Structure and function of L-threonine-3-dehydrogenase from the parasitic protozoan *Trypanosoma brucei* revealed by X-ray crystallography and geometric simulations

Eyram Adjogatse*[a], Peter Erskine[a], Stephen Wells[b], John M Kelly[c], Jonathan D Wilden[d], A. W. Edith Chan[e], David Selwood[e], Alun Coker[a], Steve Wood[a], Jonathan B. Cooper[a]

*Corresponding Author (rmhaead@ucl.ac.uk)

[a]Laboratory for Protein Crystallography, Wolfson Institute for Biomedical Research, University College London, London, WC1E 6BT, United Kingdom

[b]Department of Chemistry, University of Bath, Bath, BA2 7AY, United Kingdom

[c]Department of Pathogen Molecular Biology, London School of Hygiene & Tropical Medicine, London, WC1E 7HT, United Kingdom

[d]Department of Chemistry, University College London, London, WC1H OAJ, United Kingdom

[e]Drug Discovery, Wolfson Institute for Biomedical Research, University College London, London, WC1E 6BT, United Kingdom

Abstract

Two of the world's most neglected tropical diseases, human African trypanosomiasis (HAT) and Chagas Disease, are caused by protozoan parasites of the genus *Trypanosoma.* These organisms possess specialised metabolic pathways, frequently distinct from those in humans, which have potential to be exploited as novel drug targets. This study elucidates the structure and function of L-threonine-3-dehydrogenase (TDH) from *T. brucei*, the causative pathogen of HAT. TDH is a key enzyme in the metabolism of L-threonine, and a TDH inhibitor has been shown to have trypanocidal activity in the procyclic form of *T. brucei*. TDH in humans is a non-functional pseudogene, suggesting that it may be possible to rationally design safe and specific therapies for trypanosomiasis by targeting this parasite enzyme. As an initial step, the TDH gene from *T. brucei* has been expressed, and the three-dimensional structure of the enzyme has been solved by X-ray crystallography. In multiple crystallographic structures, *T. brucei* TDH is revealed to be a dimeric short-chain dehydrogenase that displays a considerable degree of conformational variation in its ligand binding regions. Geometric simulations of the structure have provided insight into the dynamic behaviour of this enzyme. Furthermore, the structures of TDH bound to its natural substrates and known-inhibitors have been determined, giving an indication of the mechanism of catalysis of the enzyme. Collectively, these results provide vital details for future drug design to target TDH or related enzymes.

Synopsis

Relationships between the structure and function of a putative drug target for human African trypanosomiasis are elucidated using X-ray crystallography and Geometric Simulations. The data presented provide insight into ligand binding and catalysis, giving direction to future rational drug design.

Keywords

Protein crystallography; structural biology; threonine metabolism; trypanosomiasis; geometric simulations

Abbreviations

ADH, alcohol dehydrogenase; AKB, 2-amino-3-ketobutyrate; Ac-CoA, acetyl-coenzyme A; BSF, bloodstream form; DLS, Diamond Light Source; ESRF, European Synchrotron Radiation Facility; GalE, UDP-galactose 4'-epimerase; HAT, human African trypanosomiasis; KBL, 2-amino-3-ketobutyrate; NTD, neglected tropical disease; PDB, protein data bank; PEG, polyethylene glycol; SEC, size-exclusion chromatography; TDH, L-threonine-3-dehydrogenase.

1. Introduction

Trypanosomiasis is a human and animal disease caused by infection with insect-transmitted parasites of the genus *Trypanosoma*, protozoa distinguished by a single flagellum. In humans, *Trypanosoma brucei* causes human African trypanosomiasis (HAT), which is endemic in sub-Saharan Africa, and *Trypanosoma cruzi* causes Chagas disease, which is widespread in South America. HAT is fatal if left untreated, whilst Chagas disease often results in chronic and life-threatening pathology many years after the initial infection (Büscher *et al.*, 2017; Bern, 2015).

Both HAT and Chagas disease have been designated as Neglected Tropical Diseases (NTDs). These debilitating infections often affect poorer populations within developing countries, and have historically received less attention for pharmaceutical R&D. As a result, there are a limited number of drugs available for both diseases, and these have numerous shortcomings, which include toxicity, limited efficacy, and acquisition of resistance by the pathogens. Developing a more detailed understanding of *Trypanosoma* biochemistry and metabolism should aid the identification of new drug targets.

The main hypothesis of this project is that a genetic difference between trypanosomes and their human hosts is exploitable for drug design. It is known that the insect gut form of *T. brucei* selectively and rapidly metabolises L-threonine by the action of L-threonine 3-dehydrogenase (TDH). Indeed, inhibitors of TDH have been shown to be lethal to the procyclic form of the parasite (Cross *et al.*, 1975; Linstead *et al.*, 1977). More recent work has shown that *TDH* is a pseudogene and non-functional in humans, and that an alternative metabolic pathway exists (Edgar, 2002). This suggests that inhibitors of trypanosome TDH may have potential as highly specific anti-parasitic agents. Recent work using reverse genetics has shown that glucose and threonine contribute almost equally to Ac-CoA production, both being crucial in bloodstream form (BSF) and procyclic form parasites, with concomitant blockade of the corresponding

pathways inducing cell death (Millerioux *et al.*, 2013; Mazet *et al.*, 2013). Hence, we expressed and crystallised *T. brucei* TDH, and derived its crystal structure at high resolution with various physiological ligands bound.

There are three known pathways of threonine catabolism, which are initiated by L-threonine aldolase (EC 4.1.2.5), L-threonine dehydratase (also known as L-threonine deaminase; EC 4.2.1.16) and L-threonine 3-dehydrogenase (TDH; EC 1.1.1.103)., which is the dominant pathway in several prokaryotes and eukaryotes (Bowyer *et al.*, 2009; Marcus & Dekker, 1993*b*; Edgar, 2002). The activity of L-threonine aldolase has been shown to be low or insignificant in prokaryotes (Lam *et al.*, 1980) and eukaryotes (Bird & Nunn, 1983) and no expression of the enzyme has been detected in humans (Edgar, 2005). In contrast, the pathway initiated by L-threonine dehydratase is a major mechanism for L-threonine catabolism and can become the dominant pathway in certain metabolic states in animals (Bird & Nunn, 1983). The other major route for L-threonine catabolism is initiated by the mitochondrial matrix enzyme TDH, and appears to be the dominant pathway in many prokaryotes and eukaryotes (Bowyer *et al.*, 2009; Marcus & Dekker, 1993*b*; Edgar, 2002). In this pathway, TDH works in tandem with a second enzyme, 2-amino-3-ketobutyrate ligase (KBL) to catabolise L-threonine to glycine and Ac-CoA (**Error! Reference source not found.**).

In the first part of this process, TDH oxidises L-threonine to 2-amino-3-ketobutyrate (AKB) in a reaction dependent on NAD(H). There is evidence that AKB is unstable and spontaneously breaks down to aminoacetone and carbon dioxide. Thus it has been suggested, and backed by other evidence, that TDH and

KBL form a multi-enzyme complex to facilitate the completion of the second reaction, the breakdown of AKB to Ac-CoA and glycine by KBL (Dale, 1978; Bowyer *et al.*, 2009).

TDH belongs to an enzyme family that is sub-divided into short-chain (SDR), medium-chain (MDR) and long-chain (LDR) dehydrogenase/reductases (Kavanagh *et al.*, 2008). Several of the most studied TDHs have been MDRs, showing a tetrameric quaternary structure and a requirement for divalent metal cations, such as zinc (Boylan & Dekker, 1978, 1981; Bowyer *et al.*, 2009; Higashi *et al.*, 2005; Ishikawa *et al.*, 2007; Machielsen & van der Oost, 2006). More recently, a second group of TDH enzymes that share features with UDP-galactose 4'-epimerase (UDP-GalE) have been characterised. These enzymes belong to the SDR family, and in contrast to the MDR TDH enzymes, they are dimeric and have no requirement for divalent cations (Yoneda *et al.*, 2010, 2012; Ueatrongchit & Asano, 2011). KBL has not been studied to the same extent, but the crystallographic structure of the enzyme from *Escherichia coli* has been published (Schmidt *et al.*, 2001).

The L-threonine degradation pathway catalysed by TDH and KBL plays a variety of roles in both prokaryotes and eukaryotes, including energy production, homeostasis and fatty acid synthesis. The pathway and its constituent enzymes have been studied in *T. brucei*, but crystal structures of the enzymes are yet to be published.

T. brucei was shown by Cross *et al.* to preferentially and exhaustively consume L-threonine when grown in mixed media (Cross *et al.*, 1975). Furthermore, it was shown that most of the absorbed threonine was catabolised via the TDH pathway, producing Ac-CoA and glycine. To further underline the importance

of this pathway, exposure of cultured *T. brucei* to tetraethyl thiuram disulphide (TETD, marketed as Antabuse), an inhibitor of human aldehyde dehydrogenase and a potent inhibitor of TDH, has been demonstrated to lead to trypanosome death in correlation with the degree of TDH inhibition (Cross *et al.*, 1975; Linstead *et al.*, 1977).

A principal reason for the detrimental effect of TDH inhibition on *T. brucei* is the importance of the pathway for fatty acid synthesis, for which Ac-CoA is a substrate. Millerioux, Mazet and colleagues have demonstrated that Ac-CoA is required for fatty acid synthesis by simultaneously inhibiting TDH and another key Ac-CoA-producing enzyme, pyruvate dehydrogenase, which proved to be lethal for the parasite (Millerioux *et al.*, 2013; Mazet *et al.*, 2013). The physiological role of L-threonine metabolism through the TDH pathway is not limited to fatty acid production and there may be additional functions not yet elucidated. For instance, glycine produced from L-threonine by TDH in *T. brucei* has been shown to be incorporated into trypanothione (Millerioux *et al.*, 2013), an antioxidant analogous to glutathione found in humans. In addition, studies of ^{13}C-labelled glucose metabolism in BSF *T. brucei* showed that glucose-derived acetate was used to acetylate amino acids, and it was suggested that L-threonine-derived acetate plays a role in this process (Creek *et al.*, 2015).

Below, we present the first complete description of the tertiary and quaternary structure of *Tb*TDH, as analysed by X-ray crystallography and a range of other biochemical and computational techniques. The results give insight into the relationship between TDH and its natural cofactor and substrate, and also the relationship between TDH and the subsequent enzyme in the pathway, KBL.

2. Materials and Methods

2.1. Expression and Purification

The genes for TDH and KBL were amplified by PCR from *T. brucei* genomic DNA and cloned into the NdeI and BamHI sites of the *E. coli* expression vector pET15b using standard methods and then expressed in *E. coli*. The protein was then purified from the supernatant solutions extracted from cell lysates using nickel affinity chromatography, employing a HisTrap HP column (GE Healthcare) to isolate His-tagged protein. Purity of protein solutions was determined by SDS-PAGE and by size-exclusion chromatography (SEC). SEC was performed on a Superdex 200 10/300 GL column (GE Healthcare), testing 200-400μl samples using Fast Performance Liquid Chromatography (FPLC) (Pharmacia Biotech) at a flow rate of 0.5ml/min and a maximum pressure of 1.5 MPa. Before any experimental samples were run, the column was calibrated according to the manufacturer's instructions with the standards. Protein elution volumes (V_e) were determined by following UV absorbance at 280 nm (A_{280}). The void volume (V_o) of the column was determined by running a sample of Blue dextran (MW 2000 kDa), which is not retained by the resin, due to its size. TDH and KBL were identified by using a standard calibration curve to predict the molecular weights of proteins eluting at a particular elution volume, with confirmation by SDS-PAGE.

Where it was desired to improve the purity of the protein solutions after affinity chromatography, ion-exchange chromatography (IEC) was employed. Appropriate conditions for IEC were determined by using the ProtParam tool on the ExPASy server to predict the isoelectric points of TDH and KBL from

their amino acid sequences. IEC was performed on a 250ml Q-sepharose column using an ÄKTA Prime (GE Healthcare) FPLC machine. Protein was purified by eluting with a gradient of 1000ml solvent progressing from low salt buffer (100 mM NaCl) to high salt buffer (1 M NaCl).

TDH and KBL solutions were stored in ice baths and refrigerated at 2-8°C. The proteins could be stored in this way for at least one month without an appreciable loss in function. Alternatively, solutions were stored for longer periods, frozen at -20°C, and loss of function was not observed after several months of freezing the proteins.

2.2 .X-ray crystallography

2.2.1. Crystallisation

Proteins were crystallised by use of the hanging drop method, using commercial crystallisation screens or manually prepared solutions. All details are provided in Table 1.

To prepare co-crystals of the proteins with ligands, two methods were attempted: substrates, inhibitors and metal ions were incorporated into the crystallisation solution and protein-ligand complexes were co-crystallised using the method described above, or alternatively, crystals were soaked in solutions containing additional substances of interest for periods of 5 minutes to 1 hour. Only the first co-crystallisation method yielded co-crystals of ligand-bound protein.

Prior to data collection, crystals were harvested using a fabric or plastic loop and cryo-cooled by one of three methods: the loop and crystal were plunged directly into a bath of liquid nitrogen, the loop and crystal were suddenly exposed to a stream of cryo-cooled nitrogen gas or the loop and crystal were plunged into liquid ethane for one second and then into liquid nitrogen. As cryo-cooling presents the risk of causing damage to crystals, cryoprotectants such as glycerol, ethylene glycol or PEG 400 were sometimes added. In these cases, crystals were transferred to a 10 µl drop of crystallisation well solution and four 1 µl drops of cryoprotectant were sequentially mixed into this solution before the crystal was cryo-cooled as described above.

X-ray data were collected at synchrotron radiation sources, the European Synchrotron Radiation Facility (ESRF) and Diamond Light Source (DLS). Early data processing was either carried out manually or by use of the integrated data reduction pipeline, xia2 (Winter, 2010), at DLS. When early data processing was carried out manually, data integration was carried out using iMosflm (Battye *et al.*, 2011). When automated processing was carried out by xia2, the program XDS (Kabsch, 2010) was used. Space group determination was initially attempted by selecting from iMosflm-provided options, or it was carried out automatically using POINTLESS (Evans, 2005, 2011). Scaling and merging were carried out using Scala (Evans, 2005) or Aimless (Evans & Murshudov, 2013), together with ctruncate (Stein & Ballard, 2009) to convert the intensities into structure factors. As an analogous TDH structure was already available in the PDB, phase determination for all TDH structures described herein was carried out by means of molecular replacement, using Molrep (Vagin & Teplyakov, 1997, 2009) or Phaser (McCoy *et al.*, 2007). In order to determine the appropriate number of molecules per asymmetric unit, Matthews_coef (Kantardjieff & Rupp, 2003) was used. Refinement was performed using Refmac5 (Murshudov *et al.*, 1997; Vagin *et al.*, 2004). To incorporate ligands into the model, their structures were produced as .pdb files, along with restraints (stored in .cif files, used by Refmac), using the online ProDRG server

(Schüttelkopf & van Aalten, 2004). Once positioned correctly, in accordance with its associated electron density, ligand files were merged with protein .pdb files.

The structure visualisation and manipulation software, Coot (Emsley & Cowtan, 2004; Emsley *et al.*, 2010), was used on Windows and Linux systems to carry out real space refinement and validation of structures. Manipulation of structures was guided by 2Fo-Fc maps at σ levels of 1.5 and by Fo-Fc maps at σ levels of 2.00. Similarly, solvent molecules were identified by placing them within unoccupied volumes of electron density and retaining molecules with accurate fits to the electron density after refinement.

To validate models, bond distances and geometries, including those in hydrogen bonding networks, were observed visually and modified accordingly. The Ramachandran plot function in Coot was used to find and correct errors in the phi (φ) and psi (ψ) angles of amino acid residues. Coot's rotamer analysis tool was also used to correct side chain orientations in TDH structures.

2.3 Structural Analysis

2.3.1. Sequence Analysis

Nucleotide and amino acid sequences from *Tb*TDH (Uniprot Q7YW97) were compared with those of TDH from other organisms using the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, 1990). Certain evolutionarily conserved sequences were also identified in the 2D sequences and visually using Coot and other visualisation software, such as UCSF Chimera (Pettersen *et al.*, 2004).

2.3.2. 3D structure

Protein structure models and their interactions with ligands were observed visually using PyMOL (Schrödinger LLC), Coot and UCSF Chimera (Pettersen *et al.*, 2004). To analyse conformational variability evident in various *Tb*TDH structures, eight representative monomers from solved structures were compared in a method similar to Huntington (Huntington, 2008). The CCP4 program baverage was used to calculate average *B*-factors per residue for all main chain atoms. UCSF Chimera was used to visualise the mean *B*-factors and the RMSDs calculated between α-carbons of the eight structures when superimposed.

The program PISA (Protein Interfaces, Surfaces and Assemblies) (Krissinel & Henrick, 2007) in the CCP4 package was used to output the quaternary structures of TDH models for further analysis.

2.3.3. Geometric simulations of flexible motion

Amplitudes of motion in models of TDH were simulated using a combination of rigidity analysis and coarse-grained elastic network normal mode analysis, as described previously (Jimenez-Roldan *et al.*, 2012; Wells *et al.*, 2015). Elnemo software was used (Suhre & Sanejouand, 2004) to obtain normal mode eigenvectors from coarse-grained elastic network modelling, and FIRST/FRODA software (Jacobs *et al.*, 2001; Wells *et al.*, 2005) to carry out rigidity analysis (FIRST) (Hespenheide *et al.*, 2002), identifying the noncovalent interaction network and labelling dihedral angles as locked or variable, and template-based geometric simulations of flexible motion (FRODA) (Wells *et al.*, 2005) which project the all-atom structure over large amplitudes of motion while maintaining local bonding and steric geometry.

Normal mode eigenvectors were generated in Elnemo in a one-site-per-residue coarse-graining using the Cα geometry of the input structure, placing springs of equal spring constant between all sites lying within an interaction distance cut-off of 12 Å. A rigidity analysis of the all-atom input structure was carried out in FIRST using the "pebble game" algorithm (Jacobs *et al.*, 2001; Jacobs & Thorpe, 1995) which matches degrees of freedom against bonding constraints in the molecular framework of the protein. Bonding constraints include covalent, hydrophobic and polar (hydrogen bond and salt bridge) interactions. As the strength of the polar interactions can be gauged from their geometry, the results of the analysis depend on an "energy cut-off" which selects the set of polar interactions to include in the constraint network (Hespenheide *et al.*, 2002). A cut-off of -2.0 kcal/mol was used in this study.

Template-based geometric simulation of flexible motion, carried out using FRODA, explores the mobility of the all-atom structure by iterative perturbation and relaxation of atomic positions in parallel and antiparallel to the direction of normal-mode eigenvectors. Several thousand iteration steps are carried out to generate large motion amplitudes. The simulation generates an initial phase of "easy" motion, where the bonding geometry is easily maintained, followed by the onset of "jamming" as the motion encounters steric and bonding constraints which naturally limit its amplitude. The conformational changes of geometric simulations of TDH projected using this method were observed and compared with superpositions of crystallographic models in PyMOL and UCSF Chimera.

2.4. Other biochemical techniques

2.4.1. Size-exclusion chromatography

SEC was used as described above to determine the oligomeric states of TDH and KBL. Before any experimental samples were run, the column was calibrated with the standards: β-amylase (MW 200 kDa), alcohol dehydrogenase (MW 150 kDa), bovine serum albumin (BSA) (dimer MW 132 kDa;

monomer MW 66 kDa), ovalbumin (MW 45 kDa), carbonic anhydrase (MW 24 kDa) and cytochrome C (12.4 kDa). To estimate the molecular weights of

proteins eluting at a particular elution volume (V_e), a plot of $\log_{10}$ molecular weight of standard protein against the elution volume: void volume ratio (V_e/V_0)

was made using Microsoft Excel. A number of experiments testing TDH and KBL in isolation, together and in the presence of substrates were carried out to

measure changes in the observed elution volumes and oligomeric states predicted by linear regression analysis.

2.4.2. Cross-linking

Cross-linking studies of TDH and KBL were carried out using a method similar to that of Davies and Stark (Davies & Stark, 1970), which employed dimethyl

suberimidate (DMS) as the cross-linking agent. As cross-linking reactions predominate within oligomers, this was used to provide an indication of the

oligomeric states of TDH and KBL. Protein concentrations and DMS concentrations were carefully chosen to avoid precipitation.

All solutions were made up to 50μl or 100μl in 200mM Tris.HCl buffer at pH 8.5. Final protein concentrations were approximately 1mg/ml. Reactions were

carried out in duplicate, at room temperature and at 2-8°C. The samples were incubated for 3 hours or overnight and the reactions were then halted by the

addition of Laemmli buffer in preparation for analysis by SDS-PAGE. Gels of 9% polyacrylamide, rather than the usual 12%, were used to increase the

migration of higher molecular weight complexes.

3. Results

3.1. Structure of TDH

3.1.1. Crystal growth and Data Collection

Diffracting TDH crystals were successfully grown under several different conditions, which are listed in Table 1, and data collection and refinement statistics for the related data sets are listed in Table 2. Diffracting TDH crystals were cryo-cooled using glycerol, PEG 400 or ethylene glycol. It was also possible to collect quality data from crystals cooled in the absence of cryoprotectant, as exemplified by the data used to solve two structures (PDB: **5K4W** and PDB: **5K4Y**).

3.1.2. *Tb*TDH primary and secondary structure

TDH primary sequence alignments showed that the TDH enzymes from *T. brucei* strain TREU 927 and *T. brucei gambiense* are essentially identical, confirming that the TDH studied in this work is relevant to a human pathogen. The animal pathogen *T. congolense* has a protein that shares 78% sequence identity with the TDH studied here. The sequence identity between TDH from *T. brucei* strain TREU 927 and the Chagas disease-causing pathogen, *Trypanosoma cruzi*, is 72%. Thus, TDH from a range of *Trypanosoma* species are closely related (Supplementary Figure 1).

Other TDH enzymes with a high sequence similarity to *Tb*TDH include the UDP-galactose 4'-epimerase-like (GalE) TDH enzymes from *Thermoplasma volcanium, Flavobacterium frigidimaris* and *Cupriavidus necator*. The corresponding enzymes from these organisms have monomeric molecular weights between 35 and 37.2kDa, matching that of TDH from *T. brucei* (37.8 kDa). Another group of related TDH enzymes originate from mammalian and avian

species: *Rattus norvegicus* (rat), *Sus scrofa* (boar/pig), *Capra hircus* (goat) and *Gallus gallus* (red junglefowl). TDH enzymes from these species have been studied previously and shown to have molecular weights in the range of 36 kDa (Yuan & Austic, 2001) to 37 kDa (Kao & Davis, 1994).

Inspection of the sequences of the various TDHs shows that they belong to the SDR superfamily of enzymes. Two highly conserved structural features that can be identified from the sequence are highlighted in Supplementary Figure 2. The first, a glycine-rich region, is represented by the sequence GxxGxxG, and the second consists of a tyrosine and a lysine residue separated by three other residues: YxxxK.

The X-ray structures of *Tb*TDH reported here revealed that each individual subunit consists of two domains displaying α/β topology. The distribution of secondary structure elements in TDH appears to be approximately 46% α-helix and 15% β-sheet.

3.1.3. *Tb*TDH tertiary structure

*Tb*TDH encompasses an NAD-binding domain, which includes the first 180-200 amino acids, and a catalytic domain at the C-terminal end of the protein (Supplementary Figure 3). In three dimensions, there is some crossover between these domains. In the N-terminal domain there is a Rossmann fold motif (parallel β-strands flanked by α-helices), which is characteristic of NAD-binding proteins, and hence SDRs. The structural motif GxxGxxG, which is common to the 'extended' class of SDRs (eSDR), is located within the NAD binding region and is represented by the sequence Gly9-Ala10-Leu11-Gly12-Gln13-Ile14-Gly15. This motif is involved in binding the diphosphate group of NAD. The YxxxK motif and several other conserved residues, including Met81, Ser82, Thr119, Thr186, Trp280, and Tyr144 and Lys148 of the SDR YxxxK motif are located in the active site.

When comparing the different crystallographic models, which were obtained with different ligands and crystallised in different space groups, no large domain movements were observed. An analysis of the average *B*-factor for each residue revealed the presence of high B-factors for some residues, supporting the notion that those regions of the protein are more disordered than others, and thus more flexible*.*

A number of regions in *Tb*TDH appear to be flexible (**Error! Reference source not found.**). Firstly, there are two disordered loop regions: Loop 1 includes residues Thr179 and Ala185, and Loop 2 lies between residues Asp35 and Asn60. There also appears to be a higher degree of flexibility in the catalytic domain (gold and red-coloured region on the right-hand side of **Error! Reference source not found.**a). The most striking evidence of conformational variability is found in the residues of Loop 1. Indeed, in different crystallographic structures, this loop was found to occupy two distinctive positions: one where the loop was in an 'open' position, lying at a distance of approximately 5 Å from the active site, and another where the loop was 'closed' and positioned adjacent to the active site (**Error! Reference source not found.**). Therefore, for each model of a TDH monomer, the structure can be classified as 'open' or 'closed', on the basis of the conformation of Loop 1. Comparing the average *B*-factors in the open and closed structures highlighted changes in flexibility or disorder that accompanied the change in the Loop 1 conformation. For example, Figure 4b shows the average *B*-factors of TDH residues in 'open' structures, whilst Figure 4c shows the average *B*-factors for the 'closed' structures.

In the 'open' structures, the highest *B*-factors are seen around both loop regions. However, in the closed structures, although these same residues have above average *B*-factors, they are much less disordered. Inspection of **Error! Reference source not found.**c demonstrates that the catalytic domain appears to be more disordered in closed structures. In three dimensions, this region can be seen as one of a pair of adjacent α-helices and β-strands, which are the gold-coloured regions on the right hand sides of the structures in **Error! Reference source not found.**a and **Error! Reference source not found.**a.

The NAD-binding region, particularly around Loop 2, also appears to be flexible. Although the B-factors are low, the RMSDs around residues 82 to 100 are particularly high (**Error! Reference source not found.**b). This region lies at the dimerization interface (discussed below), indicating some conformational variability there.

These findings indicate that there may be a relationship between the conformation of Loop 1, and the flexibility of other regions of the protein, particularly the catalytic domain. Although a variety of different crystallisation conditions were used, the different conformational changes do not appear to correlate with the composition of the crystallisation mixture, the use of a cryoprotectant or the space-group of the crystal (

	Model								
RCSB PDB ID	**5L9A**	**5LC1**	**5K4Q**	**5K4T**	**5K4V**	**5K4U**	**5K50**	**5K4W**	**5K4Y**
Crystallisation Conditions									
Conditions	0.1 M HEPES; 20 % w/v PEG 10K; pH 7.5; TDH (5.28 x 10^{-2}mM [2.0 mg/ml])	0.1 M HEPES; 20 % w/v PEG 10K; pH 7.5; TDH (5.28 x 10^{-2}mM [2.0 mg/ml]); NAD$^+$ (1mM); pyruvate (30mM)	0.2M lithium sulphate; 0.1M Tris; 30% w/v PEG 4K; pH 8.5; TDH (6.4 x 10^{-2}mM [2.4mg/ml]); KBL (3.8 x 10^{-2}mM [1.8mg/ml]); NAD$^+$ (0.09mM); PLP (0.18mM); L-Threonine (7.33mM)	0.1M MES; 0.2M ammonium sulphate; 30% w/v PEG 5K MME; pH 6.5; TDH (9.0 x 10^{-2}mM [3.4mg/ml]); L-serine (50mM)	0.1M tri-sodium citrate; 30% w/v PEG 4K; 0.2M ammonium acetate; pH 8.5; TDH (3.7 x 10^{-2}mM [1.4mg/ml]); NAD$^+$ (10mM); BPOB (0.5mM)	0.1M tri-sodium citrate; 30% w/v PEG 4K; 0.2M ammonium acetate; pH 5.6; TDH (3.7 x 10^{-2}mM [1.4mg/ml]); NAD$^+$ (8.5mM); quinine (2.5mM [hydrochloride dehydrate])	0.1M tri-sodium citrate; 30% w/v PEG 4K; 0.2M ammonium acetate; pH 5.6; TDH (10.4 x 10^{-2}mM [3.9mg/ml]); NAD$^+$ (10mM); L-*allo*-threonine (30mM)	0.2M sodium acetate; 0.1M Tris; 30% w/v PEG 4K; pH 8.5; TDH (9.3 x 10^{-2}mM [3.5mg/ml]); NADH (10mM); L-Threonine (30mM)	0.1M tri-sodium citrate; 25% w/v PEG 4K; 0.2M ammonium acetate; pH 5.6 TDH (3.7 x 10^{-2}mM [1.4mg/ml]); NAD$^+$ (10mM); methylglyoxal (8mM)
Cryoprotectant	Glycerol	Glycerol	Glycerol	Glycerol	Glycerol	Glycerol	Glycerol	None	None

Table 2 and Table 2). Thus, the specific conformation of Loop 1 and the flexibility of the catalytic region may actually be influenced by ligand binding. Indeed, three of the four 'closed' structures (from PDB: **5K4U**, PDB: **5K4W** and PDB: **5K50**) examined were bound to NAD/NADH and another ligand, whilst only one of the four 'open' structures (from PDB: **5K50**) examined possessed bound NAD and another ligand. Despite the evident changes in disorder that may be associated with ligand binding and catalysis, the crystallographic structures did not capture any significant changes in the relative positions of residues important in NAD and L-threonine binding.

3.1.4. *Tb*TDH quaternary structure and complex formation

TDH was shown to exist as a dimer following refinement of all the X-ray crystallographic models. The dimerization interface covers a surface area of approximately 1200-1300 $Å^2$ and consists of two α-helices (Pro89-Tyr110 and Val143-Tyr162) and a loop (Lys126-Thr142), with the same regions on each monomer making contact in the dimer. As can be seen in **Error! Reference source not found.**, the dimer is stabilised by a large number of hydrophobic, polar and ionic interactions. In particular, polar side chains of residues on each monomer form hydrogen bonds with each other and charged residues form salt bridges. Interactions across this interface include the following: Glu86 with Trp157; Asp90 with Arg101; Asp94 with Arg101; Lys126 with Asp135. This probably gives greater stability to the dimer, whilst the length of some of the side chains allows their interactions to be maintained in the case of any change in the orientation of the two monomers.

Further evidence of the existence of the dimer and that this is the largest oligomer present in all of the crystallographic structures, was provided using the software PISA (Krissinel & Henrick, 2007). Supplementary Table 1 lists a number of different interfaces between different TDH subunits in the different structural models and it can be seen that the only stable association predicted by PISA is the dimer relationship previously described. There is no clear relationship between the crystal space group and the nature of the dimerisation interface.

3.2. Oligomerisation studies

The cross-linking studies carried out using dimethyl suberimidate (DMS) provide more evidence in regard to the oligomeric states of TDH, and the subsequent enzyme in this threonine degradative pathway, KBL. SDS-PAGE analysis of the samples revealed protein bands representing monomeric and dimeric forms of TDH and KBL (Supplementary Figures 6-8). A single two-fold dilution of the concentration did not have an effect on the apparent oligomeric states detected by the cross-linking experiments. Experiments conducted with TDH and KBL together do not suggest that the two proteins are complexed in solution, as there are no additional bands present (Supplementary Figure 6, lanes 7 and 8), compared to when samples of the two proteins were tested separately (Supplementary Figure 6, lanes 1, 2, 4 and 5). The pull-down assay using His-tagged KBL and native TDH was in agreement with the data from the cross-linking studies, suggesting that TDH and KBL do not appear to form a complex with each other in solution. When His-tagged KBL was applied to the Ni-NTA column, non-tagged TDH washed through with the wash buffer. Meanwhile, KBL was retained on the column, and elution with the 250mM imidazole elution buffer caused KBL to elute alone, demonstrating that the two proteins had not bound to each other under the conditions tested (Supplementary Figure 9).

Size-exclusion chromatography experiments (SEC) experiments initially reinforced the results described thus far: when assayed separately, elution volumes of TDH and KBL indicated that they exist in dimeric forms in solution (Supplementary Table 2). The TDH results were therefore consistent with the findings gained by X-ray crystallography. Interestingly, when mixed and assayed together, TDH and KBL co-eluted predominantly at a volume suggesting the existence of the proteins as monomers. Miniscule amounts of both proteins eluted at a volume corresponding to a large oligomer of approximately 188 kDa , but the low protein concentrations make this observation

inconclusive. Overall, the SEC findings confirm the dimeric structure of TDH. Additionally, an indication of potential interactions between TDH and KBL was obtained, however further study is required to confirm the dissociation of TDH and KBL dimers or the existence of a TDH:KBL complex.

3.3. Structural insights into ligand binding and catalysis

To gain a complete picture of the TDH structure as it relates to function, various ligands, including the natural substrates, were co-crystallised with TDH prior to obtaining diffraction data. Crystallographic models containing NAD(H), L-threonine, L-allothreonine and pyruvate, were all solved with those ligands at an occupancy of 1. As can be seen in **Error! Reference source not found.** and **Error! Reference source not found.**, the TDH substrates, NAD^+ and L-threonine, occupy a deep cleft in the protein. The L-threonine binding pocket and the pocket occupied by the nicotinamide group of NAD^+ are partially separated by the side chain of Met81 (Figure 7).

NAD binds TDH through a combination of hydrogen bonds and hydrophobic interactions with several main chain and side chain atoms of TDH residues. The interactions with important TDH residues are displayed in **Error! Reference source not found.**b. In addition, the cofactor also hydrogen-bonds with several water molecules. This is particularly the case for the phosphate groups, the ribose hydroxyl groups and the nicotinamide amide (**Error! Reference source not found.**a). Three-dimensional analysis shows that NAD's nicotinamide and the adjacent ribose groups bind in deep pockets of the protein, whilst the remainder of the NAD molecule faces the exterior. Additionally, unoccupied pockets can be observed adjacent to the nicotinamide amide and C2 of the adenine ring (see Supplementary Figure 10), suggesting molecules of a different size and shape could be accommodated into the same binding site.

To gain insight into L-threonine binding, TDH was co-crystallised with it and the reduced form of the cofactor (NADH) to prevent turnover, which involves oxidation of the amino acid. The L-threonine-bound structure (PDB: **5K4W**) at 1.72 Å resolution (**Error! Reference source not found.**) reveals that the carboxyl group of L-threonine hydrogen-bonds with the side chain and main chain atoms of Ser82 and Thr186. The amine group of L-threonine is seen to hydrogen-bond with the side chain hydroxyl group of Thr186, with the nicotinamide carbonyl of NADH and with a nearby water molecule. The side chain hydroxyl group of L-threonine is hydrogen-bonded to Thr119 and Tyr144, whilst the methyl group is directed towards the hydrophobic amino acid Trp280. This aligns the β-carbon with the C4 atom of the nicotinamide ring (discussed in more detail below). Another interesting feature of this structure was that the flexible loop, Loop 1, is closed over the active site, meaning that it is completely shielded by the enzyme (**Error! Reference source not found.**c). This was the case in both TDH monomers in the asymmetric unit of the structure, both of which were bound to NADH and L-threonine.

X-ray crystallographic structures of TDH bound to compounds reported to be inhibitors, L-*allo*-threonine (Klein *et al.*, 1980) and pyruvate (Kazuoka *et al.*, 2003; Yoneda *et al.*, 2012), were also successfully solved. These compounds were shown to occupy the L-threonine binding site, forming interactions with some of the same active site residues as the natural TDH substrate, thereby demonstrating their modes of inhibition (Supplementary Figure 11).

The putative mechanism of reaction of TDH involves transfer of a hydride from the β-carbon of L-threonine to the C4 atom of the nicotinamide ring (Yoneda *et al.*, 2010). The data presented here confirm that the atoms involved in this reaction are well aligned and at a distance that would permit such a reaction to happen (**Error! Reference source not found.**). X-ray structures described herein confirmed the presence and interaction of key putative catalytic residues, such as Thr119, Tyr144 and Thr186 (Figure 7). Another structural feature that is conserved in SDRs and deemed important for catalysis is a chain of water molecules that is believed to form a proton relay that transfers the

proton extracted from L-threonine to the bulk solution, allowing the enzyme to catalyse another reaction. This water chain requires a certain orientation of residues to bind the water molecules. The feature observed in several SDRs is an interaction between an asparagine side chain and the main chain of another residue. In this case, the interaction between Asn96 and Ile80 (**Error! Reference source not found.**a) causes a kink in the α-helix spanning residues Pro89-Lys109 and changes the orientation of the Asn96 carbonyl group. This allows the Asn96 carbonyl group to hydrogen-bond one of the water molecules in the water chain (**Error! Reference source not found.**b).

The sidechain of Lys129 in *Tb*TDH hydrogen-bonds with Asn282 and may have a stabilising role in this region of the enzyme, which is located on the exterior, behind the active site (Supplementary Figure12). Lys129 corresponds to Arg180 in the homologous murine TDH (Supplementary Figure 12). An R180K variant of this enzyme has been shown to have reduced thermostability and impaired recognition of L-threonine , as indicated by reduced catalytic efficiency (He *et al.*, 2015). Inspection of the structure suggests that replacement of arginine with lysine in *Tb*TDH and bacterial TDH probably results in relatively fewer interactions with surrounding residues, and thus, a weaker stabilising effect.

3.4. Other features observed in crystallographic models

A number of molecular species from the storage buffer and from the crystallisation solutions were found to co-crystallise with TDH; glycerol, acetate, sulphate and sodium ions were all observed in TDH structure models. These molecules occupied various positions, and a few positions were occupied in a consistent manner and observed in multiple structures. Most notably, glycerol and acetate bound within the L-threonine binding site (Supplementary Figure 13). Glycerol was also found to bind to the same regions in different TDH models (Supplementary Figure 14).

Typically, different TDH dimers in each of the solved structures are separated from other monomers by an ordered array of water molecules. However, one direct interaction between two monomers in separate dimers was observed in one model (PDB: **5K50**). Interestingly, the interaction is formed

between residues of the flexible Loop 2 region (Supplementary Figure 15). The interaction does not appear to be strong enough to be an important physiological interaction and is likely to be a result of the crystallisation process.

3.5. Geometric Simulations

3.5.1. Noncovalent constraint identification using FIRST
We have examined the dimeric crystal structures using FIRST, which identifies noncovalent polar and hydrophobic interactions and assesses their effect in rigidifying the structure. FIRST identifies a well-formed network of hydrophobic tethers, hydrogen bonds and salt bridges in the dimer interface region, consistent with the PISA analysis. These interactions do not, however, make the dimer into a rigid unit; instead, flexibility remains, consistent with the observed differences between the *apo, holo*, and threonine-bound crystal structures. The Loop 1 region appears not to be constrained by any noncovalent interactions; this is consistent with the crystallographic observations of flexibility in this region.

3.5.2. Low-frequency flexible motion identified using Elnemo and FIRST/FRODA
We have identified low-frequency mode eigenvectors using coarse-grained elastic networking with the Elnemo software, and we have carried out all-atom simulations of flexible motion using the FRODA geometric simulation software (incorporated in FIRST) to explore motion biased along these low-frequency mode directions. The most intriguing large-scale motion we identify is that along the lowest-frequency nontrivial mode of motion (Elnemo's mode 7: modes 1-6 are trivial rigid body motions). This motion is predominantly a twisting motion about an axis running perpendicularly through the dimer interface, i.e. as if the dimer were being twisted about its long axis. The motion is easily permitted by the constraints in the interface region, with large amplitudes of motion (several Å RMSD) being achieved in the geometric simulations; the long flexible sidechains which form the

inter-dimer interactions easily accommodate the motion. The character of this motion is essentially identical for the *apo*, *holo* and L-threonine-bound forms.

This twisting motion appears to account for most of the global differences observed between the *apo*, *holo*, and L-threonine-bound forms of the dimer. The *holo* and L-threonine-bound forms can be superimposed on the *apo* form with fitted backbone atom RMSDs of 1.13 Å and 1.58 Å, respectively. However, the geometric simulation of motion of the *apo* form biased along the lowest-frequency nontrivial mode generates a conformation which resembles the *holo* form more closely, with a fitted backbone atom RMSD of 0.49 Å, and, shortly thereafter, a conformation which resembles the L-threonine-bound form, with a fitted backbone atom RMSD of 0.63 Å. These superpositions are shown in **Error! Reference source not found.**. Simulations beginning from the *holo* and L-threonine-bound forms (not shown) give consistent results and visit states closely resembling the *apo* form.

3.5.3. Random (diffusive) motion in the flexible Loop 1 region

To assess the freedom of motion in Loop 1, which as noted is not restrained by any identifiable noncovalent interaction, we have carried out unbiased geometric simulations of random diffusive motion in the protein starting from the *apo* state. In one such simulation, the residues making up Loop 1 were observed to drift from their initial "open" state to a much more "closed" state, as shown in in Figure 10. This confirms that the loop has the conformational freedom required to visit both open and closed states regardless of the binding of substrate, and is suggestive of a conformational-selection mechanism rather than a directly induced-fit mechanism of active site closure by Loop 1.

4. Discussion

4.1. *Tb*TDH as a GalE-like TDH

*Tb*TDH is a GalE-like TDH, sharing many features with the enzymes from *F. frigidimaris* (*Ff*TDH), *C. necator* (*Cn*TDH) and *T. volcanium* (*Tv*TDH) (Kazuoka *et al.*, 2003; Yoneda *et al.*, 2010; Ueatrongchit &

Asano, 2011; Yoneda *et al.*, 2012). We have shown that *Tb*TDH possesses features that are conserved amongst these enzymes, as well as other enzymes within the SDR enzyme superfamily: a monomeric length of around 320 residues, a glycine-rich region (GxxGxxG) within the NAD-binding domain, a structural motif including lysine and tyrosine (YxxxK) near the active site, and several other conserved active site residues (**Error! Reference source not found.**). These features contrast with TDH in other species, such as *E. coli*, *Pyrococcus horikoshii* and *Thermococcus kodakarensis*, which are homotetramers from a different class of enzyme, rather than homodimers, and which specifically bind divalent metal cations such as zinc (Johnson *et al.*, 1998; Higashi *et al.*, 2005; Bowyer *et al.*, 2009).

Another distinguishing feature of GalE-like TDHs is the presence of a loop adjacent to the active site which is directly followed by a motif represented by the sequence GTTDY (Yoneda *et al.*, 2012). In *Tb*TDH, this 'flexible Loop 1' is followed by a similar sequence, GATDY, but it is unclear if this confers any structural or functional difference between these and the other known GalE-like TDHs.

We have provided multiple lines of evidence that *Tb*TDH, like its aforementioned analogues, exists in dimeric form in solution. In SEC studies, we observed a change in elution volumes which suggested that TDH and KBL may interact in solution, possibly leading to the dissociation of homodimers and the potential formation of a multi-enzyme complex. The formation of a TDH:KBL complex has been explored elsewhere (Bowyer *et al.*, 2009; Dale, 1978; Marcus & Dekker, 1993*a*; Jamil, 2012), but it has not been confirmed for a GalE-like TDH and it remains a subject of ongoing investigation for *Tb*TDH.

Overall, the structural characteristics observed place *Tb*TDH in the 'extended' category of SDRs (eSDR), and thus the enzyme is assigned the identifier SDR14E under the SDR nomenclature initiative (Persson *et al.*, 2009). Details of other SDR-type TDH enzymes can be found along with their PDB accession codes in Supplementary Table 3.

4.2. *Tb*TDH conformational flexibility

4.2.1. Conformational variability

Figures 3, 4 and 9 indicate the various regions of flexibility and dynamic behaviour that *Tb*TDH may possess in solution. Comparisons of different X-ray crystallographic structures and the geometric simulations show that global motions in the TDH homodimer consist of a combination of hinge and shear motions, where the two subunits appear to twist in opposite directions about the dimer interface. These conformational variations are likely to predominate, as exemplified by the fact that the models arising from various geometric simulations can be superimposed over crystallographic structures of both *apo-* and *holo-TbTDH*. Other important conformational changes may involve alterations in the relative positions of the NAD-binding and catalytic domain.

Recent studies on other dimeric enzymes (Wells *et al.*, 2015, 2014) have shown that low-frequency motions can be intimately involved in the opening and closing of enzyme active sites. In the case of *Tb*TDH, however, the lowest-frequency motion gives a good description of the global differences among the *apo/holo*/threonine-bound series of structures, but the motion has no particular effect on the active site geometry. It is therefore plausible that the different degrees of twisting in the different dimer structures represent states from a broad conformational ensemble of flexible variation, selected by crystallisation conditions, i.e. the twisting motion may not be directly related to the enzyme function.

The most obvious regions of conformational variability in *Tb*TDH are the flexible Loops 1 and 2. Whereas it is unclear whether Loop 2 has any functional role, it appears that Loop 1 may have an important role in substrate binding. We will discuss this in more detail below.

4.3. Substrate binding and catalytic Mechanism of Action

The NADH- and L-threonine-bound structure presented here (**Error! Reference source not found.**) is only accompanied in the literature by one other threonine-bound wild-type TDH structure (Nakano *et al.*, 2014). Both structures provide valuable insights into the proposed mechanism of action of TDH, in which a hydride is transferred from the β-carbon of L-threonine to the C4 atom of the nicotinamide ring of NAD^+; both L-threonine and the co-factor are bound in the necessary orientation for catalysis. Furthermore, the nicotinamide ring was bound in a *syn* conformation, meaning that the hydride is transferred to the pro-S position on the C4 carbon.

The residues of the "catalytic triad" (Winberg *et al.*, 1999; Gani *et al.*, 2008), which have been described in alcohol dehydrogenases (ADH) and SDRs in general, have been identified as Thr119, Tyr144 and Lys148 in *Tb*TDH. The residues Thr119 and Tyr144 appear to hydrogen-bond the side chain hydroxyl group of L-threonine. The interatomic distances between the oxygen atoms of the relevant groups on L-threonine and Tyr144 also seem to support the hypothesis that Tyr144 acts as a base and withdraws a proton from the substrate. Molecular dynamics simulations in *Drosophila* ADH have demonstrated that this is made possible by the fact that the protonation states of the active site tyrosine and lysine residues are coupled (Koumanov *et al.*, 2003; Winberg *et al.*, 1999; Gani *et al.*, 2008). Extrapolating this theory to the TDH mechanism of action, the lysine residue (Lys148 in *Tb*TDH) extracts a proton from a hydroxyl group on a NAD ribose group, which in turn extracts the proton from the tyrosine residue, thus allowing it to extract the proton from L-threonine. In ADH, this proton is then proposed to be transferred to a chain of eight water molecules, which relay the proton until it is removed from the interior of the enzyme. In this way, the tyrosine residue is deprotonated again and is able to participate in the catalysis of another reaction (Koumanov *et al.*, 2003; Wuxiuer *et al.*, 2012). The interatomic distances between the relevant groups observed in TDH structure models all seem to lend support to this theory. The pH dependence of the protonation state of lysine and tyrosine may also explain why the optimum pH in

*Tb*TDH (unpublished research) and other GalE-like TDHs is around pH 8-9 (Green & Elliott, 1964; Linstead *et al.*, 1977; Aoyama & Motokawa, 1981; Ray & Ray, 1985; Wagner & Andreesen, 1995; Kazuoka *et al.*, 2003; Ueatrongchit & Asano, 2011; Yoneda *et al.*, 2012). A short chain of water molecules close to Lys148 was identified in several of the *Tb*TDH structures obtained here, but a complete chain of eight water molecules, which has been shown to be essential for ADH activity (Wuxiuer *et al.*, 2012), was not observed. It is possible that a shorter water chain is sufficient for activity in TDH. Alternatively, it may be that the conformational state of TDH in which this water chain is complete has not been captured in any of the crystallographic models to date.

Loops analogous to the flexible Loop 1 identified in *Tb*TDH have been highlighted in TDH from other species (Yoneda *et al.*, 2010; Nakano *et al.*, 2014). Indeed, the amino acid sequence corresponding to this loop appears to be a conserved feature among GalE-like TDHs. The studies here provide crystallographic evidence of distinct conformations adopted by Loop 1 which are characterised by a dramatic shift of approximately 5 Å in distance. In the X-ray crystallographic structures of *Tb*TDH, Loop 1 was observed in the 'open' conformation and the 'closed' conformation in a number of contexts, including in cofactor-bound structures, and structures bound to a cofactor plus L-threonine or another ligand.

Although the data are not conclusive, one can hypothesise that Loop 1 can be in an open or closed conformation, regardless of whether a substrate is bound in the L-threonine binding site or not. It is possible that a change to a closed conformation is induced by binding of a ligand to the L-threonine binding site, as was observed with L-threonine- and pyruvate-bound structures. However, one structure (PDB: **5K50**) shows that binding of L-*allo*-threonine did not induce a closed conformation. The observation that different Loop 1 conformations can be adopted in different subunits of the same TDH dimer (as observed in PDB: **5K4U** [Supplementary Figure 16]) suggests that the Loop 1 conformation in each monomer is adopted independently of the conformation in the other subunit.

The random walk geometric simulations of *Tb*TDH indicate that the loop may have some conformational freedom due to a lack of restraints on this motif. Alternatively, there may be a mechanism whereby the conformation of one subunit affects the conformation in the other subunit.

He *et al.* proposed that the residue Arg180 in *Mm*TDH, which corresponds to Lys129 in *Tb*TDH, may act as a remote "switch" between the open and closed forms of TDH (He *et al.*, 2015). Although these residues are not located in regions of conformational variability (Figure 3 and Figure 4), Lys 129 does interact with Asn282, which is in close proximity to the flexible region at residues 255-275. Further investigation will be required to understand if Lys129 plays a role in the stability and catalysis of TDH from *T. brucei*.

Considering the effect that it has on the active site of TDH, it is possible that Loop 1 plays an important role in the mechanism of TDH binding to its ligands. The loop may remain open before and after NAD$^+$ binding to accommodate L-threonine, but then close over the active site once L-threonine is bound. Recently, data from studies of TDH from *C. necator* have also demonstrated that this loop region adopts different conformations in 'open' and 'closed' structures of TDH. Using a combination of crystallographic data from *apo* and *holo* forms of *Cn*TDH, molecular dynamics simulations and enzyme kinetics were used to build a hypothesis of ligand binding and structural changes in this enzyme. Nakano *et al.* proposed a model whereby NAD$^+$ binding caused a rigidification of flexible regions in the NAD-binding domain, but the flexible loop remained open. Then, in the same model L-threonine binding causes Loop 1 to close over the active site, and this accounts for the selectivity of the enzyme for L-threonine (Nakano *et al.*, 2014).

These findings are in some agreement with those presented herein, which suggest that changes in the conformation of Loop 1 and the catalytic domain of *Tb*TDH happen in response to L-threonine binding. The findings of Nakano *et al.* also support the existence of an induced fit mechanism of ligand binding for TDH, as Loop 1 seems to form a "lid" over the L-threonine active site, something that would prevent L-threonine binding if it happened prior to its encounter with the enzyme. Some

researchers assert that all enzymes that form a lid over their active sites must act by induced fit mechanisms (Sullivan & Holyoak, 2008); the findings on TDH structure would seem to illustrate this well. However, some structures presented herein show that *Tb*TDH can adopt a 'closed' conformation in the absence of L-threonine (e.g. PDB: **5K4V** or PDB: **5K4U**; see Table 2). Furthermore, rigidity analysis and geometric simulations of *apo-Tb*TDH suggest that Loop 1 has the conformational freedom to enter the closed state in the absence of ligands. Therefore, in the case of *Tb*TDH, an induced fit mechanism may occur as part of the wider process of conformational selection, which has been proposed as a way in which proteins with variable conformations can adopt the necessary state required to carry out a function (Hammes *et al.*, 2009). Although the flexibility of Loop 1 may be linked to the selectivity of TDH for binding and catalysis of L-threonine, the possibility that the enzyme may sample several different conformations in solution may allow a greater variety of ligands to bind, beyond analogues of the natural substrates. This could have significant implications for drug discovery efforts involving TDH as a target.

4.4 Relevance of findings to future structure-based drug design

The TDH studied here is identical to that from HAT-causing *T. brucei gambiense* and shares high sequence similarity with the Chagas disease-causing *T. cruzi* and animal trypanosomiasis-causing *T. congolense*. Thus, the data presented could be value for human or veterinary trypanosomiasis drug discovery efforts.

We have shown that *Tb*TDH exhibits conformational variability. However, this does not seem to have a significant effect on the relative positions of important substrate binding residues, providing a useful basis for the rational design of TDH inhibitors.   In contrast, the flexibility of Loop 1 has a profound effect on the nature of the L-threonine binding site, and its conformational variability would be an important consideration during the design of an inhibitor targeting that site. In one respect, the closure of Loop 1 is likely to provide greater specificity of TDH for L-threonine or for a given inhibitor. However, it restricts access to the active site, which could be a barrier to the binding of larger inhibitors or those that target the TDH-L-threonine complex.

The NAD binding site consists of a deep cleft, where several hydrophobic and hydrophilic residues can be targeted for binding. The binding pocket is larger than the NAD molecule itself, providing pockets and additional residues that could be exploited to increase the affinity of a designed ligand. Furthermore, our data show that an inhibitor designed to simultaneously bind within the NAD and L-threonine binding pockets is feasible and could be useful strategy for achieving high specificity, though the presence of Met81 may limit the extent to which the proximity of the two binding sites can be used.

4.5 Conclusion

In this study, the structure of *Tb*TDH has been described in detail, providing insights into the dynamic behaviour of the enzyme and how this may relate to its function. The relationship between the enzyme and its natural substrates has been highlighted, and the possible interaction between TDH and KBL has been explored. This new information will contribute to the interpretation of further functional data on TDH, which could in turn help to guide drug discovery efforts. Furthermore, these findings add to the knowledge-base of the related enzymes of the SDR family.

Accession numbers

Coordinates and structure factors for all structures cited herein have been deposited in the Protein Data Bank with the following accession numbers: **5L9A**; **5LC1**; **5K4Q**; **5K4T**; **5K4V**; **5K4U**; **5K50**; **5K4W**; **5K4Y**.

Acknowledgements

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.

Aoyama, Y. & Motokawa, Y. (1981). *J. Biol. Chem.* **256**, 12367–12373.

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Crystallogr. D Biol. Crystallogr.* **67**, 271–281.

Bern, C. (2015). *N. Engl. J. Med.* **373**, 456–466.

Bird, M. I. & Nunn, P. B. (1983). *Biochem. J.* **214**, 687–694.

Bowyer, A., Mikolajek, H., Stuart, J. W., Wood, S. P., Jamil, F., Rashid, N., Akhtar, M. & Cooper, J. B. (2009). *J. Struct. Biol.* **168**, 294–304.

Boylan, S. A. & Dekker, E. E. (1978). *Biochem. Biophys. Res. Commun.* **85**, 190–197.

Boylan, S. A. & Dekker, E. E. (1981). *J. Biol. Chem.* **256**, 1809–1815.

Büscher, P., Cecchi, G., Jamonneau, V. & Priotto, G. (2017). *The Lancet*. **390**, 2397–2409.

Creek, D. J., Mazet, M., Achcar, F., Anderson, J., Kim, D.-H., Kamour, R., Morand, P., Millerioux, Y., Biran, M., Kerkhoven, E. J., Chokkathukalam, A., Weidt, S. K., Burgess, K. E. V., Breitling, R., Watson, D. G., Bringaud, F. & Barrett, M. P. (2015). *PLOS Pathog*. **11**, e1004689.

Cross, G. A., Klein, R. A. & Linstead, D. J. (1975). *Parasitology*. **71**, 311–326.

Dale, R. A. (1978). *Biochim. Biophys. Acta BBA - Gen. Subj.* **544**, 496–503.

Davies, G. E. & Stark, G. R. (1970). *Proc. Natl. Acad. Sci.* **66**, 651–656.

Edgar, A. J. (2002). *BMC Genet.* **3**, 18.

Edgar, A. J. (2005). *BMC Genomics*. **6**, 32.

Emsley, P. & Cowtan, K. (2004). *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501.

Evans, P. (2005). *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82.

Evans, P. R. (2011). *Acta Crystallogr. D Biol. Crystallogr.* **67**, 282–292.

Evans, P. R. & Murshudov, G. N. (2013). *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214.

Gani, O. A. B. S. M., Adekoya, O. A., Giurato, L., Spyrakis, F., Cozzini, P., Guccione, S., Winberg, J.-O. & Sylte, I. (2008). *Biophys. J.* **94**, 1412–1427.

Green, M. L. & Elliott, W. H. (1964). *Biochem. J.* **92**, 537–549.

Hammes, G. G., Chang, Y.-C. & Oas, T. G. (2009). *Proc. Natl. Acad. Sci.* **106**, 13737–13741.

He, C., Huang, X., Liu, Y., Li, F., Yang, Y., Tao, H., Han, C., Zhao, C., Xiao, Y. & Shi, Y. (2015). *J. Struct. Biol.* **192**, 510–518.

Hespenheide, B. M., Rader, A. J., Thorpe, M. F. & Kuhn, L. A. (2002). *J. Mol. Graph. Model.* **21**, 195–207.

Higashi, N., Matsuura, T., Nakagawa, A. & Ishikawa, K. (2005). *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* **61**, 432–434.

Huntington, J. A. (2008). *Biol. Chem.* **389**, 1025–1035.

Ishikawa, K., Higashi, N., Nakamura, T., Matsuura, T. & Nakagawa, A. (2007). *J. Mol. Biol.* **366**, 857–867.

Jacobs, D. J., Rader, A. j., Kuhn, L. A. & Thorpe, M. f. (2001). *Proteins Struct. Funct. Bioinforma.* **44**, 150–165.

Jacobs, D. J. & Thorpe, M. F. (1995). *Phys. Rev. Lett.* **75**, 4051–4054.

Jamil, F. (2012). Mechanistic and Stereochemical Studies on 2-Amino-3-Ketobutyrate CoA Ligase and Related Enzymes. PhD thesis. University of Lahore.

Jimenez-Roldan, J. E., Freedman, R. B., Römer, R. A. & Wells, S. A. (2012). *Phys. Biol.* **9**, 016008.

Johnson, A. R., Chen, Y.-W. & Dekker, E. E. (1998). *Arch. Biochem. Biophys.* **358**, 211–221.

Kabsch, W. (2010). *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132.

Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.

Kao, Y. C. & Davis, L. (1994). *Protein Expr. Purif.* **5**, 423–431.

Kavanagh, K. L., Jörnvall, H., Persson, B. & Oppermann, U. (2008). *Cell. Mol. Life Sci.* **65**, 3895–3906.

Kazuoka, T., Takigawa, S., Arakawa, N., Hizukuri, Y., Muraoka, I., Oikawa, T. & Soda, K. (2003). *J. Bacteriol.* **185**, 4483–4489.

Klein, R. A., Angus, J. M., Amadife, A. E. & Smith, L. (1980). *Comp. Biochem. Physiol. Part B Comp. Biochem.* **66**, 143–146.

Koumanov, A., Benach, J., Atrian, S., Gonzàlez-Duarte, R., Karshikoff, A. & Ladenstein, R. (2003). *Proteins Struct. Funct. Bioinforma.* **51**, 289–298.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Lam, V. M. S., Chan, I. P. R. & Yeung, Y. G. (1980). *J. Gen. Microbiol.* **117**, 539–542.

Linstead, D. J., Klein, R. A. & Cross, G. A. M. (1977). *J Gen Microbiol*. **101**, 243–251.

Machielsen, R. & van der Oost, J. (2006). *FEBS J.* **273**, 2722–2729.

Marcus, J. P. & Dekker, E. E. (1993*a*). *Biochem. Biophys. Res. Commun.* **190**, 1066–1072.

Marcus, J. P. & Dekker, E. E. (1993*b*). *J. Bacteriol.* **175**, 6505–6511.

Mazet, M., Morand, P., Biran, M., Bouyssou, G., Courtois, P., Daulouède, S., Millerioux, Y., Franconi, J.-M., Vincendeau, P., Moreau, P. & Bringaud, F. (2013). *PLoS Negl Trop Dis*. **7**, e2587.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Crystallogr.* **40**, 658–674.

Millerioux, Y., Ebikeme, C., Biran, M., Morand, P., Bouyssou, G., Vincent, I. M., Mazet, M., Riviere, L., Franconi, J.-M., Burchmore, R. J. S., Moreau, P., Barrett, M. P. & Bringaud, F. (2013). *Mol. Microbiol.* **90**, 114–129.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255.

Nakano, S., Okazaki, S., Tokiwa, H. & Asano, Y. (2014). *J. Biol. Chem.* jbc.M113.540773.

Persson, B., Kallberg, Y., Bray, J. E., Bruford, E., Dellaporta, S. L., Favia, A. D., Duarte, R. G., Jörnvall, H., Kavanagh, K. L., Kedishvili, N., Kisiela, M., Maser, E., Mindnich, R., Orchard, S., Penning, T. M., Thornton, J. M., Adamski, J. & Oppermann, U. (2009). *Chem. Biol. Interact.* **178**, 94–98.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.

Ray, M. & Ray, S. (1985). *J. Biol. Chem.* **260**, 5913–5918.

Schmidt, A., Sivaraman, J., Li, Y., Larocque, R., Barbosa, J. A. R. G., Smith, C., Matte, A., Schrag, J. D. & Cygler, M. (2001). *Biochemistry (Mosc.)*. **40**, 5151–5160.

Schüttelkopf, A. W. & van Aalten, D. M. F. (2004). *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1355–1363.

Stein, N. & Ballard, C. (2009). Intensity to amplitude conversion using CTRUNCATE.

Suhre, K. & Sanejouand, Y.-H. (2004). *Nucleic Acids Res.* **32**, W610–W614.

Sullivan, S. M. & Holyoak, T. (2008). *Proc. Natl. Acad. Sci.* **105**, 13829–13834.

Ueatrongchit, T. & Asano, Y. (2011). *Anal. Biochem.* **410**, 44–56.

Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F. & Murshudov, G. N. (2004). *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2184–2195.

Vagin, A. & Teplyakov, A. (1997). *J. Appl. Crystallogr.* **30**, 1022–1025.

Vagin, A. & Teplyakov, A. (2009). *Acta Crystallogr. D Biol. Crystallogr.* **66**, 22–25.

Wagner, M. & Andreesen, J. R. (1995). *Arch. Microbiol.* **163**, 286–290.

Wells, S. A., Crennell, S. J. & Danson, M. J. (2014). *Proteins Struct. Funct. Bioinforma.* **82**, 2657–2670.

Wells, S. A., Kamp, M. W. van der, McGeagh, J. D. & Mulholland, A. J. (2015). *PLOS ONE*. **10**, e0133372.

Wells, S., Menor, S., Hespenheide, B. & Thorpe, M. F. (2005). *Phys. Biol.* **2**, S127.

Winberg, J. O., Brendskag, M. K., Sylte, I., Lindstad, R. I. & McKinley-McKee, J. S. (1999). *J. Mol. Biol.* **294**, 601–616.

Winter, G. (2010). *J. Appl. Crystallogr.* **43**, 186–190.

Wuxiuer, Y., Morgunova, E., Cols, N., Popov, A., Karshikoff, A., Sylte, I., Gonzàlez-Duarte, R., Ladenstein, R. & Winberg, J.-O. (2012). *FEBS J.* **279**, 2940–2956.

Yoneda, K., Sakuraba, H., Araki, T. & Ohshima, T. (2012). *J. Biol. Chem.* **287**, 12966–12974.

Yoneda, K., Sakuraba, H., Muraoka, I., Oikawa, T. & Ohshima, T. (2010). *FEBS J.* **277**, 5124–5132.

Yuan, J. H. & Austic, R. E. (2001). *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **130**, 65–73.

Table 1. Details of crystallisation conditions used in the preparation of crystals from which data were collected. Note that all co-crystals were obtained by co-crystallisation.

	Model								
RCSB PDB ID	**5L9A**	**5LC1**	**5K4Q**	**5K4T**	**5K4V**	**5K4U**	**5K50**	**5K4W**	**5K4Y**
Crystallisation Conditions									
Conditions	0.1 M HEPES; 20 % w/v PEG 10K; pH 7.5; TDH (5.28 x 10^{-2}mM [2.0 mg/ml])	0.1 M HEPES; 20 % w/v PEG 10K; pH 7.5; TDH (5.28 x 10^{-2}mM [2.0 mg/ml]); NAD$^+$ (1mM); pyruvate (30mM)	0.2M lithium sulphate; 0.1M Tris; 30% w/v PEG 4K; pH 8.5; TDH (6.4 x 10^{-2}mM [2.4mg/ml]); KBL (3.8 x 10^{-2}mM [1.8mg/ml]); NAD$^+$ (0.09mM); PLP (0.18mM); L-Threonine (7.33mM)	0.1M MES; 0.2M ammonium sulphate; 30% w/v PEG 5K MME; pH 6.5; TDH (9.0 x 10^{-2}mM [3.4mg/ml]); L-serine (50mM)	0.1M tri-sodium citrate; 30% w/v PEG 4K; 0.2M ammonium acetate; pH 8.5; TDH (3.7 x 10^{-2}mM [1.4mg/ml]); NAD$^+$ (10mM); BPOB (0.5mM)	0.1M tri-sodium citrate; 30% w/v PEG 4K; 0.2M ammonium acetate; pH 5.6; TDH (3.7 x 10^{-2}mM [1.4mg/ml]); NAD$^+$ (8.5mM); quinine (2.5mM [hydrochloride dehydrate])	0.1M tri-sodium citrate; 30% w/v PEG 4K; 0.2M ammonium acetate; pH 5.6; TDH (10.4 x 10^{-2}mM [3.9mg/ml]); NAD$^+$ (10mM); L-allo-threonine (30mM)	0.2M sodium acetate; 0.1M Tris; 30% w/v PEG 4K; pH 8.5; TDH (9.3 x 10^{-2}mM [3.5mg/ml]); NADH (10mM); L-Threonine (30mM)	0.1M tri-sodium citrate; 25% w/v PEG 4K; 0.2M ammonium acetate; pH 5.6 TDH (3.7 x 10^{-2}mM [1.4mg/ml]); NAD$^+$ (10mM); methylglyoxal (8mM)
Cryoprotectant	Glycerol	Glycerol	Glycerol	Glycerol	Glycerol	Glycerol	Glycerol	None	None

Table 2. Details of data collected and structures solved by X-ray crystallography. *Data collection statistics for the outer resolution shell are provided in brackets. †The high R-free values of some of the complexes probably stem from the challenges of processing data and refining structures with a unit cell parameter greater than 270 Angstroms.

	Model								
RCSB PDB ID	**5L9A**	**5LC1**	**5K4Q**	**5K4T**	**5K4V**	**5K4U**	**5K50**	**5K4W**	**5K4Y**
Data Collection									
Beamline	ESRF ID23-2	ESRF ID29	ESRF	DLS I04-1	DLS I04-1	DLS	DLS I04-1	DLS I02	DLS I04-1
Wavelength (Å)	0.8726	1.07356	0.9334	0.9173	0.9173	0.9173	0.92	0.9795	0.92001
Space group	P1	$P2_12_12$	$P2_12_12$	$P4_32_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$
Unit cell parameters									
a (Å)	46.98	132.03	132.04	91.76	90.40	90.07	133.03	83.45	133.45
b (Å)	57.72	276.49	276.49	91.76	131.53	133.10	273.06	136.12	278.63
c (Å)	70.07	55.74	55.74	93.60	55.02	55.61	55.80	55.69	56.27
α (°)	72.46	90	90	90	90	90	90	90	90
β (°)	70.38	90	90	90	90	90	90	90	90
γ (°)	73.20	90	90	90	90	90	90	90	90
No. of reflections	409937	874324	624537	301645	109277	203173	297998	435135	438691
No. of unique reflections (used)	109054	117204	93173	23982	32475	61881	91289	67909	193067
Solvent content (%)	44.20	45.05	46.34	52.83	42.88	43.94	44.98	41.24	46.7
Matthews coefficient ($Å^3.Da^{-1}$)	2.20	2.24	2.29	2.61	2.15	2.19	2.23	2.09	2.31
Resolution range (Å)*	53.82 - 1.45 (1.53 - 1.45)	95.48 - 2.10 (2.21 - 2.10)	48.14 - 2.30 (2.42 - 2.30)	29.54 – 1.91 (1.96 - 1.91)	28.23 - 2.20 (2.32 - 2.20)	29.29 - 1.79 (1.89 - 1.79)	31.09 - 2.26 (2.34 – 2.26)	46.32 - 1.72 (1.76 - 1.72)	40.12 - 1.77 (1.80 - 1.77)
Multiplicity*	3.8 (2.3)	7.5 (3.3)	6.7 (4.6)	12.6 (10.0)	3.4 (3.0)	3.3 (3.3)	1.9 (1.9)	6.4 (6.6)	6.2 (6.3)
Completeness (%)*	95.0 (82.1)	97.3 (81.9)	99.2 (95.3)	100.0 (95.2)	96.0 (95.1)	98.0 (97.5)	90.4 (92.5)	99.5 (100.0)	99.8 (99.9)
Mean I/σ(I) *	10.4 (2.7)	11.4 (3.2)	11.2 (4.4)	33.6 (3.7)	13.4 (3.9)	12.7 (2.8)	5.6 (1.9)	11.0 (2.5)	12.7 (4.1)
R_{merge} (%)*	8.9 (39.8)	14.5 (52.5)	12.0 (27.6)	5.4 (70.9)	6.7 (30.5)	7.4 (46.0)	15.1 (64.6)	9.4 (76.5)	8.6 (41.4)
R_{pim} (%)*	5.1 (33.1)	5.5 (26.9)	4.9 (14.2)	1.6 (24.3)	4.0 (19.7)	4.8 (29.9)	15.1 (64.6)	4.0 (32.2)	3.7 (17.9)
R_{meas} (%)*	10.3 (52.2)	15.6 (59.4)	13.0 (31.3)	5.6 (78.7)	7.9 (36.6)	8.8 (55.1)	21.4 (91.4)	10.2 (83.1)	9.4 (45.2)
Refinement									
Max. resolution (Å)	1.45	2.10	2.3	2.1	2.2	1.9	2.26	1.72	1.77

	Model								
RCSB PDB ID	**5L9A**	**5LC1**	**5K4Q**	**5K4T**	**5K4V**	**5K4U**	**5K50**	**5K4W**	**5K4Y**
R_{work} (%)	16.5	18.1	17.5	17.4	16.4	18.6	20.7	15.2	16.1
R_{free} (%) [†]	22.2	28.7	26.2	24.0	24.1	23.2	28.1	19.8	20.1
RMS bond lengths (Å)	0.011	0.008	0.0178	0.0245	0.0201	0.0241	0.0128	0.0198	0.0201
RMS bond angles (°)	2.198	1.837	1.764	2.059	1.826	1.99	1.6492	2.1037	1.9749
Average B-factor (Å^2)	18.68	22.20	21.09	26.769	26.234	22.138	33.76	27.689	19.218
Ramachandran favoured, % of total	696, 97.6%	1872, 97.0%	1830, 96.62%	309, 96.87%	613, 96.69%	618, 97.48%	1801, 95.14%	622, 97.65%	1824, 97.91%
Ramachandran allowed, % of total	711, 99.7%	1923, 99.6%	61, 3.22%	8, 2.51%	18, 2.84%	14, 2.21%	82, 4.33%	13, 2.04%	33, 1.77%
Ramachandran outliers, % of total	2, 0.3%	7, 0.4%	3, 0.16%	2, 0.63%	3, 0.47%	2, 0.32%	10, 0.53%	2, 0.31%	6, 0.32%
Molecules in the asymmetric unit	2	6	6	1	2	2	6	2	6
Quaternary structure	Dimer	Dimer	Dimer	Dimer	Dimer	Dimer	Dimer	Dimer	Dimer
Residue range	His1-Leu321	His1-Leu321	Pro3-Leu321	His1-Leu321	Pro3-Leu321	Pro3-Leu321	Pro3-Leu321	His1-Leu321	Met2-Leu321
No. protein atoms per monomer	2621	2511	2493	2532	2493	2493	2500.5	2507	2533
No. water molecules per monomer	651	328	252.3	152	220.5	100.5	73.7	210	241.1
NAD$^+$ bound?	No	Yes	Yes	No	Yes	Yes	Yes	Yes (NADH)	Yes
Ligands bound, Mean B-factor	-	NAD, 17.69 Pyruvate, 23.07 NAD, 21.54 Pyruvate, 32.95 NAD, 16.75 Pyruvate, 45.43 NAD, 14.71 Pyruvate, 28.07 NAD, 21.78 Pyruvate, 41.96 NAD, 17.43	NAD, 21.38 NAD, 23.70 NAD, 18.70 NAD, 15.34 NAD, 22.09 NAD, 16.01		NAD, 30.90 NAD, 26.30	NAD, 15.76 NAD, 17.58	NAD, 26.36 NAD, 20.73 NAD, 21.75 NAD, 24.45 NAD, 24.58 NAD, 23.92 L-*allo*-threonine, 39.53 L-*allo*-threonine, 26.88	NAD, 22.22 NAD, 22.68 L-threonine, 26.06 L-threonine, 30.58	NAD, 11.34 NAD, 12.22 NAD, 12.52 NAD, 11.02 NAD, 12.44 NAD, 12.82

RCSB PDB ID				Model					
	5L9A	**5LC1**	**5K4Q**	**5K4T**	**5K4V**	**5K4U**	**5K50**	**5K4W**	**5K4Y**
		Pyruvate 30.47							
Other solvent molecules/ions	acetate	acetate	glycerol, sulphate	glycerol	glycerol, acetate, Na^+	glycerol, acetate	glycerol, acetate	glycerol, Na^+	glycerol, acetate
Conformation(s) of flexible Loop 1	Open	Closed	Open	Open	Closed	Open and Closed	Open and Closed	Closed	Open

Figure Legends

Figure 1. The L-threonine degradation pathway. The enzyme TDH oxidises L-threonine to 2-amino-3-ketobutyrate (AKB). KBL then converts AKB to glycine and Ac-CoA. Alternatively, AKB may undergo spontaneous breakdown to aminoacetone and carbon dioxide.

Figure 2. NAD+-bound TbTDH coloured by secondary structure (gold = α-helix; dark green = β-strand; light blue = loop; NAD coloured by heteroatom). Cartoon representation produced using UCSF Chimera.

Figure 3. Conformational flexibility as indicated by RMSD of Cα positions of the different TbTDH structures (PDB: 5K4Q; PDB: 5K4T; PDB: 5K4V; PDB: 5K4U; PDB: 5K50; PDB: 5K4W) when superimposed in UCSF Chimera. (a) A ribbon representation of TDH, coloured by the RMSD of α-carbon positions in each residue. The scale at the bottom of the diagram indicates the colour scheme relating to the RMSD measured in Ångstrom. (b) A bar chart showing the RMSD of α-carbons per TDH residue.

Figure 4. A colour representation of the average B-factors of TDH residues, derived from structures PDB: 5K4Q, PDB: 5K4T, DB ID: 5K4V, PDB: 5K4U, PDB: 5K50 and PDB: 5K4W. (a) A ribbon diagram of TDH coloured by residue-average B-factor (scale in horizontal bar). The distinct 'closed' position of Loop 1 is displayed in red. (b) A chart showing the average B-factor of residues in 'open' TDH structures. (c) A chart showing average B-factors of TDH residues in 'closed' structures.

Figure 5. The TDH dimer and its dimerisation interface. Lines between amino acids indicate Van der Waals, polar or ionic interactions. Key residues involved in the interaction include Glu86, Asp90, Asp94, Arg101, Lys126, Asp135 and Trp157.

Figure 6. Binding of NAD+ to TDH in 3D (a) and 2D (b) representations. In (a) hydrogen bonds are indicated by red lines; water molecules are depicted as red spheres. In (b) hydrogen bonds are

indicated by dotted lines, whilst Van der Waals interactions are represented by green curved lines, with the corresponding TDH residues in green text.

Figure 7. Binding of L-threonine to TDH in 3D (a) and 2D (b) representations, derived from structure 5K5W. In (a) hydrogen bonds are indicated by red lines, electron density of L-threonine is represented by a blue mesh and a water molecule is depicted by a red sphere. In (b) hydrogen bonds are indicated by dotted lines, whilst Van der Waals interactions are represented by green curved lines, with the corresponding TDH residues in green text. (c) shows a surface representation of threonine-bound TDH derived from PDB: 5K4W; the surface of loop 1, which is closed over the active site, is coloured red and is semi-transparent.

Figure 8. Interaction between Asn96 and Ile80 (a), enabling hydrogen bonding of a water atom in a larger chain of hydrogen-bonded water molecules (b). Hydrogen bonds are indicated by red lines; water molecules are depicted as red spheres.

Figure 9. Geometric simulations of apo-TbTDH along Elnemo's mode 7: (a) overlay of multiple projections of apo-TbTDH conformations, superposed on holo- (gold) and L-threonine-bound (dark red) TbTDH, with arrows indicating direction of the twisting motion; (b) superposition of holo-TbTDH (gold) over a projection of the apo form conformation (fitted backbone RMSD = 0.49 Å); (c) superposition of L-threonine-bound TbTDH (dark red) over a projection of the apo form conformation (backbone fitted RMSD = 0.63 Å).

Figure 10. Overlays of multiple states of apo-TbTDH during simulations of random diffusive motion. The arrow indicates the direction of progressive movement of Loop 1 (in gold) during the simulation. States showing extremes of the Loop 1 position are in solid colour, whilst intermediate states are translucent.

Figures

Fig. 1.
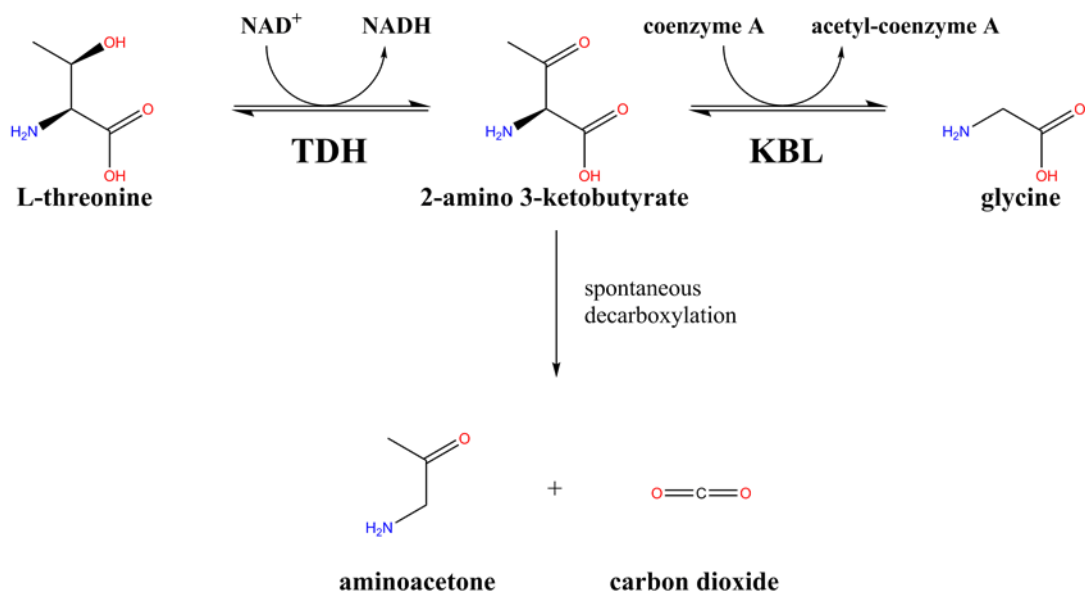
Fig. 2.
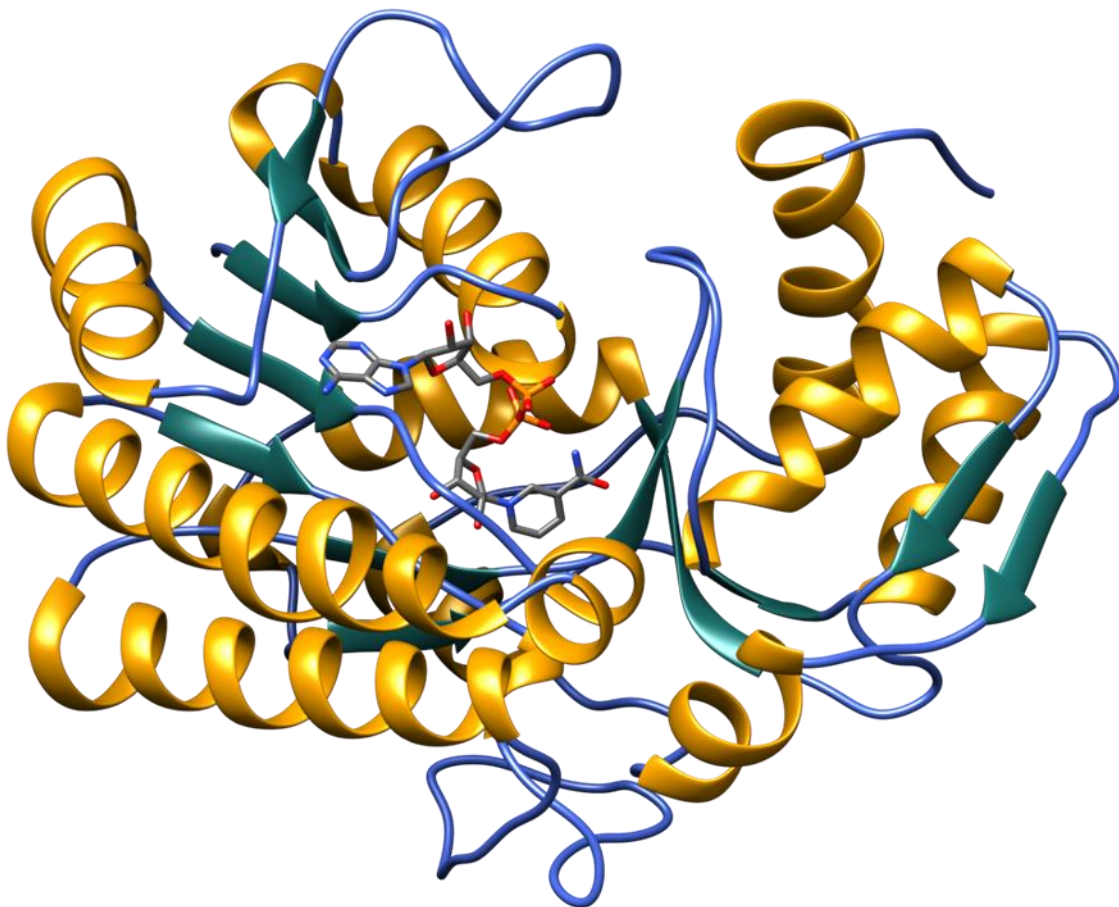
Fig. 3.
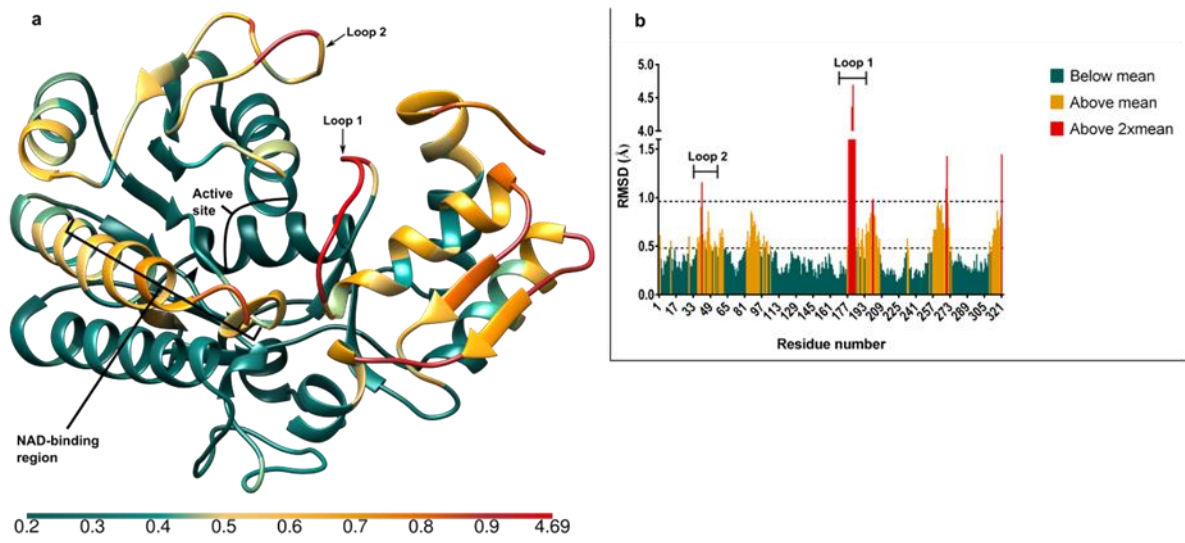
Fig. 4.
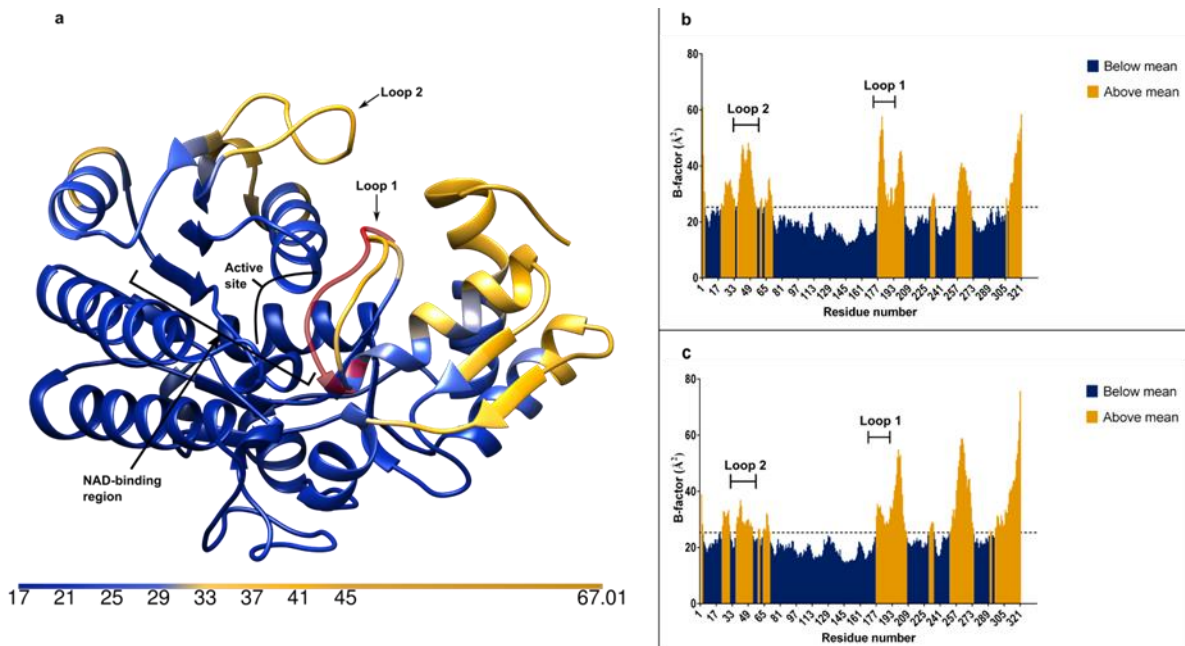
Fig. 5
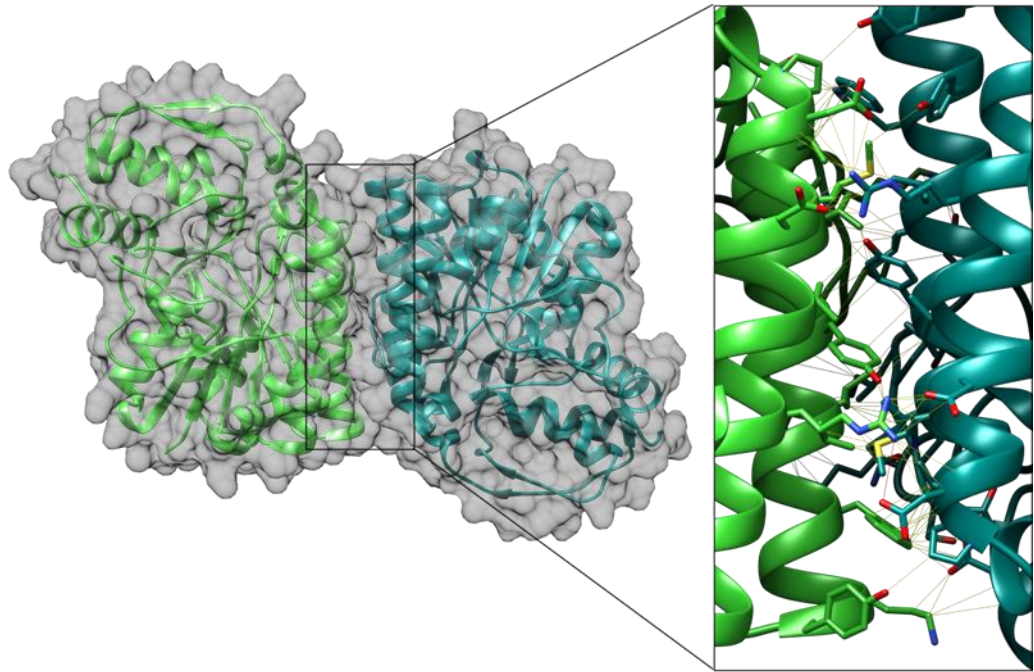
Fig. 6.
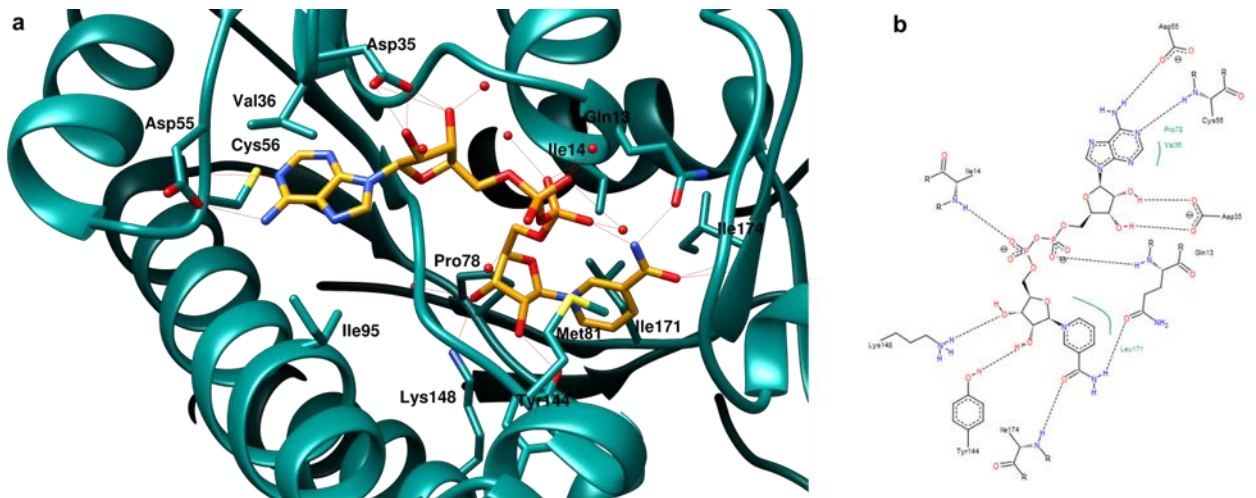
Fig. 7.
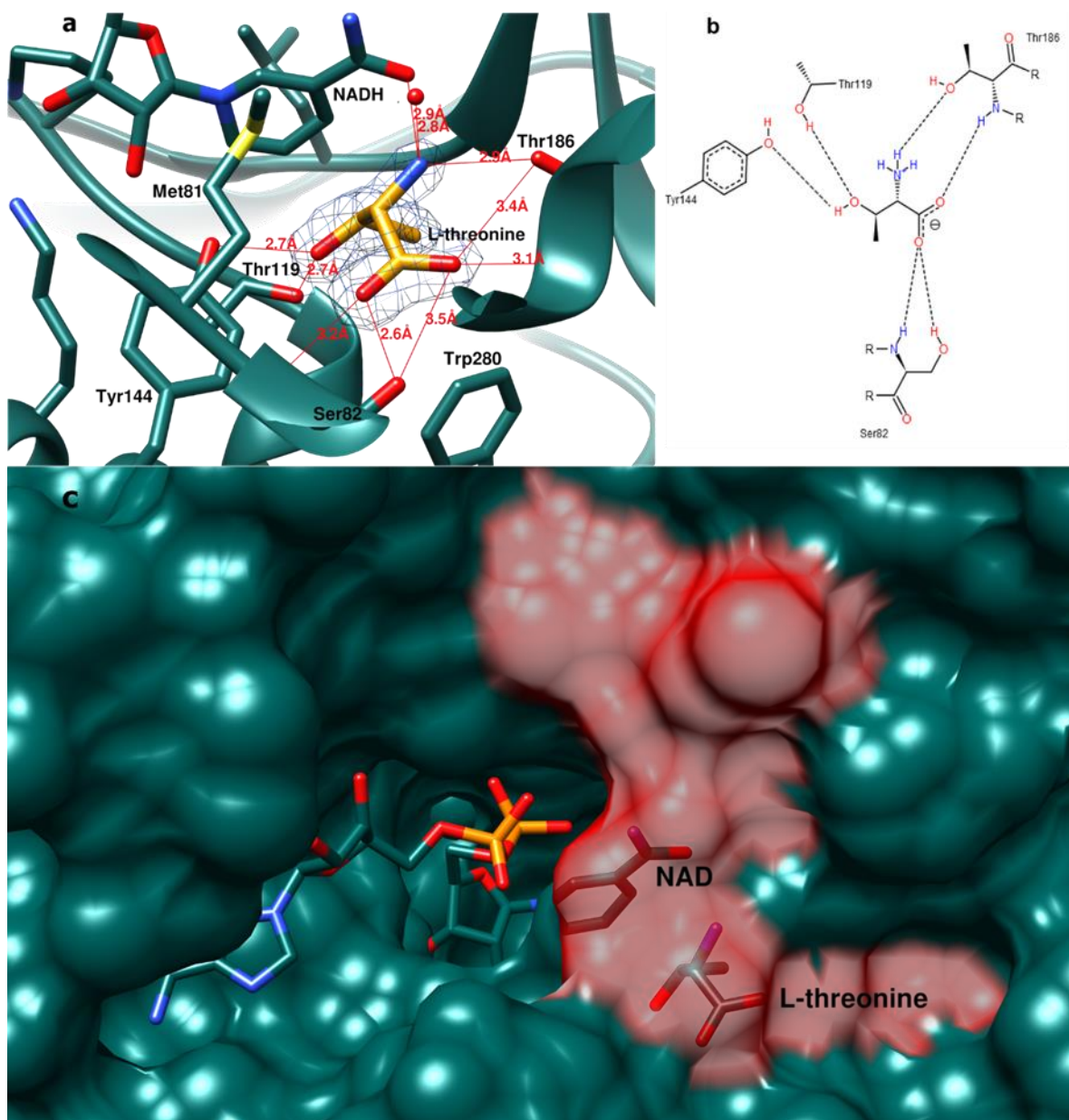
Fig. 8.
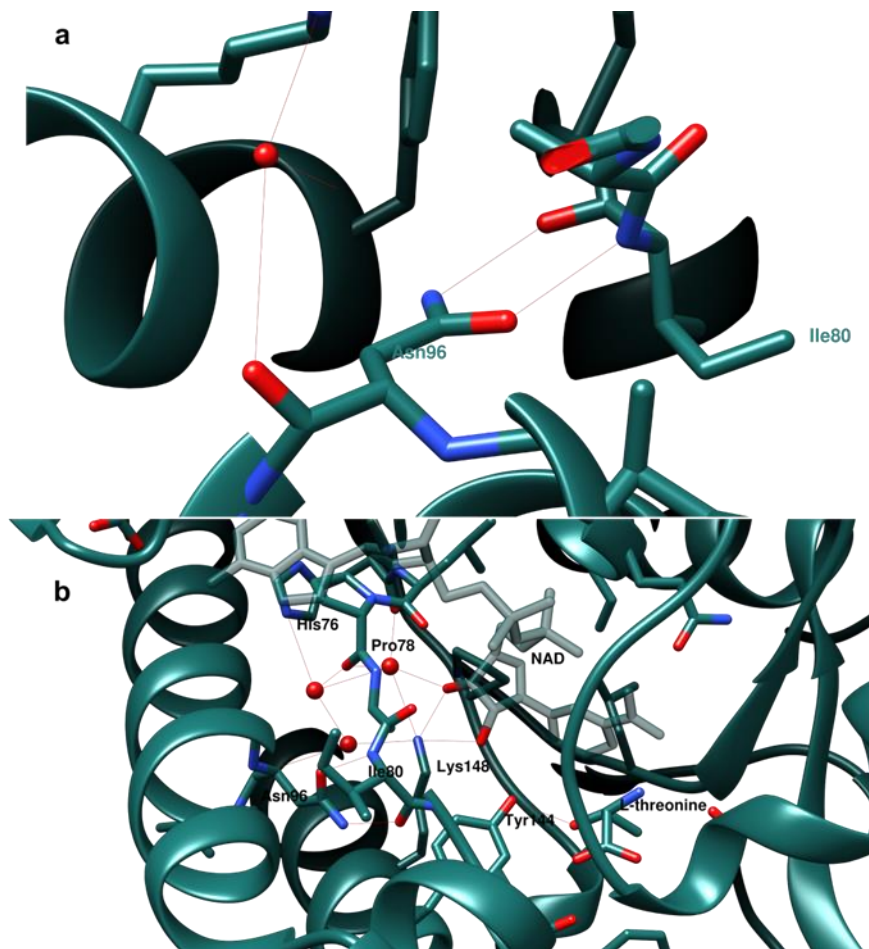
Fig. 9.
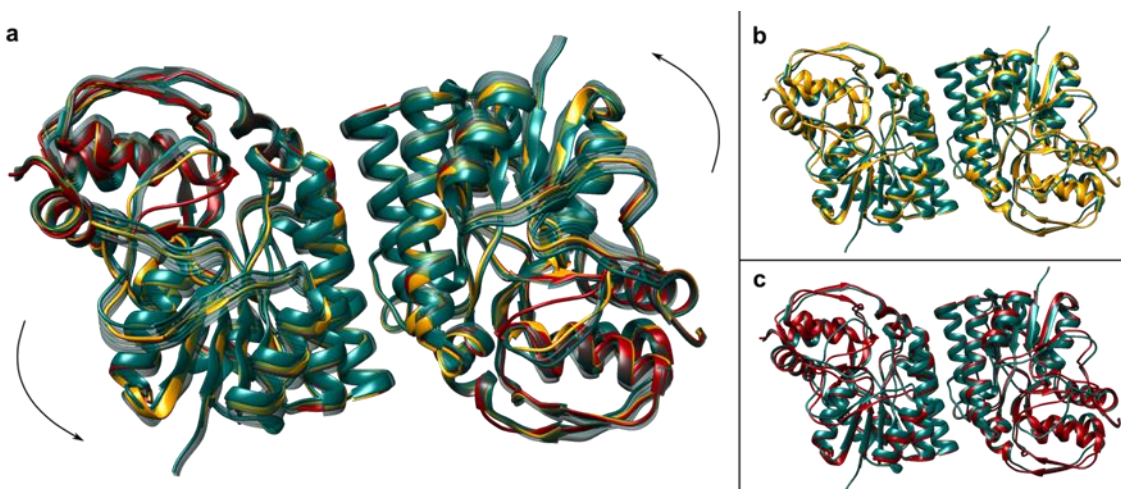
Fig. 10.
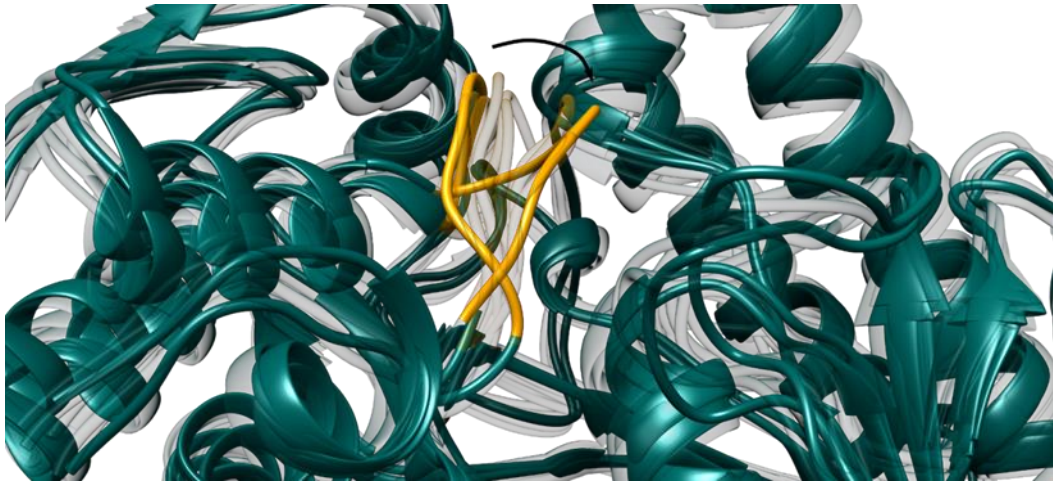
Supplementary Material

Model	Oligomer	Accessible surface area (Å^2)	Buried surface area (Å^2)	ΔG_{diss}* (kcal.mol^{-1})	Stable in solution?	Visual confirmation
Model: 5L9A Space group: P1 Unit cell parameters: a=46.98, b=57.72, c=70.07; α=72.46°, β=70.38°, γ=73.20° Molecules in the asymmetric unit: 2	Dimer	25869.5	3161.9	2.5	Yes	✓
Model: 5LC1 Space group: P2$_1$2$_1$2 Unit cell parameters: a=132.03, b=276.49, c=55.74; αβγ=90° Molecules in the asymmetric unit: 6	Dimer	24444.8	5980.5	3.8	Yes	✓
	Dimer	24278.4	6117.1	3.6	Yes	✓
	Dimer	24600.3	5880.2	3.3	Yes	✓
Model: 5K4Q Space group: P2$_1$2$_1$2 Unit cell parameters: a=132.04, b=276.49, c=55.74; αβγ=90° Molecules in the asymmetric unit: 6	Dimer	24918.1	5470.9	3.3	Yes	✓
	Dimer	24832.7	5449.2	2.6	Yes	✓
	Dimer	24993.7	5428.6	2.3	Yes	✓
Model: 5K4T Space group: P4$_3$2$_1$2 Unit cell parameters: ab=91.76, c=93.60; αβγ=90° Molecules in the asymmetric unit: 1	Dimer	25239.1	5028.6	3.2	Yes	✓

Model	Oligomer	Accessible surface area (Å^2)	Buried surface area (Å^2)	ΔG_{diss}* (kcal.mol^{-1})	Stable in solution?	Visual confirmation
Model: 5K4V Space group: P2$_1$2$_1$2 Unit cell parameters: a=90.40, b=131.53, c=55.02; αβγ=90° Molecules in the asymmetric unit: 2	Dimer	24519.8	4880.8	4.4	Yes	✓
Model: 5K4U Space group: P2$_1$2$_1$2 Unit cell parameters: a=133.03,b=273.06, c=55.80; αβγ=90° Molecules in the asymmetric unit: 2	Dimer	24264.3	7309.4	4.1	Yes	✓
Model: 5K50 Space group: P2$_1$2$_1$2 Unit cell parameters: a=133.03, b=273.06, c=55.80; αβγ=90° Molecules in the asymmetric unit: 6	Dimer	24167.7	4813.6	4.3	Yes	✓
	Dimer	24294.1	4772.0	3.1	Yes	✓
	Dimer	24217.8	4790.1	2.4	Yes	✓
Model: 5K4W Space group: P2$_1$2$_1$2 Unit cell parameters: a=83.45, b=136.12, c=55.69; αβγ=90° Molecules in the asymmetric unit: 2	Dimer	24587.0	5788.1	0.3	Yes	✓
	Tetramer	46831.2	13919.2	-1.6	Uncertain	x (see Supplementary Figure 12)
Model: 5K4Y Space group: P2$_1$2$_1$2 Unit cell parameters: a=133.45, b=278.63, b=56.27; αβγ=90°	Dimer	24779.8	6050.2	4.5	Yes	✓
		24400.6	6351.1	4.1	Yes	✓

50

Model	Oligomer	Accessible surface area (Å²)	Buried surface area (Å²)	ΔG_{diss}* (kcal.mol⁻¹)	Stable in solution?	Visual confirmation
Molecules in the asymmetric unit: 6		24521.6	6357.1	3.1	Yes	✓

Supplementary Table 2. Samples, elution volumes, predicted molecular weights and corresponding oligomeric forms, as indicated by results of SEC experiments.

Sample Contents	Ve (ml)	Ve/V$_0$	Interpolated MW (kDa)	Multiple of MW	Suggested Oligomeric State
TDH (1.8 x10^{-1} mM [6.8mg/ml])	15	1.79	68.8	1.82	Dimer
TDH (1.1 x10^{-1} mM [4mg/ml]) + 10mM NAD$^+$	15	1.79	68.8	1.82	Dimer
KBL (4.4 x10^{-2} mM [2mg/ml])	14.4	1.71	87.5	1.91	Dimer
TDH (5.3 x10^{-2} mM [2mg/ml]) + KBL (4.4 x10^{-2} mM [2mg/ml]) - first peak	12.5	1.49	187.7	4.97 (TDH); 4.09 (KBL)	Multi-enzyme complex
TDH (5.3 x10^{-2} mM [2mg/ml]) + KBL (4.4 x10^{-2} mM [2mg/ml]) - second peak	16.8	1.99	34.0	0.90 (TDH); 0.74 (KBL)	Monomers (TDH and/or KBL)

Supplementary Table 3. SDR and GalE-like TDH deposited in the Protein Data Bank.

Organism	Amino acids	Quaternary structure	Wild-type/mutant	Ligands	PDB ID	Literature reference(s)
Trypnanosoma brucei	332	Dimer	Wild-type	-	5L9A	This manuscript, released 22 JUN 16
	332	Dimer	Wild-type	NAD; pyruvate	5LC1	This manuscript, deposited 22 JUN 16
	332	Dimer	Wild-type	NAD	5K4Q	This manuscript, released 15 NOV 17
	332	Dimer	Wild-type	-	5K4T	This manuscript, released 15 NOV 17
	332	Dimer	Wild-type	NAD	5K4V	This manuscript, released 15 NOV 17
	332	Dimer	Wild-type	NAD	5K4U	This manuscript, released 15 NOV 17
	332	Dimer	Wild-type	NAD, L-*allo*-threonine	5K50	This manuscript, released 15 NOV 17
	332	Dimer	Wild-type	NAD, L-threonine	5K4W	This manuscript, released 04 JAN 18
	332	Dimer	Wild-type	NAD	5K4Y	This manuscript, released 17 JAN 18
Flavobacterium frigidimaris	312	Dimer	Wild-type	NAD; glycerol	2YY7	Febs J 277: 5124-5132 (2010)
Thermoplasma volcanium	317	Dimer	Wild-type	NAD	3A1N	J Biol Chem 287:12966-12974 (2012)
	317	Dimer	Wild-type	NAD; pyruvate	3A4V	
	317	Dimer	Y137F mutant	NAD; L-threonine	3A9W	
	317	Dimer	Y137F mutant	NAD; L-3-Hydroxynorvaline	3AJR	
Mus musculus	373	Dimer	Wild-type	NAD$^+$	4YR9	J Struct Biol 192:510-518 (2015)
	373	Dimer	Wild-type	-	4YRA	
	373	Dimer	R180K mutant	NAD	4YRB	
Cupriavidus necator	318	Dimer	Wild-type	NAD; L-threonine	3WMX	J Biol Chem 289:10445-10454 (2014)
	318	Dimer	Wild-type	-	3WMW	
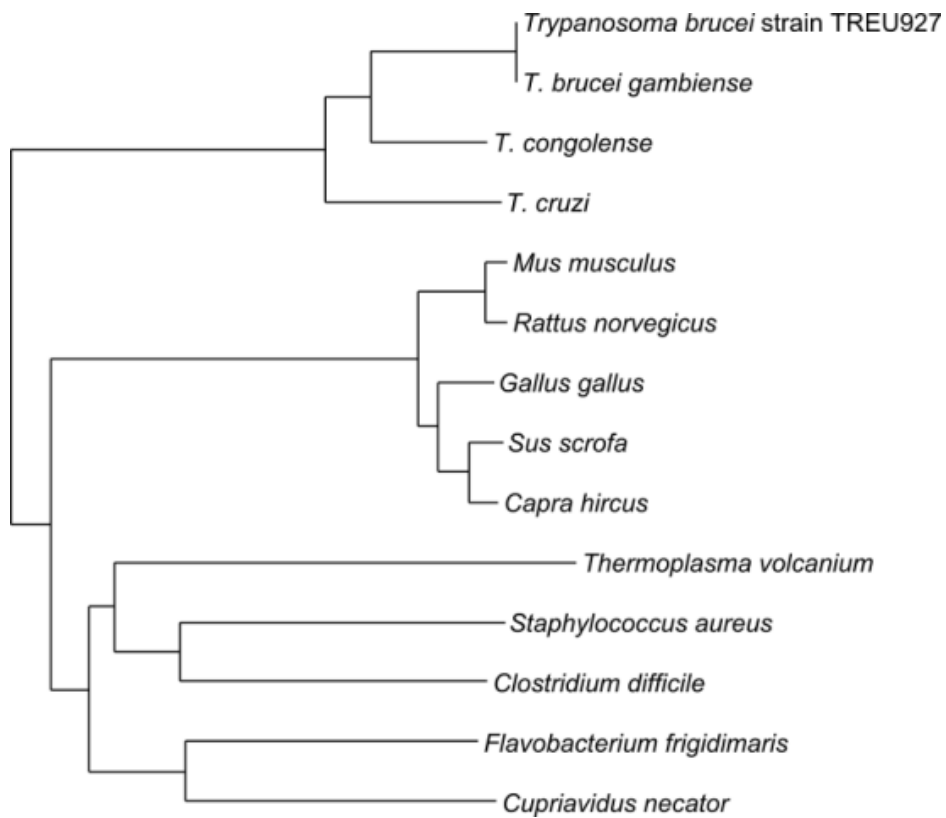
Supplementary Figure 1. Phylogenetic Tree of TDH from *Trypanosoma* and other organisms. The lengths of the horizontal lines represent the relative genetic distances between the TDH genes from the named organisms. Genes connected to a common branch share a common genetic lineage. The phylogenetic tree was generated using SeaView 4.

Supplementary Figure 2. Primary sequence alignment of TDH from *Trypanosoma* species and other organisms. Each residue is represented by the corresponding single-letter code, and residues are coloured according to 8 default families of amino acids sharing biochemical characteristics. Sequences of non-aligning residues are represented by a dash. Key conserved sequence segments are indicated by an arrow and bold text. The alignment was created using the software SeaView 4.
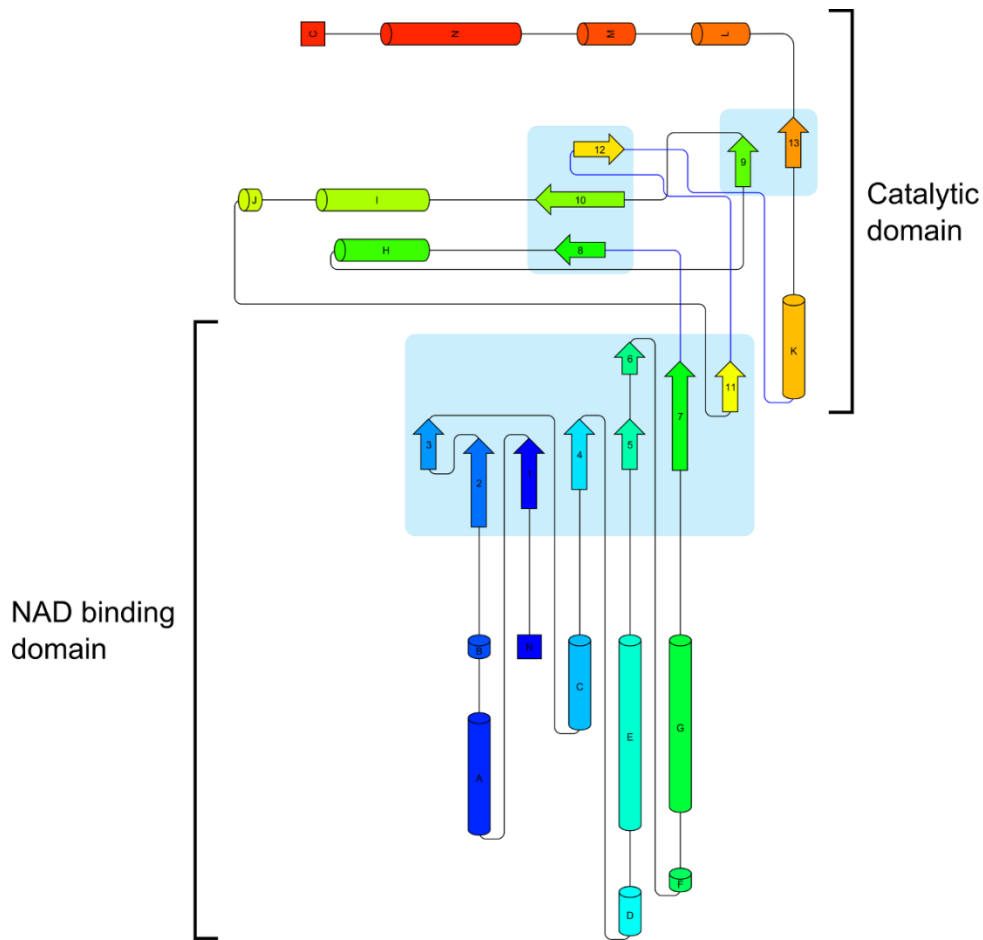
Supplementary Figure 3. Protein topology diagram of TDH from *T. brucei*. From the N-terminus (N) to C-terminus (C), alpha helices are lettered from A to Z and beta strands are numbered in ascending order. The secondary structure elements are coloured from blue, starting at the N-terminus, through green, yellow and orange to red at the C-terminus. This diagram was drawn using Pro-origami software and a monomeric TDH structure (from PDB: 5K4Q).
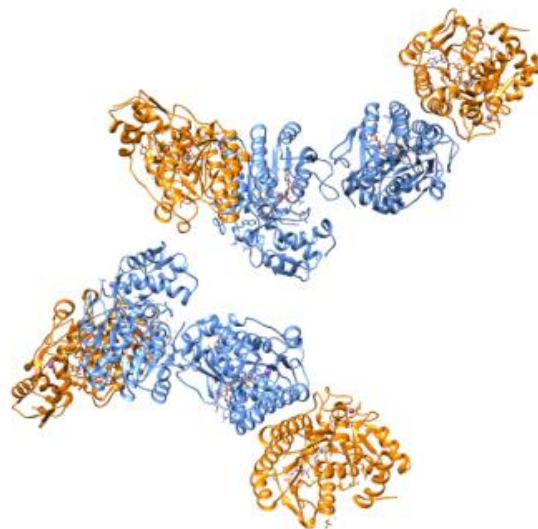
Supplementary Figure 4. Visualisation of TDH structure (PDB: 5K4W) produced by pdbset, showing all symmetry-related molecules in the crystallographic unit cell. There is one dimer in the asymmetric unit, and four dimers are created by crystallographic symmetry. Subunits within the same dimer are coloured differently (blue and gold).
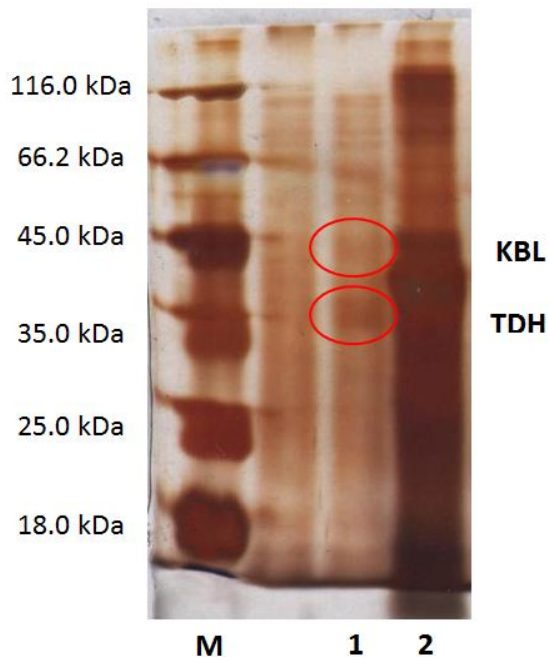
Supplementary Figure 5. Silver stain of an SDS-PAGE gel indicating the presence of TDH and KBL in fractions that corresponded to peaks on the chromatogram resulting from the size-exclusion chromatography of a mixture of TDH and KBL (See Supplementary Table 2). M = molecular weight standards; 1 = fraction at 12-14ml elution (peak at 12.5ml). 2 = fraction at 16-18ml(peak at 16.75ml). The protein bands in Lane 1 are light, suggesting that the concentrations of both proteins were very low. In Lane 2, which analyses a fraction corresponding to a much larger peak, the bands are less distinguishable, due to over-staining in this lane. As the protein concentrations were high, bands corresponding to both proteins can still be seen.
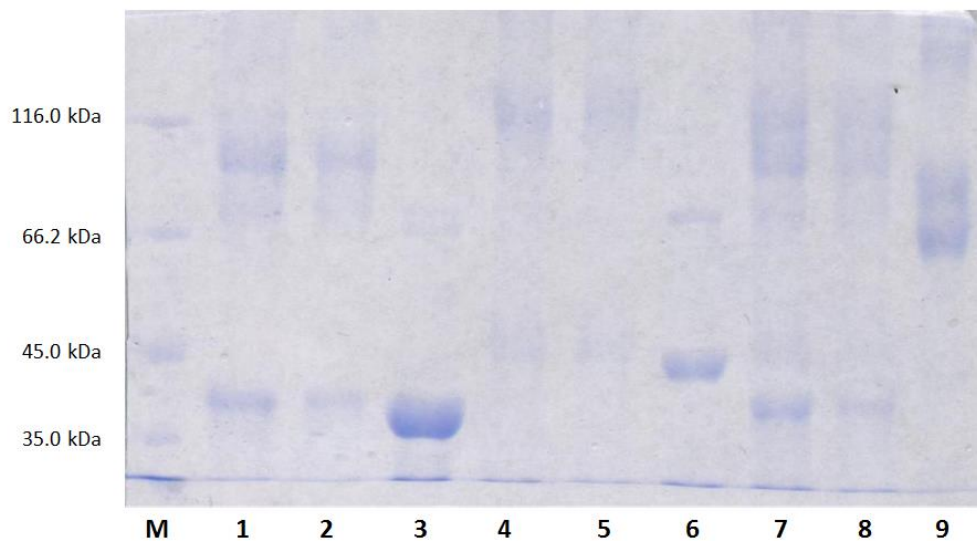
Supplementary Figure 6. SDS PAGE analysis of cross-linking experiments with TDH and KBL, on a 9% polyacrylamide gel. M = molecular weight standards; 1 = TDH (2.6 x 10-2 mM[1mg/ml]); 2 = TDH (1.3×10^{-2} mM [0.5mg/ml]); 3 = TDH (2.6 x 10^{-2} mM [1mg/ml]) non-cross-linked control; 4 = KBL (2.2×10^{-2} mM [1mg/ml]); 5 = KBL (1.1×10^{-2} mM [0.5mg/ml]); 6 = KBL (2.2×10^{-2} mM [1mg/ml]) non-cross-linked control; 7 = TDH (2.6×10^{-2} mM [1mg/ml])and KBL (2.2×10^{-2} mM [1mg/ml]); 8 = TDH (1.3×10^{-2} mM [0.5mg/ml]) and KBL (1.1×10^{-2} mM [0.5mg/ml]); 9 = BSA control (1.5×10^{-2} mM [1mg/ml]). Note that the protein bands of samples exposed to DMS appear higher due to covalent binding with DMS, leading to a higher molecular weight. This is also the cause of there being more than one band corresponding to a particular oligomer.
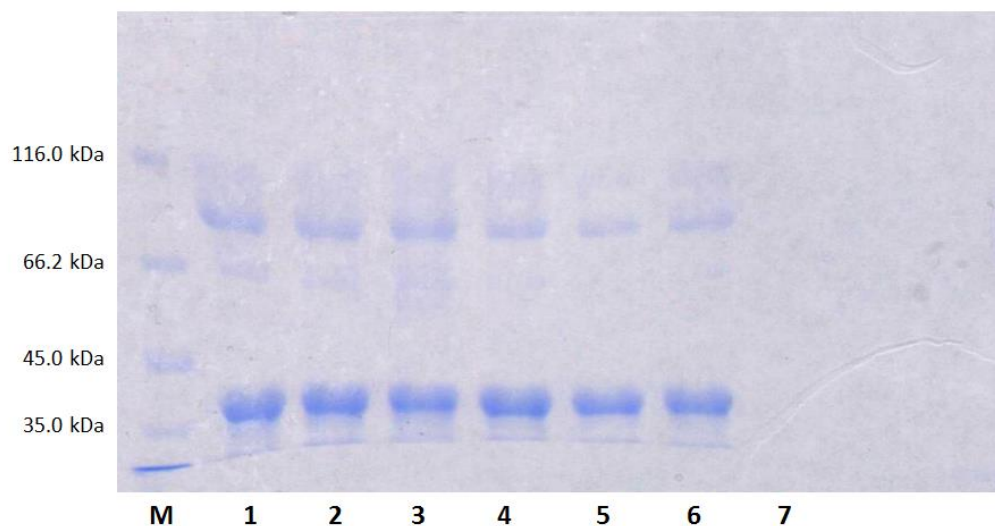
Supplementary Figure 7. SDS PAGE analysis of cross-linking experiments with TDH (at 2.6 x 10^{-2} mM [1mg/ml] concentration) on a 9% polyacrylamide gel. M = molecular weight standards; 1 = TDH; 2 = TDH and 1mM NAD$^+$; 3 = TDH and 10mM NAD$^+$; 4 = TDH, 1mM NAD+ and 30mM L-threonine; 5 = TDH, 10mM NAD$^+$ and 30mM L-threonine; 6 = TDH, 10mM NAD$^+$ and 15mM L-threonine; 7 = 10mM NAD$^+$ and 30mM L-threonine control.
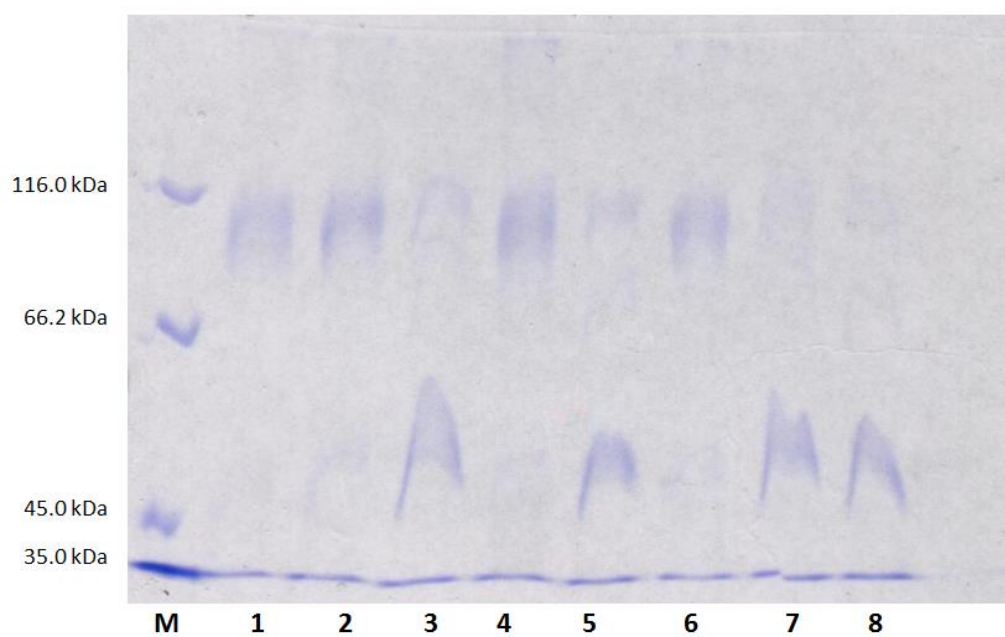

Supplementary Figure 8. SDS PAGE analysis of cross-linking experiments with KBL (2.2 x 10^{-2} mM [1mg/ml]) on a 9% polyacrylamide gel. M = molecular weight standards; 1 = KBL; 2 = KBL and 1mM PLP; 3 = KBL and 30mM glycine; 4 = KBL and 5mM A-CoA; 5 = KBL, 1mM PLP and 30mM glycine; 6 = KBL, 1mM PLP and 5mM A-CoA; 7 = KBL, 1mM PLP, 5mM A-CoA and 30mM glycine; 8 = KBL, 5mM A-CoA and 30mM glycine.
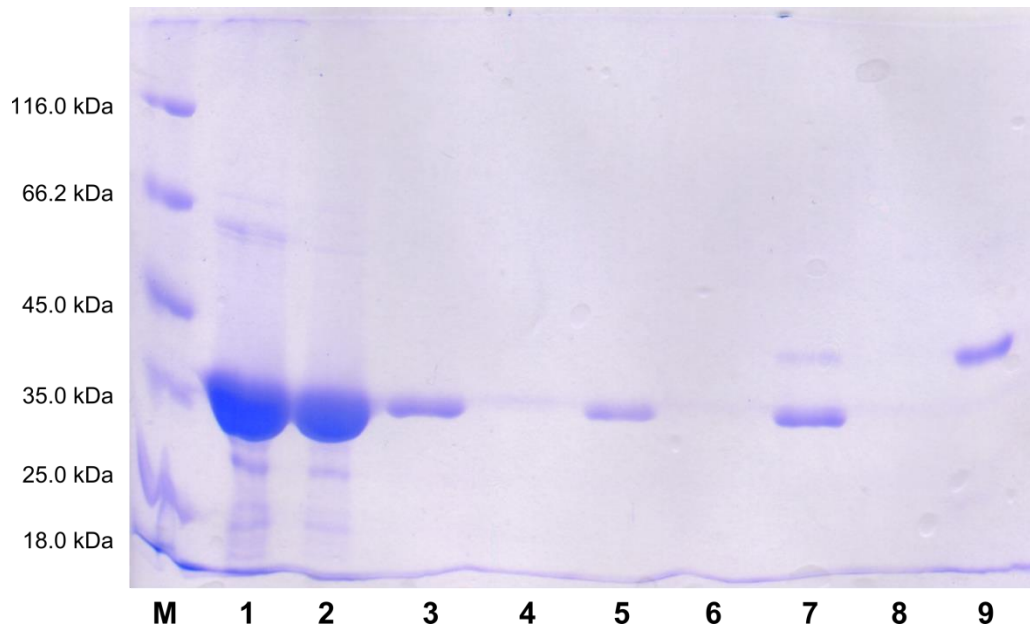
Supplementary Figure 9. SDS PAGE analysis of the pull-down assay to investigate complex formation between TDH and KBL. M = molecular weight standards; 1 = original TDH sample; 2 = TDH and thrombin; 3 = TDH with His-tag removed by washing the solution through a Ni-NTA column and benzamidine column attached end-to-end; 4 = eluate collected on elution of histidine tag from Ni-NTA column; 5 = elution of thrombin (MW approx. 37 kDa) from the benzamidine column; 6 = solution flowing through Ni-NTA column as KBL is loaded; 7 = solution flowing through the Ni-NTA column as TDH is loaded; 8 = solution flowing through the Ni-NTA column as the column is washed with buffer; 9 = eluate containing KBL after applying elution buffer to the column.
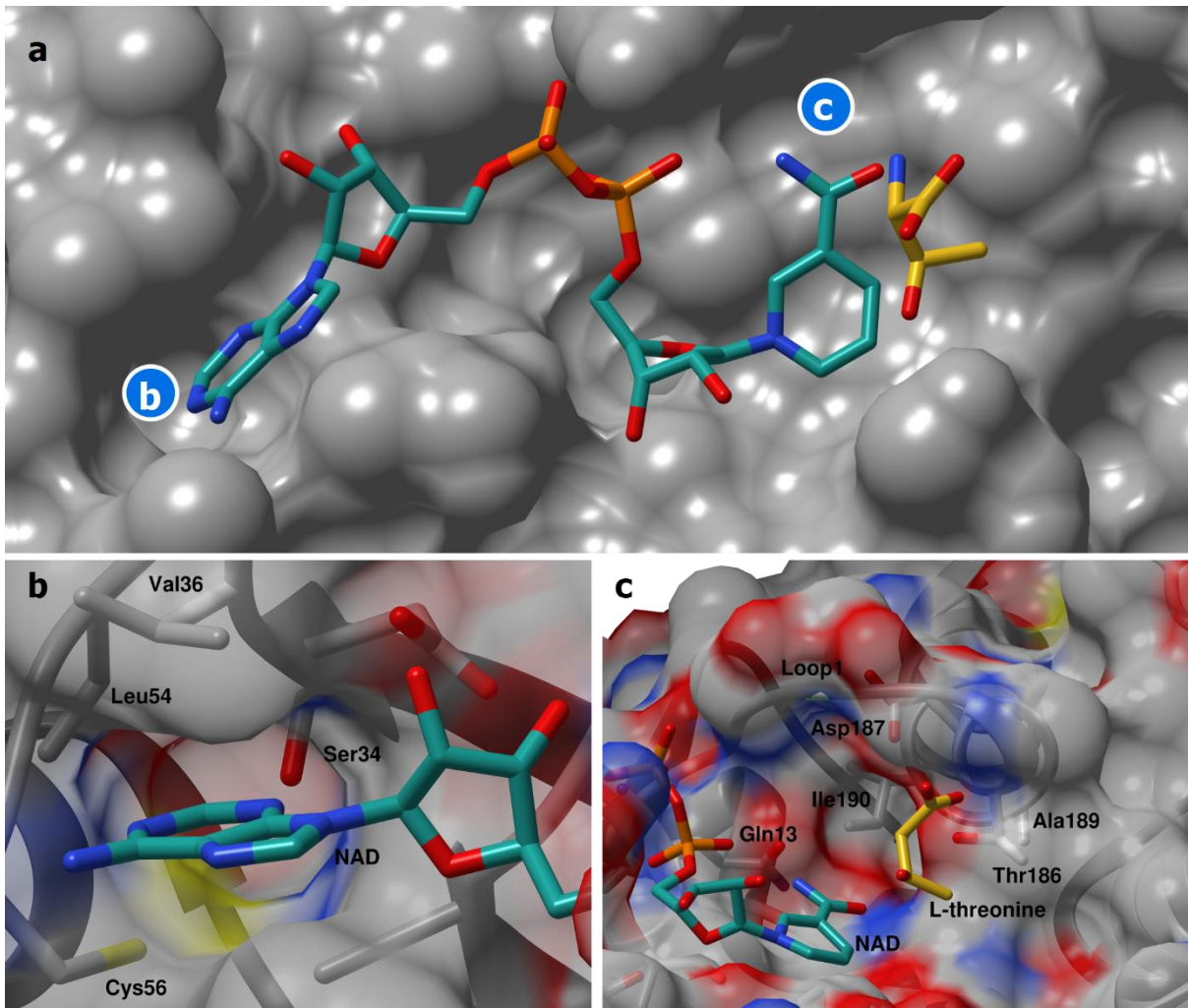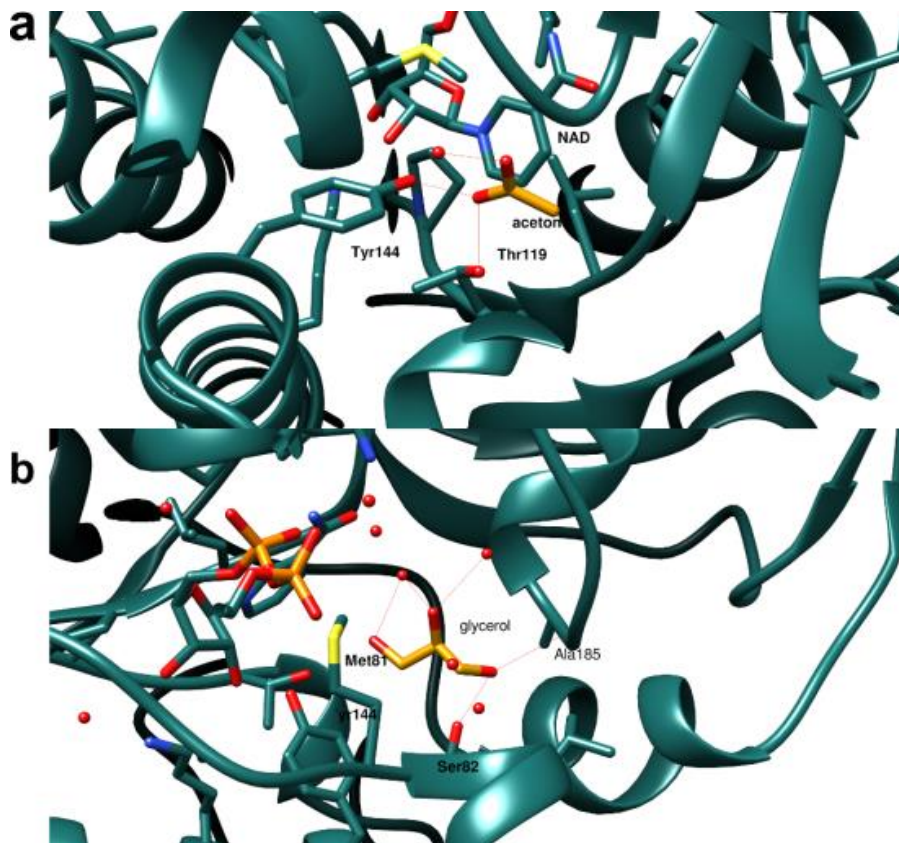
Supplementary Figure 10. Additional binding pockets in the NAD binding site. Panel a shows a cross-section of TDH, focusing on the NAD binding site. The foci of panels b (additional pocket adjacent to C2 atom of adenine) and c (additional pocket close to nicotinamide amide) are highlighted in panel A.
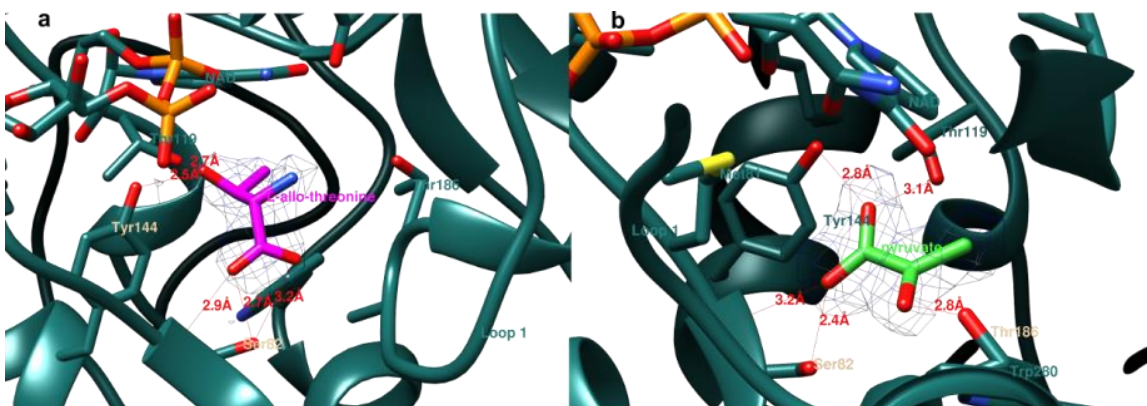

Supplementary Figure 11. Binding of L-*allo*-threonine (a, PDB: 5K50) and pyruvate (b, PDB: 5LC1) to TDH in the L-threonine binding pocket. Hydrogen bonds are indicated by red lines, electron density of bound ligands is represented by a blue mesh.
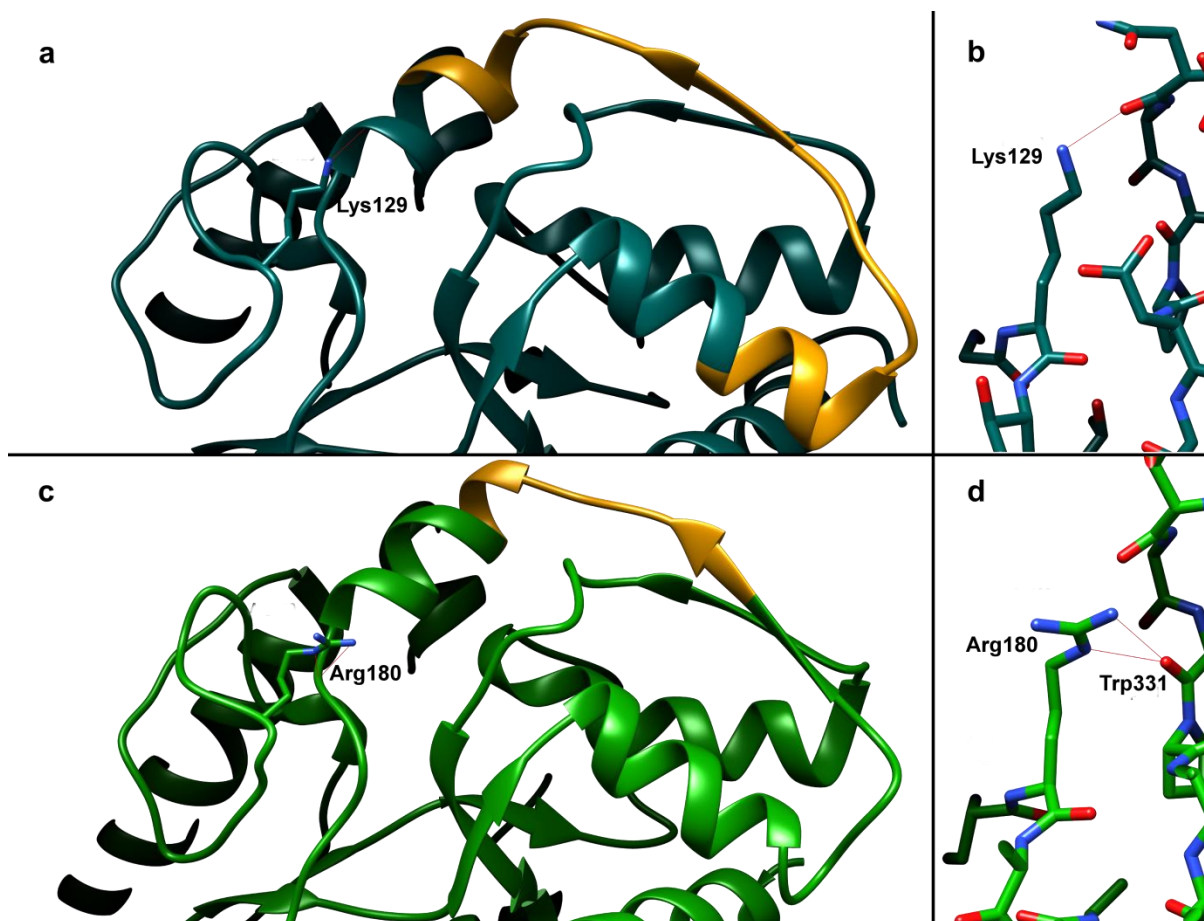
Supplementary Figure 12. Equivalent positions of Lys129 and Arg180 in *Tb*TDH (top, PDB: 4YR9) and *Mm*TDH (bottom, PDB: 5K4W), respectively. Panels a and b correspond to *Tb*TDH and panels c and d correspond to *Mm*TDH. Regions identified as conformationally variable in this study (for TbTDH) and by He et al. (for *Mm*TDH) are coloured gold in panles a and b. Hydrogen bonds are represented by red lines.

Supplementary Figure 13. Binding poses of acetone (a) and glycerol (b) in the L-threonine binding pocket. These positions were observed in more than one monomer and/or structure model.
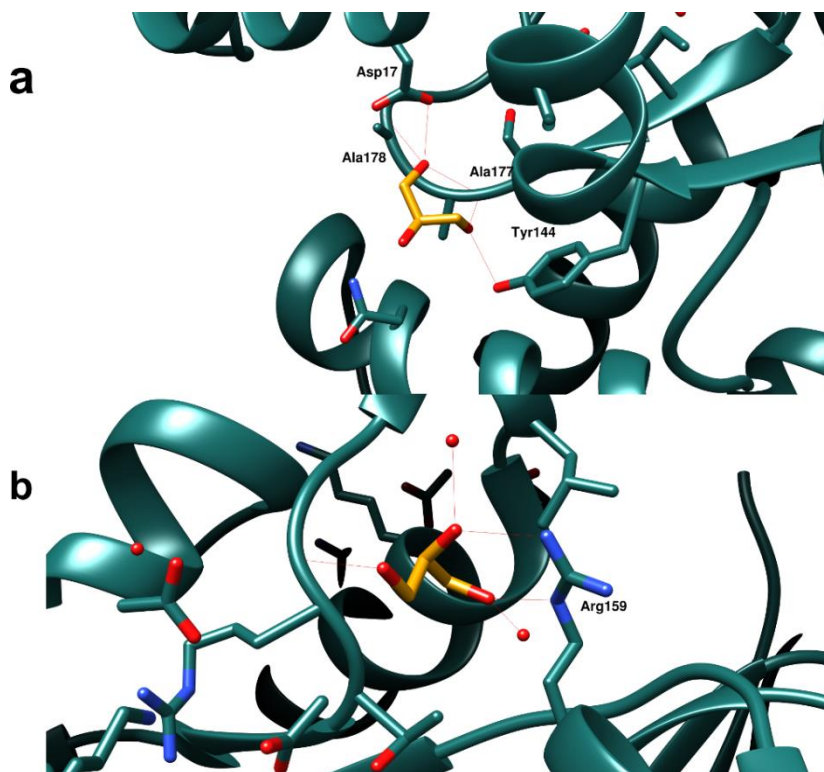
Supplementary Figure 14. Additional binding positions of glycerol on TDH. Glycerol molecules were often found adjacent to residues Asp17 (a) and Arg159 (b).
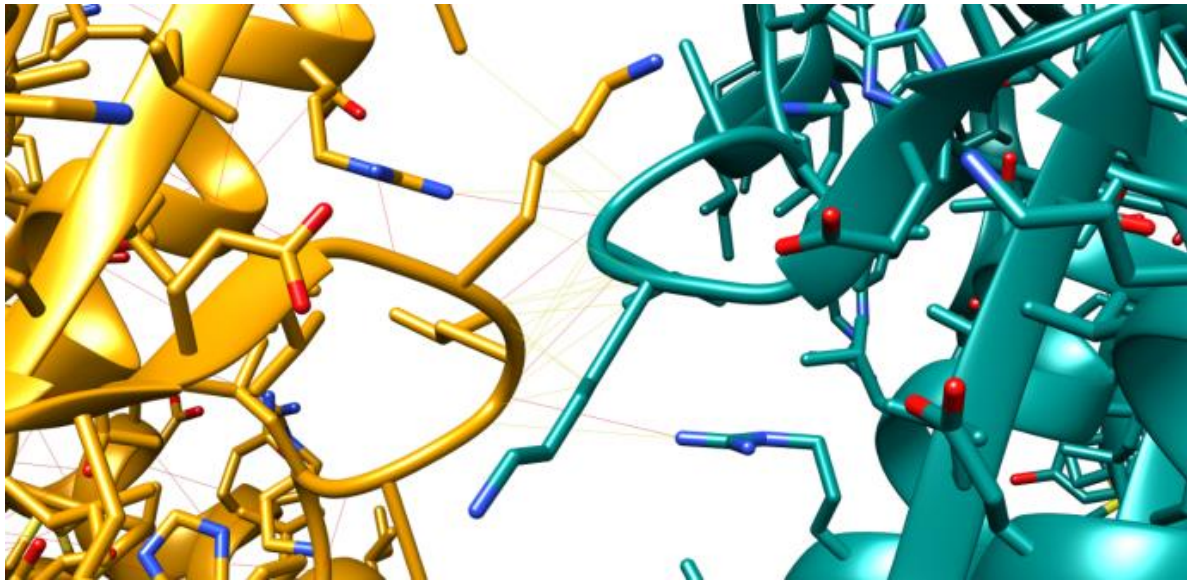
Supplementary Figure 15. An interaction between two TDH monomers (PDB: 5K50) at the Loop 2 region (Pro44-Gly50). Hydrogen bonds between residues on different subunits are depicted with red lines.
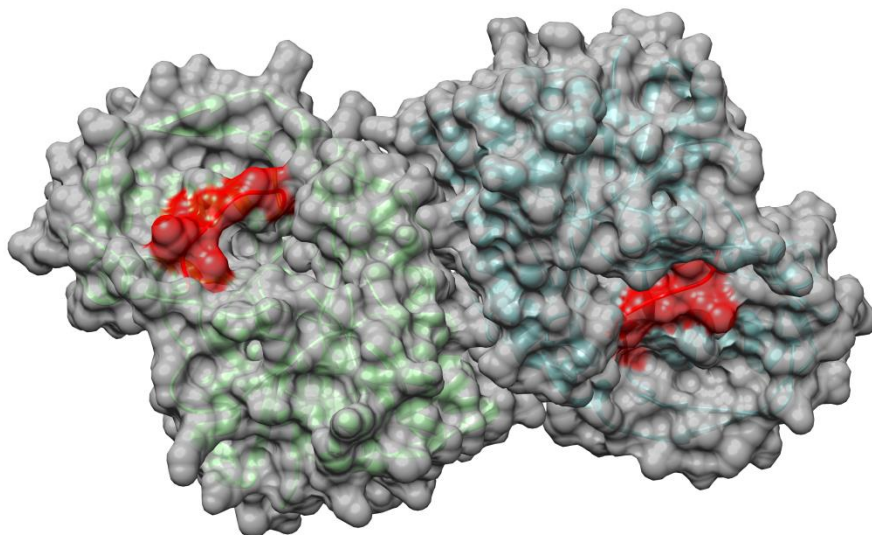
Supplementary Figure 16. Combined ribbon and surface representation of the TDH dimer, determined in crystallographic model PDB: 5K4U. In one subunit (green ribbon, left) Loop 1 is in the open conformation. In the second subunit (blue ribbon, right) Loop 1 is in the closed conformation. Loop 1 is coloured red in both subunits.