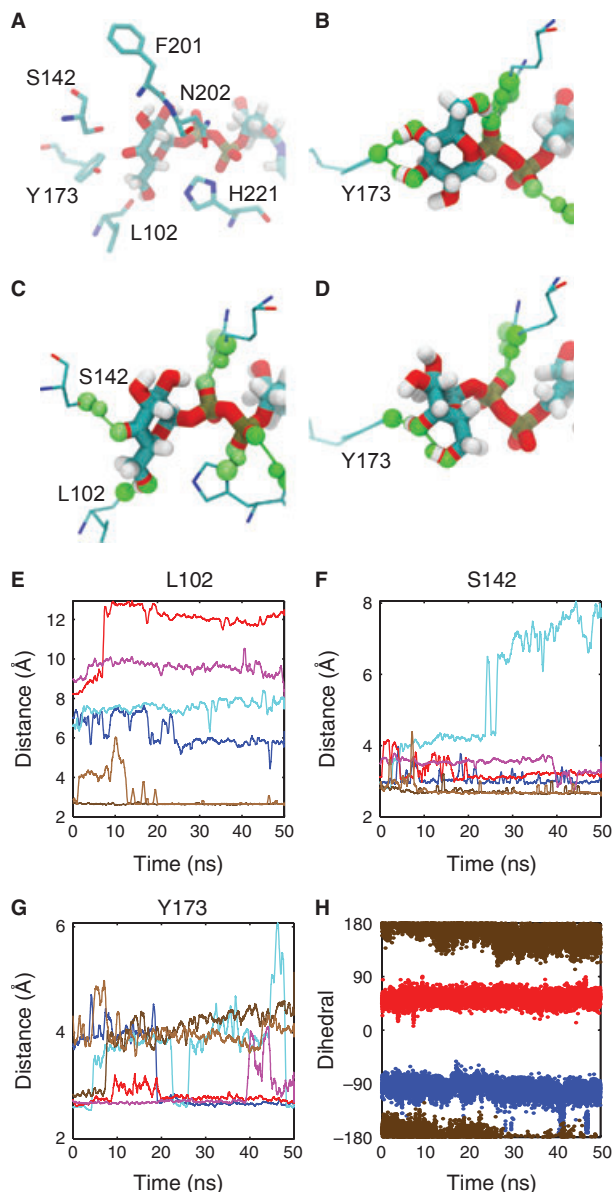


Figure 4: Ligand hydrogen bonding. Hydrogen bonds present in over 75% of the 1000 equidistant frames extracted from each MD simulation are shown in green. In all charts, UDP-galactose is shown in blue, UDP-glucose is shown in red, and the UDP-ketose intermediate is shown in brown. For D–F, chain A is represented by the darker color, and chain B by the lighter color. (A) UDP-glucose. The glucose is held in place by two hydrogen bond interactions with Y173. (B) UDP-ketose intermediate. The orientation of the sugar is flipped from that of UDP-glucose. The sugar C6 hydroxyl group interacts with L102, and the intermediate carbonyl oxygen is tethered to S142. (C) UDP-galactose. The now epimerized sugar, still flipped, forms a single hydrogen bond with Y173. (D–F) Moving average using 500 pseconds windows. (D) Distance from the L102 carbonyl oxygen to the C6 hydroxyl on the UDP-sugar. (E) Distance from the S142 side chain hydroxyl to the C4 hydroxyl on the UDP-sugar. (F) Distance from the Y173 side chain hydroxyl to the C4 hydroxyl on the UDP-sugar. (G) [C1–O–P–O] torsion angle of the UDP-sugar. C1 belongs to the sugar.

dated the important role L102 may play in stabilizing this intermediate. These results suggest that L102 may be a good residue to target in future drug discovery efforts, in addition to the already identified Y173 and S142 (20).

Identification of conserved residues

As current trypanocidal compounds are subject to growing resistance (50–54), future drug design strategies should also attempt to anticipate mutations that may reduce therapeutic efficacy. One strategy to avoid resistance is to pharmacologically target protein residues that are conserved across related members of the same protein family. Conserved residues are likely to be essential for catalysis and/or the binding of natural substrates and thus are less likely to undergo resistance-conferring mutations.

After considering the 392 reviewed members of the sugar epimerase family listed in UniProt, five representative proteins spanning both the eukaryotic and bacterial domains of life were selected: *Trypanosoma brucei* GalE (PDB ID: 2CNB), a chloroplastic protein from *Arabidopsis thaliana* (UniParc: Q8H124), Gal10 from *Schizosaccharomyces pombe* (UniParc: Q9HUU3), bifunctional polymyxin resistance protein ArnA from *Yersinia enterocolitica* (UniParc: A1JPN5), and probable rhamnose biosynthetic enzyme 1 from *Arabidopsis thaliana* (UniParc: Q9SYM5). The sequences of these five proteins were aligned, and the following *TbGalE* active site residues were found to be highly conserved: G7, G10, I12, D75, A100, N117, S141, S142, A143, Y173, K177, and R268. Of these, S142, A143, Y173, K177, and R268 appear to associate with the UDP-sugar.

A143 is particularly noteworthy. While A143 did not form a persistent hydrogen bond with the UDP-sugar C3 hydroxyl group in our simulations (i.e., the bond was present in fewer than 75% of all simulation frames analyzed), this hydrogen bond did form transiently, consistent with the 2CNB crystal structure. As S142 and A143 form hydrogen bonds with adjacent UDP-sugar hydroxyl groups, concurrent binding likely serves to stabilize the conformation of the UDP-sugar in a conformation amenable to catalysis. Thus, small-molecule inhibitors that have adjacent hydroxyl groups or similar hydrogen bonding moieties might be good candidates targeting these two conserved residues.

Allosteric binding

In a recent study, Durrant *et al.* (12) used virtual screening to identify *TbGalE* ligands from among the compounds of the NCI Diversity Set II. Fourteen low-micromolar inhibitors were ultimately reported. However, in addition to these antagonists, seven agonists, unreported at the time, were also identified (Table S1). We found this result interesting, as the agonists were identified by docking small-molecule models into the *TbGalE* active site and therefore should compete with the UDP-sugar substrate rather than enhance its catalysis.

Although *TbGalE* agonists are ill suited as HAT therapeutics, a human condition known as type III galactosemia arises from a deficiency in *HsGalE*. Understanding GalE agonism is therefore of great therapeutic interest. To determine whether agonism might arise from allosteric binding, the *holo* and *apo* monomer trajectories

were clustered into five groups using RMSD cutoffs of 0.7 and 0.75 Å, respectively (Figure 5A). A representative protein conformation was then taken from each cluster; together, these representative conformations are said to constitute an *ensemble*.

The FT-MAP server (55) was then used to computationally flood the entire surface of the ensemble conformations with models of small organic probes in an attempt to identify potential binding pockets. FT-MAP identified the NAD(H) and UDP-sugar binding pockets in every structure; however, no other potential allosteric pockets were consistently recognized (Figure 5B), suggesting that agonism does not likely occur by allosteric modulation.

While the possibility of allosteric binding at an uncharacterized secondary site could not be ruled out, the fact that FT-MAP revealed no such site suggests cooperative agonism mediated by binding to the principal site. It may be that agonist binding to one *TbGalE*

monomer alters the affinity of the binding site on the other dimeric monomer, thus facilitating additional binding. The steep Hill coefficients associated with several of the inhibitors, even in the presence of a detergent that disrupts colloidal aggregates, support this theory (56,57). Hemoglobin is the classic example of cooperative binding, but many other examples exist in nature as well. As an alternate explanation, agonist binding to the principal site of one *TbGalE* monomer might drive dimerization by stabilizing the four-helix bundle at the dimer interface. This theory is supported by previous evidence in related proteins suggesting that, while the monomer is functional, full functionality is achieved only through dimerization (58,59).

Unfortunately, when the identified *TbGalE* agonists were tested against *HsGalE*, no agonism was noted (Figure S1, Tables S2 and S3). However, it is possible that the lack of activity arose from differences in the human and *T. brucei* versions of the GalE protein rather than fundamental differences in the agonistic mechanism. For example, the NAD⁺ adenine-binding pocket is more closed in *TbGalE*, and *HsGalE* G237 in the UDP-sugar binding domain is replaced by C266 in *TbGalE*, a potentially reactive residue that may prove useful in future drug design efforts (11). Further efforts to identify *HsGalE* agonists may therefore be justified.

Conclusions and Future Directions

In this work, we have used molecular simulations to probe the dynamics of *TbGalE* and to specifically analyze the mechanisms governing ligand-binding and enzymatic conversion. Our MD simulations suggest that the conformations sampled by *TbGalE* are highly dependent on the composition of the ligand, as even the chirality of the UDP-sugar C4 atom greatly affected the conformations explored.

Additionally, we have identified an active site residue, L102, that may be important in the stabilization of the UDP-ketose intermediate. While this residue has been previously identified in crystallographic studies as a potential mediator of UDP-galactose binding (20), to our knowledge, its possible role in stabilizing the UDP-ketose intermediate has not been previously recognized.

Finally, as none of our simulations revealed any *TbGalE* allosteric sites, we anticipate that dimeric agonism likely results from either cooperative binding or dimeric stabilization. We hoped that the results presented here will not only provide insight into the function of this and related enzymes, but also assist future computer-aided drug discovery efforts targeting *TbGalE* and *HsGalE*.

Acknowledgements

Funding for this work was provided by NIH 5T32GM007752-32 to AJF and NIH GM31749, NSF MCB-0506593, and MCA93S013 to JAM. Support from the Howard Hughes Medical Institute, the NSF Supercomputer Centers, the San Diego Supercomputer Center, the W.M. Keck Foundation, the National Biomedical Computational Resource, and the Center for Theoretical Biological Physics is gratefully acknowledged. TJM was in receipt of a Department of

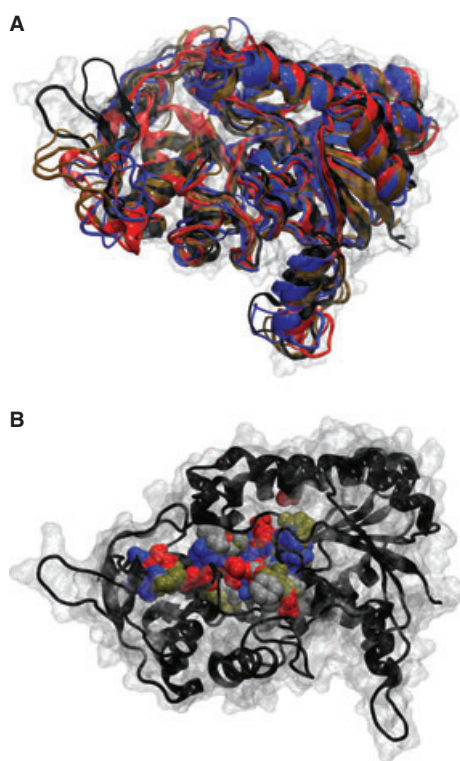


Figure 5: Binding pocket identification. (A) Active site clustering of the *TbGalE* monomers. The frames of each simulation were clustered by the active site C α 's using 0.70 and 0.75 Å RMSD cutoffs for the *holo* and *apo* simulations, respectively. The top two representative structures of chain A from each simulation are shown; *apo* is depicted in black, UDP-galactose in blue, UDP-glucose in red, and the UDP-ketose intermediate in brown. Darker and lighter colors correspond to the most populated and the second most populated cluster, respectively. (B) FT-MAP analysis. The top five clusters from each chain were submitted to the FT-MAP server. Shown are the results for the top chain A cluster of each simulation. These results suggest that *TbGalE* contains no allosteric sites; the observed agonism likely results from ligand-induced dimerization and/or cooperativity between the two monomers.

Employment and Learning (Northern Ireland) PhD studentship. We would also like to thank Michael Urbaniak and Paul Gasper for helpful discussions. The authors declare no conflicts of interest.

References

- Legros D., Ollivier G., Gastellu-Etcheberry M., Paquet C., Burri C., Jannin J., Büscher P. (2002) Treatment of human African trypanosomiasis – present situation and needs for research and development. *Lancet Infect Dis*;2:437–440.
- Maser P., Luscher A., Kaminsky R. (2003) Drug transport and drug resistance in African trypanosomes. *Drug Resist Updat*;6:281–290.
- Delespau V., de Koning H.P. (2007) Drugs and drug resistance in African trypanosomiasis. *Drug Resist Updat*;10:30–50.
- Vansterkenburg E.L., Coppens I., Wilting J., Bos O.J., Fischer M.J., Janssen L.H., Opperdoes F.R. (1993) The uptake of the trypanocidal drug suramin in combination with low-density lipoproteins by *Trypanosoma brucei* and its possible mode of action. *Acta Trop*;54:237–250.
- Gehrig S., Efferth T. (2008) Development of drug resistance in *Trypanosoma brucei rhodesiense* and *Trypanosoma brucei gambiense*. Treatment of human African trypanosomiasis with natural products (Review). *Int J Mol Med*;22:411–419.
- Control and surveillance of African trypanosomiasis. (1998) Report of a WHO Expert Committee. *World Health Organ Tech Rep Ser*;881:I–VI, 1–114.
- Chappuis F., Udayraj N., Stietenroth K., Meussen A., Bovier P.A. (2005) Eflornithine is safer than melarsoprol for the treatment of second-stage *Trypanosoma brucei gambiense* human African trypanosomiasis. *Clin Infect Dis*;41:748–751.
- Iten M., Matovu E., Brun R., Kaminsky R. (1995) Innate lack of susceptibility of Ugandan *Trypanosoma brucei rhodesiense* to DL-alpha-difluoromethylornithine (DFMO). *Trop Med Parasitol*;46:190–194.
- Mehlert A., Zitzmann N., Richardson J.M., Treumann A., Ferguson M.A. (1998) The glycosylation of the variant surface glycoproteins and procyclic acidic repetitive proteins of *Trypanosoma brucei*. *Mol Biochem Parasitol*;91:145–152.
- Tetaud E., Barrett M.P., Bringaud F., Baltz T. (1997) Kinetoplastid glucose transporters. *Biochem J*;325(Pt 3):569–580.
- Shaw M.P., Bond C.S., Roper J.R., Gourley D.G., Ferguson M.A., Hunter W.N. (2003) High-resolution crystal structure of *Trypanosoma brucei* UDP-galactose 4'-epimerase: a potential target for structure-based development of novel trypanocides. *Mol Biochem Parasitol*;126:173–180.
- Durrant J.D., Urbaniak M.D., Ferguson M.A., McCammon J.A. (2010) Computer-aided identification of *Trypanosoma brucei* uridine diphosphate galactose 4'-epimerase inhibitors: toward the development of novel therapies for African sleeping sickness. *J Med Chem*;53:5025–5032.
- Fischer E. (1894) Einfluss der Configuration auf die Wirkung der Enzyme. *Ber Dtsch Chem Ges*;27:2895–2993.
- Durrant J.D., McCammon J.A. (2010) Computer-aided drug-discovery techniques that account for receptor flexibility. *Curr Opin Pharmacol*;10:770–774.
- Koshland D.E. Jr, Ray W.J. Jr, Erwin M.J. (1958) Protein structure and enzyme action. *Fed Proc*;17:1145–1150.
- Kumar S., Ma B., Tsai C.J., Wolfson H., Nussinov R. (1999) Folding funnels and conformational transitions via hinge-bending motions. *Cell Biochem Biophys*;31:141–164.
- Ma B., Kumar S., Tsai C.J., Nussinov R. (1999) Folding funnels and binding mechanisms. *Protein Eng*;12:713–720.
- Ma B., Shatsky M., Wolfson H.J., Nussinov R. (2002) Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. *Protein Sci*;11:184–197.
- Tsai C.J., Ma B., Nussinov R. (1999) Folding and binding cascades: shifts in energy landscapes. *Proc Natl Acad Sci USA*;96:9970–9972.
- Alphey M.S., Burton A., Urbaniak M.D., Boons G.J., Ferguson M.A., Hunter W.N. (2006) *Trypanosoma brucei* UDP-galactose-4'-epimerase in ternary complex with NAD⁺ and the substrate analogue UDP-4-deoxy-4-fluoro-alpha-D-galactose. *Acta Crystallogr Sect F Struct Biol Cryst Commun*;62(Pt 9):829–834.
- Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. (2000) The protein data bank. *Nucleic Acids Res*;28:235–242.
- Dolinsky T.J., Nielsen J.E., McCammon J.A., Baker N.A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res*;32(Web Server issue):W665–W667.
- Li H., Robertson A.D., Jensen J.H. (2005) Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Struct, Funct, Bioinf*;61:704–721.
- Bas D.C., Rogers D.M., Jensen J.H. (2008) Very fast prediction and rationalization of pKa values for protein–ligand complexes. *Proteins: Struct, Funct, Bioinf*;73:765–783.
- Humphrey W., Dalke A., Schulten K. (1996) VMD: visual molecular dynamics. *J Mol Graph*;14:33–38, 27–8.
- Pavelites J.J., Gao J., Bash P.A. Jr (1997) ADM. A molecular mechanics force field for NAD⁺, NADH, and the pyrophosphate groups of nucleotides. *J Comput Chem*;18:221–239.
- Walker R.C., de Souza M.M., Mercer I.P., Gould I.R., Klug D.R. (2002) Large and fast relaxations inside a protein: ÅÅ calculation and measurement of reorganization energies in alcohol dehydrogenase. *J Phys Chem B*;106:11658–11665. [doi: 10.1021/jp0261814].
- Wang J., Wolf R.M., Caldwell J.W., Kollman P.A., Case D.A. (2004) Development and testing of a general amber force field. *J Comput Chem*;25:1157–1174.
- Hornak V., Abel R., Okur A., Strockbine B., Roitberg A., Simmerling C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*;65:712–725.
- Jorgensen W.L., Chandrasekhar J., Madura J.D., Impey R.W., Klein M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys*;79:926–935.
- Phillips J.C., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E., Chipot C., Skeel R.D., Kalé L., Schulten K. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem*;26:1781–1802.
- Andersen H.C. (1983) Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *J Comput Phys*;52:24–34. [doi: 10.1016/0021-9991(83)90014-1].
- Miyamoto S., Kollman P.A. (1992) Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comput Chem*;13:952–962.

34. Feller S.E., Zhang Y., Pastor R.W., Brooks B.R. (1995) Constant pressure molecular dynamics simulation: the Langevin piston method. *J Chem Phys*;103:4613–4621.
35. Darden T., York D., Pedersen L. (1993) Particle mesh Ewald: An N [center-dot] $\log(N)$ method for Ewald sums in large systems. *J Chem Phys*;98:10089–10092.
36. Christen M., Hunenberger P.H., Bakowies D., Baron R., Burgi R., Geerke D.P. *et al.* (2005) The GROMOS software for biomolecular simulation: GROMOS05. *J Comput Chem*;26:1719–1751.
37. Durrant J.D., McCammon J.A. (2011) HBonanza: a computer algorithm for molecular-dynamics-trajectory hydrogen-bond analysis. *J Mol Graph Model*;31:5–9.
38. Mongan J. (2004) Interactive essential dynamics. *J Comput Aided Mol Des*;18:433–436.
39. Bairoch A., Apweiler R., Wu C.H., Barker W.C., Boeckmann B., Ferro S., Gasteiger E., Huang H., Lopez R., Magrane M., Martin M.J., Natale D.A., O'Donovan C., Redaschi N., Yeh L.S. (2005) The universal protein resource (UniProt). *Nucleic Acids Res*;33(Database issue):D154–D159.
40. Roberts E., Eargle J., Wright D., Luthey-Schulten Z. (2006) MultiSeq: unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics*;7:382.
41. Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J., Higgins D.G. (2007) Clustal W and Clustal X version 2.0. *Biochimie*;93:1747–1754.
42. McCorvie T.J., Wasilenko J., Liu Y., Fridovich-Keil J.L., Timson D.J. (2011) In vivo and in vitro function of human UDP-galactose 4'-epimerase variants. *Biochimie*;93:1747–1754.
43. Timson D.J. (2005) Functional analysis of disease-causing mutations in human UDP-galactose 4-epimerase. *FEBS J*;272:6170–6177.
44. Pantoliano M.W., Petrella E.C., Kwasnoski J.D., Lobanov V.S., Myslik J., Graf E., Carver T., Asel E., Springer B.A., Lane P., Salemm F.R. (2001) High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J Biomol Screen*;6:429–440.
45. Pey A.L., Ying M., Cremades N., Velazquez-Campoy A., Scherer T., Thony B., Sancho J., Martinez A. (2008) Identification of pharmacological chaperones as potential therapeutic agents to treat phenylketonuria. *J Clin Invest*;118:2858–2867.
46. Ericsson U.B., Hallberg B.M., Detitta G.T., Dekker N., Nordlund P. (2006) ThermoFluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem*;357:289–298.
47. Matulis D., Kranz J.K., Salemm F.R., Todd M.J. (2005) Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor. *Biochemistry*;44:5258–5266.
48. Allard S.T., Giraud M.F., Naismith J.H. (2001) Epimerases: structure, function and mechanism. *Cell Mol Life Sci*;58:1650–1665.
49. Liu Y., Thoden J.B., Kim J., Berger E., Gulick A.M., Ruzicka F.J., Holden H.M., Frey P.A. (1997) Mechanistic roles of tyrosine 149 and serine 124 in UDP-galactose 4-epimerase from *Escherichia coli*. *Biochemistry*;36:10675–10684.
50. Carter N.S., Fairlamb A.H. (1993) Arsenical-resistant trypanosomes lack an unusual adenosine transporter. *Nature*;361:173–176.
51. Vincent I.M., Creek D., Watson D.G., Kamleh M.A., Woods D.J., Wong P.E., Burchmore R.J., Barrett M.P. (2010) A molecular mechanism for eflornithine resistance in African trypanosomes. *PLoS Pathog*;6:e1001204.
52. Stewart M.L., Krishna S., Burchmore R.J., Brun R., de Koning H.P., Boykin D.W., Tidwell R.R., Hall J.E., Barrett M.P. (2005) Detection of arsenical drug resistance in *Trypanosoma brucei* with a simple fluorescence test. *Lancet*;366:486–487.
53. Matovu E., Geiser F., Schneider V., Maser P., Enyaru J.C., Kaminsky R., Gallati S., Seebeck T. (2001) Genetic variants of the TbAT1 adenosine transporter from African trypanosomes in relapse infections following melarsoprol therapy. *Mol Biochem Parasitol*;117:73–81.
54. Barrett M.P., Fairlamb A.H. (1999) The biochemical basis of arsenical-diamidine crossresistance in African trypanosomes. *Parasitol Today*;15:136–140.
55. Brenke R., Kozakov D., Chuang G.Y., Beglov D., Hall D., Landon M.R., Mattos C., Vajda S. (2009) Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics*;25:621–627.
56. Langmead C.J. (2007) Screening for positive allosteric modulators: assessment of modulator concentration-response curves as a screening paradigm. *J Biomol Screen*;12:668–676.
57. Surig U., Gaal K., Kostenis E., Trankle C., Mohr K., Holzgrabe U. (2006) Muscarinic allosteric modulators: atypical structure-activity-relationships in bispyridinium-type compounds. *Arch Pharm (Weinheim)*;339:207–212.
58. Brahma A., Banerjee N., Bhattacharyya D. (2009) UDP-galactose 4-epimerase from *Kluyveromyces fragilis* – catalytic sites of the homodimeric enzyme are functional and regulated. *FEBS J*;276:6725–6740.
59. Hay M., Bhaduri A. (1975) UDP-glucose 4-epimerase from *Saccharomyces fragilis*. Allosteric kinetics with UDP-glucose as substrate. *J Biol Chem*;250:4373–4375.

Notes

^aFrisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, *et al.* Gaussian 03, Revision C.02.

^bCase DA, DT, Cheatham TE III, Simmerling CL, Wang J, *et al.* (2008) AMBER 10. San Francisco: University of California.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Thermal scanning fluorimetry of *HsGalE*.

Table S1. *TbGalE* Agonists

Table S2. Percentage activity of 20 nM *HsGalE* in the presence of different DTP compounds

Table S3. Melting temperatures of *HsGalE* in the presence of different DTP compounds

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.