

M. Ferrell,^{a,b*} J. Abendroth,^{a,c}
Y. Zhang,^{a,b} B. Sankaran,^d
T. E. Edwards,^{a,c} B. L. Staker,^{a,c}
W. C. Van Voorhis,^{a,e}
L. J. Stewart^{a,c} and P. J. Myler^{a,b}

^aSeattle Structural Genomics Center for Infectious Disease, USA, ^bSeattle Biomedical Research Institute, 307 Westlake Avenue North, Suite 500, Seattle, WA 98109, USA, ^cEmerald BioStructures Inc., 7869 NE Day Road West, Bainbridge Island, WA 98110, USA, ^dBerkeley Center For Structural Biology, Ernest Orlando Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA, and ^eSchool of Medicine, University of Washington, Seattle, WA 98195, USA

Correspondence e-mail:
micah.ferrell@seattlebiomed.org

Received 1 March 2011
Accepted 1 August 2011

PDB Reference: aldose reductase, 3krb.

Structure of aldose reductase from *Giardia lamblia*

Giardia lamblia is an anaerobic aerotolerant eukaryotic parasite of the intestines. It is believed to have diverged early from eukarya during evolution and is thus lacking in many of the typical eukaryotic organelles and biochemical pathways. Most conspicuously, mitochondria and the associated machinery of oxidative phosphorylation are absent; instead, energy is derived from substrate-level phosphorylation. Here, the 1.75 Å resolution crystal structure of *G. lamblia* aldose reductase heterologously expressed in *Escherichia coli* is reported. As in other oxidoreductases, *G. lamblia* aldose reductase adopts a TIM-barrel conformation with the NADP⁺-binding site located within the eight β-strands of the interior.

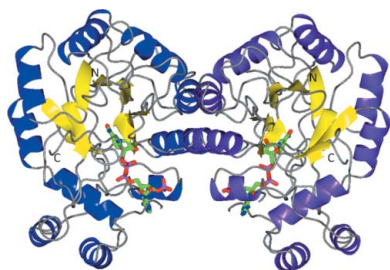
1. Introduction

1.1. *Giardia lamblia*

G. lamblia is the etiological agent of giardiasis, a common cause of diarrheal disease in the developing world. It is an anaerobic aerotolerant eukaryotic parasite of the lower intestine. *Giardia* sp. are believed to be basal eukaryotes, with a minimized genome and simplified metabolic pathways (Morrison *et al.*, 2007). Mitochondria and the machinery of oxidative phosphorylation are absent, with energy being derived from substrate-level phosphorylation. Carbohydrate metabolism is largely limited to D-glucose (Lindmark, 1980), while other sugars, sugar alcohols, Krebs-cycle intermediates and organic acids have not been found to be utilized. Glucose is oxidized to ethanol and acetate as the fermentation end products (Lindmark, 1980; Jarroll *et al.*, 1981; Schofield *et al.*, 1991), with ethanol being the dominant product under anaerobic conditions and acetate predominating under aerobic conditions. When cultured in glucose-depleted media, *Giardia* grow slowly, achieve lower cell densities and secrete little ethanol (Schofield *et al.*, 1990). While glucose appears to be the preferred carbon source, it is not however essential for *Giardia* growth. In the absence of glucose the carbon skeletons of amino acids are used as a substitute, with alanine secreted as the waste product. Ethanol production is largely absent under glucose-limited conditions. Alcohol dehydrogenase (ADH), while present in *Giardia*, does not appear to play a role in the production of the ethanol waste product. Cultures grown in the presence of pyrazole, an inhibitor of ADH, show no reduction in ethanol production. Instead, it is believed that ethanol is produced by the reduction of acetaldehyde by aldose reductase in an NADPH-dependent reaction. *Giardia* grown in the presence of sodium valproate, an inhibitor of aldose reductase, show diminished ethanol production and limited growth (Schofield *et al.*, 1991).

1.2. Aldose reductase

Aldose reductases (EC 1.1.1.21) are ubiquitous NADPH-dependent oxidoreductases that catalyze the reduction of aldehydes to alcohols. As the name suggests, the aldehyde substrates have been classically characterized as aldose sugars; however, in recent years there has been an increased appreciation of the potential of these



enzymes to detoxify small aldehydes. Typical aldose reductases exhibit the TIM-barrel fold with eight α -helices on the exterior and an eight-stranded β -barrel in the interior. Binding of the NADPH cofactor is facilitated by a deep elliptical pocket near the C-terminal end of the β -barrel (Borhani *et al.*, 1992; Wilson *et al.*, 1992; Rondeau *et al.*, 1992). Mutagenic studies have shown that tight cofactor binding is dependent on a conserved arginine residue (Arg269 in PDB entry 3kbr; Bohren *et al.*, 2005). The active site consists of two distinct pockets: a rigid anion-binding pocket and a flexible hydrophobic specificity pocket. A third region near the C-terminus has been shown to change position upon binding certain ligands (Urzhumtsev *et al.*, 1997; Kinoshita *et al.*, 2002; Podjarny *et al.*, 2004). Aldose reductase is common to both prokaryotes and eukaryotes, with TIM-barrel folds being reported from a diversity of species. Aldose reductase functions by a sequential ordered mechanism (Kubiseski *et al.*, 1992; Grimshaw *et al.*, 1995), with NADPH being bound before substrate. Activity involves a stereospecific transfer of the 4-pro-*R* hydride from NADPH to the substrate carbonyl C atom followed by the protonation of the substrate carbonyl O atom by a conserved tyrosine residue (Tyr40 in 3kbr). Proton transfer in the active site is assisted by a hydrogen bond between the active tyrosine and the ϵ -amino group of a conserved lysine (Lys71 in 3kbr), which is itself linked by a salt bridge to a conserved aspartic acid (Asp35 in 3kbr). The substrate is oriented in the active site by a conserved histidine residue (His104 in 3kbr) (Del Corso *et al.*, 2008).

Owing to its role in human disease, human aldose reductase has been the subject of over thirty years of intense research. In diabetic hyperglycemia, the hexokinase of insulin-independent tissues becomes saturated with glucose, causing excess glucose to instead be shunted into the polyol pathway, where aldose reductase converts it to sorbitol. Disease results from sorbitol-induced hyperosmotic swelling and oxidative stress owing to the lowered glutathione concentration caused by the depletion of cellular NADPH reserves. A large number of inhibitors have been identified that bind the active site and many crystal structures of mammalian aldose reductases with bound inhibitors have been solved. Unfortunately, despite these efforts, very few of these compounds have made it past clinical trials (Del Corso *et al.*, 2008).

2. Methods

2.1. Protein expression and purification

Full-length aldose reductase (Morrison *et al.*, 2007) from *G. lamblia* ATCC 50803 was cloned into pAVA0421 vector (Alexandrov *et al.*, 2004) by ligand-independent cloning (LIC; Aslanidis & de Jong, 1990) to produce a construct with an N-terminal hexahistidine tag that is cleavable with 3C protease (the entire tag sequence was MAHHHHHMGTLAQTO/GPGS-ORF, in which the 3C cleavage site is marked by a prime). Protein was expressed in *Escherichia coli* BL21 (DE3) cells in 2 l auto-induction medium (Studier, 2005) in a LEX bioreactor (Harbinger, Markham, Ontario, Canada) at 293 K for 72 h, after which the harvested cells were flash-frozen in liquid nitrogen. The frozen cell pellet was thawed and resuspended by vortexing in 200 ml lysis buffer [20 mM HEPES pH 7.4, 300 mM NaCl, 5% glycerol, 30 mM imidazole, 0.5% CHAPS, 10 mM MgCl₂, 3 mM β -mercaptoethanol, 1.3 mg ml⁻¹ protease-inhibitor cocktail (Roche, Basel, Switzerland) and 0.05 mg ml⁻¹ lysozyme]. The cell suspension was packed on ice and disrupted by sonication for 15 min in 5 s pulses at 70% amplitude using a Branson 450D Sonifier (Branson Ultrasonics, Danbury, Connecticut, USA). The lysate was incubated with 20 μ l Benzonase nuclease (EMD Chemicals,

Table 1

Data-collection statistics.

Values in parentheses are for the highest resolution shell.

Beamline	ALS 5.0.1
Wavelength (Å)	0.9774
Space group	C2
Unit-cell parameters (Å, °)	$a = 196.77, b = 66.09, c = 56.29,$ $\beta = 92.26$
Resolution range (Å)	50–1.75 (1.80–1.75)
Mean $I/\sigma(I)$	17.0 (2.5)
$R_{\text{merge}}^{\dagger}$	0.073 (0.558)
Completeness (%)	99.9 (99.8)
Multiplicity	4.6 (4.1)
No. of unique reflections	72824 (5374)

$$\dagger R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$$

Gibbstown, New Jersey, USA) for 40 min at room temperature under gentle agitation. The lysate was clarified by centrifugation with a Sorvall RC5 at 10 000 rev min⁻¹ for 60 min at 277 K in a F14S rotor (Thermo Fisher, Waltham, Massachusetts, USA). The clarified solution was syringe-filtered through a 0.45 μ m cellulose acetate filter (Corning Life Sciences, Lowell, Massachusetts, USA). The tagged protein was purified by affinity chromatography using a HisTrap FF 5 ml column (GE Biosciences, Piscataway, New Jersey, USA) equilibrated in binding buffer (25 mM HEPES pH 7.0, 300 mM NaCl, 5% glycerol, 30 mM imidazole, 1 mM DTT) and eluted with 500 mM imidazole in the same buffer. To cleave the N-terminal affinity tag, peak fractions were pooled and assayed for concentration by 280 nm spectrophotometry; 3C protease (Alexandrov *et al.*, 2001) was mixed with the target in a 1:50 ratio and dialyzed overnight at 277 K against cleavage buffer (20 mM HEPES pH 7.6, 200 mM NaCl, 5% glycerol, 1 mM DTT). Uncleaved target, protease and cleaved tag were removed by a second round of affinity chromatography using a 5 ml HisTrap column. The flowthrough and wash from secondary affinity chromatography were pooled and concentrated using an Amicon Ultra-15 30 kDa molecular-weight cutoff concentrator (Millipore, Billerica, Massachusetts, USA). The concentrated sample was further purified by size-exclusion chromatography (SEC) using a Superdex 75 26/60 column (GE Biosciences) equilibrated in SEC buffer (20 mM HEPES pH 7.0, 300 mM NaCl, 5% glycerol and 2 mM DTT) attached to an ÄKTAprime plus FPLC system (GE Biosciences). Peak fractions were collected and assessed for purity by SDS-PAGE on 4–12% NuPAGE gels (Invitrogen, Carlsbad, California, USA) with Coomassie staining using SimplyBlue Safestain (Invitrogen). Pure fractions were pooled, concentrated to 23 mg ml⁻¹ and flash-frozen in liquid nitrogen. The final concentration was determined by 280 nm spectrophotometry and the final purity was assayed by SDS-PAGE.

2.2. Crystallization

With the purified protein at a concentration of 25.2 mg ml⁻¹ in SEC buffer, two sparse-matrix screens were set up using JCSG+ (Emerald BioSystems, Bainbridge Island, Washington, USA), PACT (Molecular Dimensions, Suffolk, England), Index and Crystal Screen (Hampton Research, Aliso Viejo, California, USA) following an extended Newman's strategy (Newman *et al.*, 2005). 0.4 μ l protein solution was mixed with 0.4 μ l well solution and equilibrated against 100 μ l reservoir using 96-well Compact Jr crystallization plates from Emerald BioSystems. Crystals were found under several PEG conditions, while crystals suitable for X-ray diffraction experiments were obtained from PACT condition H6: 20% PEG 3350, 200 mM sodium formate, 100 mM Bis-Tris propane. The crystals were cryoprotected by soaking them in a buffer consisting of 25% ethylene glycol mixed

Table 2
Refinement and model statistics.

Beamline	ALS 5.0.1
Resolution range (Å)	50–1.75
$R_{\text{cryst}}^{\dagger}$	0.144
$R_{\text{free}}^{\dagger}$	0.173
R.m.s.d. bonds (Å)	0.016
R.m.s.d. angles (°)	1.56
Protein atoms	4908
Nonprotein atoms	767
Mean B factor (Å ²)	16.7
Residues in favoured region	580 (97.2%)
Residues in allowed region	17 (2.9%)
Residues in disallowed region	0 (0%)
<i>MolProbity</i> ‡ score (percentile)	1.13 (99th)
PDB code	3krb

[†] $R_{\text{cryst}} = \sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|$. The free R factor was calculated using 5% of the reflections omitted from the refinement (Winn *et al.*, 2011). [‡] Chen *et al.* (2010).

with reservoir solution. The crystals were vitrified by plunging them directly into liquid nitrogen.

2.3. Data collection and structural determination

Diffraction data were collected on the Berkeley Center for Structural Biology ALS 5.0.1 beamline as part of the Collaborative Crystallography program. The beamline uses a wavelength of 0.9774 Å and is equipped with an ADSC Quantum 210 CCD detector. The data were reduced in the monoclinic space group $C2$ to 1.75 Å resolution with *XDS/XSCALE* (Kabsch, 1988, 2010; see Table 1).

Packing density (Matthews, 1968) suggested the presence of two molecules of aldose reductase in the asymmetric unit ($V_M = 1.58 \text{ Å}^3 \text{ Da}^{-1}$, 52% solvent). A search of the PDB for sequence homology yielded a human aldose-reductase-like protein (AKR1B10) as the closest homolog with known structure (PDB entry 1zua; Gallego *et al.*, 2007), with 43% sequence identity. A search model was generated from molecule *A* of PDB entry 1zua using the *CCP4* program *CHAINS*AW (Stein, 2008; Winn *et al.*, 2011). The structure was solved by molecular replacement with the *CCP4* program *Phaser* (McCoy *et al.*, 2007); two molecules could be placed with high Z

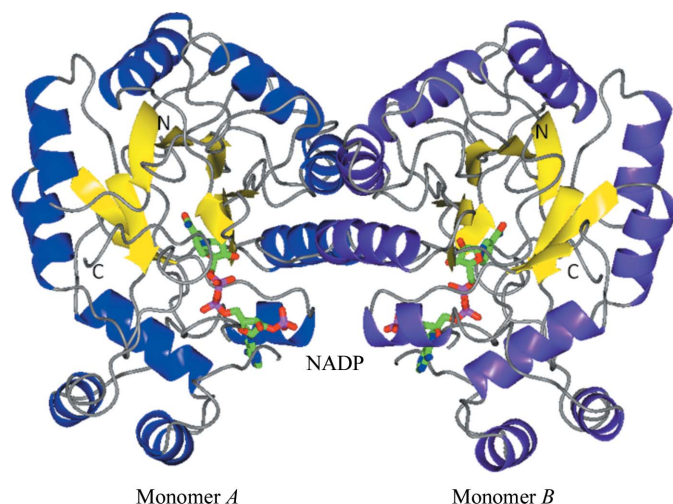


Figure 1
Dimer of aldose reductase from *G. lamblia* (GilaA.01452.aA1). Aldose reductase from *G. lamblia* forms a dimer of two TIM barrels. In this figure the twofold axis runs approximately vertically. The helices of the two monomers are shown as blue and purple ribbons. In each monomer, an NADP molecule is visible at the C-terminal end of the strands of the β -barrel. This figure was prepared with *CCP4mg* (McNicholas *et al.*, 2011).

scores. The model was then iteratively extended manually using *Coot* (Emsley & Cowtan, 2004) followed by cycles of reciprocal-space refinement with the *CCP4* program *REFMAC5* (Murshudov *et al.*, 2011). The final model was validated with the validation tools within *Coot* and with *MolProbity* (Chen *et al.*, 2010).

The final model was refined at 1.75 Å resolution to $R_{\text{work}} = 0.144$ and $R_{\text{free}} = 0.173$ with good stereochemistry (see Table 2). The observed structure extends from residue Ser0 (part of the purification tag) to Asp313 in both chains. The loop between residues Ala111 and Thr119 could not be modeled owing to weak electron density. One NADP⁺ molecule and one ethylene glycol molecule were modelled in each chain. The nicotinamide ring of the cofactor is planar; hence, we assume that NADP⁺ is bound to the protein in the oxidized form. The identity of a slightly prolate spheroid of electron density close to the nicotinamide ring of NADP⁺ could not be established and was therefore modelled as ‘unknown atoms’ (UNK). A total of 611 water molecules were located.

3. Results and discussion

G. lamblia aldose reductase adopts a TIM-barrel fold with the NADP⁺-binding site located within the eight β -strands of the interior

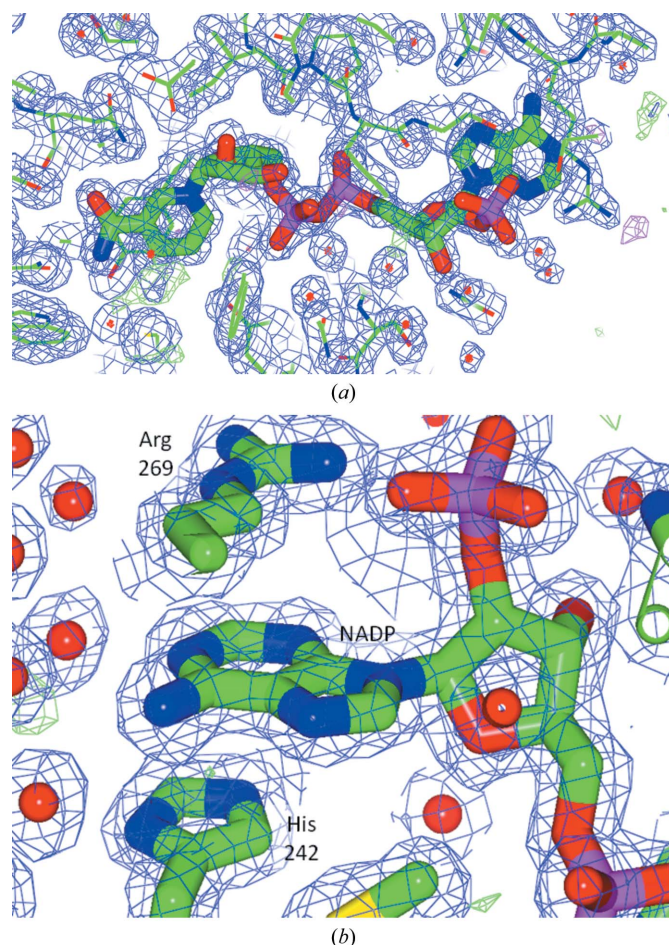


Figure 2
(a) NADP is well defined in the structure of *G. lamblia* aldose reductase. (b) The adenine ring of NADP stacks hydrophobically with the side chains of His242 and Arg269. While Arg269 is conserved between human and *Giardia* aldose reductase, His242 of *Giardia* aldose reductase is an alanine in the human enzyme. This figure was prepared with *CCP4mg* (McNicholas *et al.*, 2011). In both cases σ_A -weighted $2|F_o| - |F_c|$ electron density is shown in blue at 1σ and σ_A -weighted $|F_o| - |F_c|$ electron density is shown in green and red at $\pm 3\sigma$.

(Fig. 1). Sequence alignment with *Homo sapiens* aldose reductase shows 44% identity; no structures in the PDB have an identity greater than 50%. The two monomers of *G. lamblia* aldose reductase superimpose with a root-mean-square deviation (r.m.s.d.) of 0.3 Å (on C α atoms); a search of the PDB for structural homologues using *SSM/PDBeFold* (Krissinel & Henrick, 2004) yielded several structures of human aldose reductase with C α -atom r.m.s.d.s of 1.0–1.1 Å. When compared with the structure of mammalian aldose reductase, subtle variations in structure are apparent.

There is a distinct structural divergence of the N-termini of *G. lamblia* aldose reductase and the human protein: while the N-terminal 14 residues of both structures are well ordered, *G. lamblia* aldose reductase lacks the small N-terminal capping β -sheet exhibited by mammalian structures (residues 2–14 of human aldose reductase).

Cofactor binding by Arg269 is conserved between the two structures. The side chain of His242 of *G. lamblia* aldose reductase hydrophobically stacks with the adenine of NADP⁺, while this residue is an alanine in the human protein (Fig. 2). This histidine residue is not found in other aldose reductases of higher organisms and may be unique to *Giardia*. A flexible loop between the strand and helix 7 of the β/α -barrel (dubbed the ‘safety belt’) has previously been reported to sequester solvent from the NADP⁺ cofactor (Wilson *et al.*, 1992). This flexible feature was not resolved in this structure. Tantalizingly, this region shows divergence between the mammalian and *Giardia* sequences. Sequence comparison with human aldose reductase by *ClustalW* (Larkin *et al.*, 2007) analysis (Fig. 3) indicates that phenylalanine residues at positions 109 and 118 are conserved participants in forming the hydrophobic substrate-specificity pocket. Other residues identified as components of the mammalian specificity pocket (Klebe *et al.*, 2004; Harrison *et al.*, 1994; Sotriffer *et al.*, 2004; Urzhumtsev *et al.*, 1997) are divergent (Table 3). The residues of the anion-binding pocket and the active site appear to be completely conserved (Table 3) and their orientation is conserved in the three-dimensional structure. The residues flanking the C-terminal mobile region appear to be conservative substitutions, while the interior residues of this short region show no conservation (Table 3).

The structure of the *G. lamblia* aldose reductase was superimposed with ligand-bound aldose reductases such as human aldose reductase in complex with fidarestat (PDB entry 1ef3; Oka *et al.*, 2010) and porcine aldose reductase in complex with tolrestat (PDB entry 1ah3;

Table 3
Conservation of residues involved in substrate binding.

	<i>H. sapiens</i>	<i>G. lamblia</i>
Specificity pocket	Thr114	Leu107
	Cys304	Ala304
	Tyr310	Pro310
Anion-binding pocket	Trp21	Trp12
	Val48	Val39
	Tyr49	Tyr40
	His111	His104
	Trp112	Trp105
	Lys78	Lys71
C-terminal mobile region	Asp44	Asp35
	Val298	Ile297
	Cys304	Ala304
	Ala299	Phe299
	Leu300	Cys301

Urzhumtsev *et al.*, 1997) or sorbinil (PDB entry 1ah0; Urzhumtsev *et al.*, 1997). All three inhibitors bind in close proximity to the nicotinamide group of NADP (Fig. 4). The protein models superimpose well overall, with r.m.s.d.s of 1.06 Å (1ef3), 1.03 Å (1ah3) and 1.13 Å (1ah0). Despite high overall structural homology, there is a distinct difference in structure, separate from the mobile regions defined earlier, between the *G. lamblia* and mammalian enzymes close to the inhibitor-binding site. The C-terminal loop formed by residues Pro303–Leu311 diverges significantly from the structure of the corresponding loop in the human and porcine structures. In the human and porcine structures this loop is much more conserved and interacts with the inhibitors. In addition, structural rearrangements would be necessary for *G. lamblia* aldose reductase to harbor any of these three inhibitors. This structural diversity of the ligand-binding pocket in turn provides opportunities for the design of drugs that are specific for *G. lamblia* aldose reductase.

4. Conclusion

The structure of *G. lamblia* aldose reductase in complex with NADP⁺ was solved to a resolution of 1.75 Å. Residues in the active site and anion-binding pocket show conservation in sequence and structure with mammalian structures, while the nonphenylalanine residues involved in forming the specificity pocket show divergence. This structure did not resolve a flexible loop involved in solvent

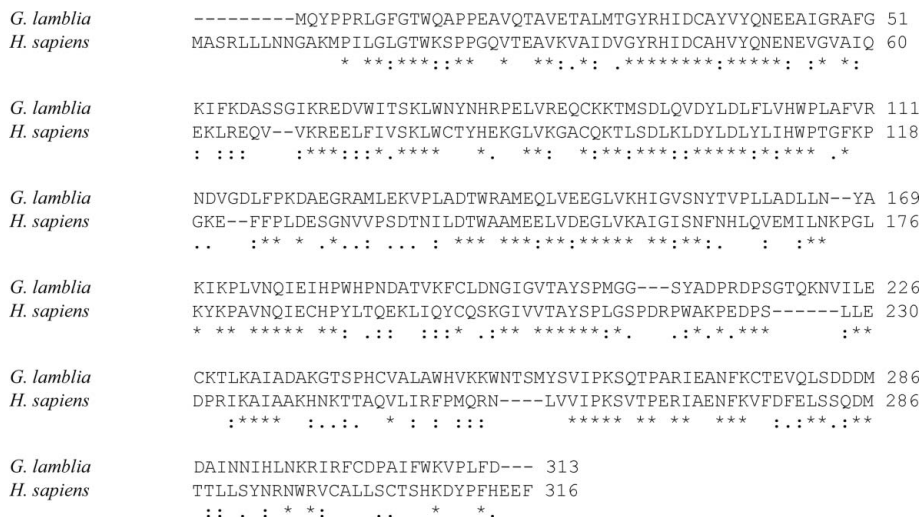


Figure 3
ClustalW (Larkin *et al.*, 2007) comparison of the amino-acid sequences of *H. sapiens* and *G. lamblia* aldose reductase.

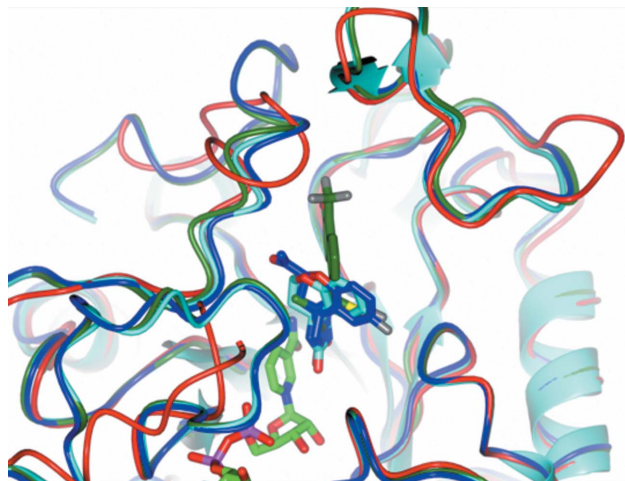


Figure 4

Inhibitor-binding pocket of *G. lamblia* aldose reductase in comparison with ligand-bound human and porcine aldose reductases. *G. lamblia* aldose reductase (red model) superimposes with small r.m.s.d.s with human aldose reductase in complex with fidarestat (PDB entry 1ef3, dark blue) and porcine aldose reductase in complex with tolrestat (PDB entry 1ah3, green) and sorbinil (PDB entry 1ah0, cyan). Despite the overall structural homology, the C-terminal loop, which interacts with the ligand, shows a distinctly different trace. For clarity, the C atoms of the inhibitors follow the same color scheme as the corresponding ribbon. The NADP molecule is shown only for *G. lamblia*, with light green C atoms. This figure was prepared with CCP4mg (McNicholas *et al.*, 2011).

sequestration of NADP⁺. A conserved arginine residue is involved in cofactor binding together with an apparently *Giardia*-specific histidine residue.

The authors wish to thank all of the members of the SSGCID team. This research was funded under Federal Contract No. HHSN272200700057C from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services. The Berkeley Centre for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences and the US Department of Energy under Contract No. DE-AC02-05CH11231.

References

- Alexandrov, A., Dutta, K. & Pascal, S. M. (2001). *Biotechniques*, **30**, 1194–1198.
- Alexandrov, A., Vignali, M., LaCount, D. J., Quartley, E., de Vries, C., De Rosa, D., Babulski, J., Mitchell, S. F., Schoenfeld, L. W., Fields, S., Hol, W. G., Dumont, M. E., Phizicky, E. M. & Grayhack, E. J. (2004). *Mol. Cell Proteomics*, **3**, 934–938.
- Aslanidis, C. & de Jong, P. J. (1990). *Nucleic Acids Res.* **18**, 6069–6074.
- Bohren, K. M., Brownlee, J. M., Milne, A. C., Gabbay, K. H. & Harrison, D. H. (2005). *Biochim. Biophys. Acta*, **1748**, 201–212.
- Borhani, D. W., Harter, T. M. & Petrash, J. M. (1992). *J. Biol. Chem.* **267**, 24841–24847.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst. D* **66**, 12–21.
- Del Corso, A., Cappiello, M. & Mura, U. (2008). *Curr. Med. Chem.* **15**, 1452–1461.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst. D* **60**, 2126–2132.
- Gallego, O., Ruiz, F. X., Ardèvol, A., Domínguez, M., Alvarez, R., de Lera, A. R., Rovira, C., Farrés, J., Fita, I. & Parés, X. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 20764–20769.
- Grimshaw, C. E., Bohren, K. M., Lai, C. J. & Gabbay, K. H. (1995). *Biochemistry*, **34**, 14356–14365.
- Harrison, D. H., Bohren, K. M., Ringe, D., Petsko, G. A. & Gabbay, K. H. (1994). *Biochemistry*, **33**, 2011–2020.
- Jarroll, E. L., Muller, P. J., Meyer, E. A. & Morse, S. A. (1981). *Mol. Biochem. Parasitol.* **2**, 187–196.
- Kabsch, W. (1988). *J. Appl. Cryst.* **21**, 67–72.
- Kabsch, W. (2010). *Acta Cryst. D* **66**, 125–132.
- Kinoshita, T., Miyake, H., Fujii, T., Takakura, S. & Goto, T. (2002). *Acta Cryst. D* **58**, 622–626.
- Klebe, G., Krämer, O. & Sotriffer, C. (2004). *Cell. Mol. Life Sci.* **61**, 783–793.
- Krissinel, E. & Henrick, K. (2004). *Acta Cryst. D* **60**, 2256–2268.
- Kubieski, T. J., Hyndman, D. J., Morjana, N. A. & Flynn, T. G. (1992). *J. Biol. Chem.* **267**, 6510–6517.
- 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). *Bioinformatics*, **23**, 2947–2948.
- Lindmark, D. G. (1980). *Mol. Biochem. Parasitol.* **1**, 1–12.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. (2011). *Acta Cryst. D* **67**, 386–394.
- Morrison, H. G. *et al.* (2007). *Science*, **317**, 1921–1926.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst. D* **67**, 355–367.
- Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J. L., Stuart, D. I. & Perrakis, A. (2005). *Acta Cryst. D* **61**, 1426–1431.
- Oka, M., Matsumoto, Y., Sugiyama, S., Tsuruta, N. & Matsushima, M. (2010). *J. Med. Chem.* **43**, 2479–2483.
- Podjarny, A., Cachau, R. E., Schneider, T., Van Zandt, M. & Joachimiak, A. (2004). *Cell. Mol. Life Sci.* **61**, 763–773.
- Rondeau, J. M., Tête-Favier, F., Podjarny, A., Reymann, J. M., Barth, P., Biellmann, J. F. & Moras, D. (1992). *Nature (London)*, **355**, 469–472.
- Schofield, P. J., Edwards, M. R. & Kranz, P. (1991). *Mol. Biochem. Parasitol.* **45**, 39–47.
- Schofield, P. J., Kranz, P. & Edwards, M. R. (1990). *Int. J. Parasitol.* **20**, 701–703.
- Sotriffer, C. A., Krämer, O. & Klebe, G. (2004). *Proteins*, **56**, 52–66.
- Stein, N. (2008). *J. Appl. Cryst.* **41**, 641–643.
- Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.
- Urzhumtsev, A., Tête-Favier, F., Mitschler, A., Barbanton, J., Barth, P., Urzhumtseva, L., Biellmann, J. F., Podjarny, A. & Moras, D. (1997). *Structure*, **5**, 601–612.
- Wilson, D. K., Bohren, K. M., Gabbay, K. H. & Quiocho, F. A. (1992). *Science*, **257**, 81–84.
- Winn, M. D. *et al.* (2011). *Acta Cryst. D* **67**, 235–242.