Using Fragment Cocktail Crystallography To Assist Inhibitor Design of *Trypanosoma brucei* Nucleoside 2-Deoxyribosyltransferase†

Jürgen Bosch,[⊗,‡,§] Mark A. Robien,[⊗,‡] Christopher Mehlin,[⊗,‡] Erica Boni,[⊗,‡] Aaron Riechers,[#] Frederick S. Buckner,[‡,#] Wesley C. Van Voorhis,[‡,#] Peter J. Myler,[||] Elizabeth A. Worthey,[||] George DeTitta,[⊥] Joseph R. Luft,[⊥] Angela Lauricella,[⊥] Stacey Gulde,[⊥] Lori A. Anderson,[⊗,‡] Oleksandr Kalyuzhniy,[⊗,‡] Helen M. Neely,[⊗,‡] Jenni Ross,[⊗,‡] Thomas N. Earnest,[∞] Michael Soltis,[×] Lori Schoenfeld,[⊗,‡] Frank Zucker,[⊗,‡] Ethan A. Merritt,[⊗,‡] Erkang Fan,[⊗,‡] Christophe L. M. J. Verlinde,[⊗,‡] and Wim G. J. Hol*[,⊗,‡,§]

Department of Biochemistry, Division of Infectious Disease, and Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, Structural Genomics of Pathogenic Protozoa (SGPP) Consortium, Seattle, Washington 98185, Seattle Biomedical Research Institute, Seattle, Washington 98109, Hauptmann-Woodward Institute, Buffalo, New York 14203, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, and SSRL, Stanford University, Stanford, California 94309

Received April 11, 2006

The 1.8 Å resolution *de novo* structure of nucleoside 2-deoxyribosyltransferase (EC 2.4.2.6) from *Trypanosoma brucei* (TbNDRT) has been determined by SAD[a] phasing in an unliganded state and several ligand-bound states. This enzyme is important in the salvage pathway of nucleoside recycling. To identify novel lead compounds, we exploited "fragment cocktail soaks". Out of 304 compounds tried in 31 cocktails, four compounds could be identified crystallographically in the active site. In addition, we demonstrated that very short soaks of ∼10 s are sufficient even for rather hydrophobic ligands to bind in the active site groove, which is promising for the application of similar soaking experiments to less robust crystals of other proteins.

Introduction

Trypanosomatids are parasitic unicellular eukaryotic organisms, infecting plants as well as animals. Several species within the family are responsible for serious, but largely neglected, diseases of humans and domestic animals. Most of these are distributed in tropical and subtropical areas of the world. In sub-Saharan Africa, *Trypanosoma brucei*, the causative agent of sleeping sickness, is transmitted by the tsetse fly. If untreated, this disease is fatal to humans. Estimations by the WHO indicate nearly 60 million people at risk.[1,2] In Central and South America *Trypanosoma cruzi* causes Chagas' disease, with an estimated 100 million people at risk. Furthermore, some *Trypanosoma brucei* subspecies have an economic impact on cattle farming, since they cause a disease called "nagana".[1,2]

Nucleoside 2-deoxyribosyltransferase (NDRT,[a] EC 2.4.2.6) is essential for nucleoside recycling of purine and pyrimidine phosphorylases in organisms missing those enzymes.[3] It catalyzes the cleavage of the glycosidic bonds of 2-deoxyribonucleosides and transfers the deoxyribose to another base, thereby playing an important role in recycling nucleosides. Biochemical evidence of NDRT activity has been reported in the kinetoplastid organism *Crithidia luciliae*.[4] NDRTs can be divided into two classes depending on their substrate specificity: class I enzymes transfer deoxyribose between two purines, whereas class II enzymes transfer deoxyribose between either purines or pyrimidines. The structures of class I and class II enzymes are very similar to each other and differ mainly in one loop shielding the active site.[3,5] Both enzymes are found in various *Lactobacillus* species where the active enzyme consists of a trimer of dimers.[6]

Trypanosoma brucei lacks enzymes of *de novo* purine biosynthesis and, thus, is entirely dependent on scavenging purine nucleosides from the host. Scavenged nucleosides undergo metabolism by purine salvage/recycling enzymes in order to provide the necessary nucleotides for cellular function and replication. Since NDRT is involved in nucleoside salvage/recycling, specific inhibition of this enzyme could have dramatic effects on the growth and replication of *Trypanosoma*. We therefore not only solved the unliganded structure of the enzyme from *T. brucei* but also employed a fragment-based cocktail screening in combination with crystal X-ray analysis in order to identify potential leads for future drug development. This work was performed as part of the Structural Genomics of Pathogenic Protozoa (SGPP) consortium effort to identify structural targets for the development of antiprotozoan therapeutics (see also www.sgpp.org).

Crystallographic screening methods have been used to sample large compound libraries in order to detect novel lead compounds for drug development. Ligands are detected by extra features in the electron density map compared to maps of unliganded protein. The crystallographic method reveals immediately the chemical environment of the bound ligand and hence can be utilized directly for lead optimization. The term "fragment" describes a low molecular weight compound (∼100−300 Da), which typically is smaller than druglike molecules. Fragments cover a wider range of "chemical space" usually leading to higher hit rates and are therefore easier to exploit. An initial and probably first exploration of the "cocktail soak" crystallographic procedure was performed in the early 1990s at the University of Groningen.[7] More sophisticated procedures are currently a popular tool for early lead discovery strategies.[8−12]

† Coordinates and structure factors have been deposited with the PDB (accession codes 2A0K, 2F2T, 2F64, 2F62, 2F67).

* To whom correspondence should be addressed. Phone: (206) 685-7044. Fax: (206) 685-7002. E-mail: wghol@u.washington.edu.
⊗ Department of Biochemistry, University of Washington.
‡ Structural Genomics of Pathogenic Protozoa (SGPP) Consortium.
§ Howard Hughes Medical Institute, University of Washington.
# Division of Infectious Disease, University of Washington.
|| Seattle Biomedical Research Institute.
⊥ Hauptmann Woodward Institute.
∞ Lawrence Berkeley National Laboratory.
× Stanford University.

[a] Abbreviations: SAD, Single-wavelength Anomalous Diffraction; NDRT, nucleoside 2-deoxyribosyltransferase; SeMet, selenomethionine.

10.1021/jm060429m CCC: $33.50
Published on Web 09/02/2006

Table 1. Data Collection and Refinement Statistics[a]

PDB ID	2A0K	2F67	2F2T	2F64	2F62
Ligand	-	Benzo[CD]indol-2(1H)-one	5-Aminoisoquinoline	1-Methylquinoline-2(1H)-one	2-Ethylbenzylalcohol
		12B	5IQ	12Q	12M
Beamline	SSRL 9-1	ALS 8.2.2	ALS 8.2.2	SSRL 1-5	ALS 8.2.2
Space group	C2	C2	C2	C2	C2
Unit cell	60.00 75.16 87.36 90.00 90.03 90.00	58.93 75.51 86.34 90.00 90.01 90.00	58.96 75.49 86.32 90.00 90.05 90.00	59.83 75.35 87.05 90.00 90.01 90.00	58.61 75.91 85.96 90.00 90.03 90.00
Wavelength (Å)	0.9798	0.99999	0.9202	0.9202	0.9202
Resolution range (Å)	50 - 1.8	30 - 1.6	30 - 1.7	30 - 1.6	30 - 1.5
Number of reflections	331137	382624	220454	247117	248477
Number of unique reflections	32301	42601	36740	46217	49499
Completeness (%)	94.42(74.2)	89.9 (56)	92.8 (67.2)	95.6 (92.5)	86.7 (45)
Redundancy	4.2 (3.4)	3.5 (2.6)	3.1 (2.2)	2.3 (2.2)	3.2 (2.0)
Mean I/σ(I)	8.6 (2.7)	7.8 (2.5)	10.7 (3.9)	11.0 (4.3)	13.4 (5.4)
R_{sym} on I	10.9 (45.8)	9.4 (28.2)	6.6 (21.7)	4.9 (16.5)	3.4 (14.3)
Wilson B ($Å^2$)	14,9	20,4	18,1	15,9	14,9
Final Refinement Statistics					
Number of protein atoms	2979	2931	3053	3067	3005
Number of water molecules	416	304	396	374	332
Number of ligand atoms	-	22	20	24 x 2	26
$B_{Protein}$ ($Å^2$)	16.6	19.2	12.5	17.1	11.9
B_{Water} ($Å^2$)	27.3	31	26.7	28.3	25
B_{Ligand} ($Å^2$)	-	40	16.9	37	23.6
Number of reflections in refinement	32301	42601	36740	46217	49499
Number of reflections in test set	1704	2931	1928	2484	2675
R_{work} (%)	17,5	17,9	17,8	16,6	16,8
R_{free} (%)	20,6	20,8	20,4	18,7	18,6
rms deviations from ideal geometry					
bonds (Å)	0.08	0.008	0.006	0.006	0.007
angles (degree)	0.953	1.026	0.863	0.934	0.968

[a] In the unliganded structure of TbDNRT, the rms deviation for the Cα atoms between the two subunits of the dimer in the asymmetric unit is 0.021 Å; for all other atoms it is 0.31 Å. The rms deviation for the *B* factors between all proteins atoms of the two subunits is 14.4 $Å^2$.

Here, commercially available compounds were chosen on the basis of molecular weight range, a calculated log *P* range, a maximum number of rotatable bonds, the presence of at least one ring, and the absence of reactive functional groups. Subsequently the remaining compounds were clustered in groups of 10 on the basis of maximum shape diversity. Of the 70 virtual cocktails obtained, 31 were made thus far. A detailed description of the cocktail selection, contents, and modus operandi for dissolution is in preparation (Verlinde et al.).

Here, we report the crystal structure of *T. brucei* NDRT (TbNDRT, SGPPID:Tbru015777AAA, GeneDB Tb05.30H13.400) and the results of probing its active site by fragment cocktail crystallography. Interestingly, NDRTs are absent in humans, where the nucleoside recycling is carried out by purine and pyrimidine phosphorylases, making this enzyme worth exploring as a drug target.

Results

Overall Structure of the Unliganded TbNDRT. The de novo structure of TbNDRT was solved to 1.8 Å by applying SAD X-ray techniques to SeMet substituted protein. The refined structure has a R_{work} of 17.5% and a R_{free} of 20.6%, with excellent geometry (Table 1). The overall fold of the monomer is an α/β domain with approximate dimensions of 76 Å × 39 Å × 42 Å, while the dimer has dimensions of 76 Å × 60 Å × 54 Å. A central, parallel, five-stranded β-sheet with topology 12347 and a short α-helix (α1) form the core of the molecule with α-helices α2 and α7 on one side and with α3 and α4 on the other side of the β-sheet. Additionally two smaller helices (α5, α6) and two β-strands (β5, β6) are observed in the proximity of the substrate binding pocket (Figure 1A). A SO_4^{2-} ion is bound to each of the N-terminal His_6 tags (Figure 1B), which are involved in crystal packing.

A dimer is observed in the asymmetric unit, with the dimer interface burying a ∼4000 $Å^2$ solvent accessible surface as reported by the PROFUNC server.[13] The dimer interface mainly involves residues from helices α3, α5, and α6. There is no evidence for a higher molecular arrangement (e.g., a hexamer); both DLS and size exclusion experiments indicate a dimer in solution for TbNDRT (data not shown). Also, crystal contacts of the dimer with neighboring subunits do not result in large buried surfaces.

The active site nucleophile Glu90[5] is buried in a 14 Å deep and 7 Å wide L-shaped pocket, which also includes one glycerol molecule at the bottom of the pocket and several waters (Figure 1B). The volume of each of the two active site pockets per dimer, calculated with a probe radius of 1.4 Å using AREAimol,[14] is ∼1180 $Å^3$ and is formed by 13 residues of one subunit and 6 residues of the other subunit. The total number of residues forming the pocket is 19, with 9 hydrophobic, 4 charged, and 6 neutral side chains.

The closest structural homologue to TbNDRT, according to the EBI SSM server,[15] is NDRT from *Lactobacillus helveticus* with a rmsd of 2.1 Å for 165 residues and a sequence identity of 26% (PDB accession code 1S2I).[5] The biological unit in *Lactobacillus helveticus* is a hexamer formed by a trimer of dimers.[6] The "building blocks" of the *Lactobacillus helveticus* hexameric enzyme correspond to the dimer of TbNDRT.

Fragment Cocktail Crystallography. "Fragment cocktail soaking" was carried out at 4 °C. A total of 69 crystals were examined probing 304 different compounds in 31 cocktails, including variations in soaking times for several of the cocktails. Diffraction was evaluated by automated screening at SSRL beamlines 1-5, 9-1, 9-2 and manual screening at ALS 8.2.1, 8.2.2, and 5.0.3. Figure 2A displays the resolution distribution for these experiments, demonstrating that very few of the

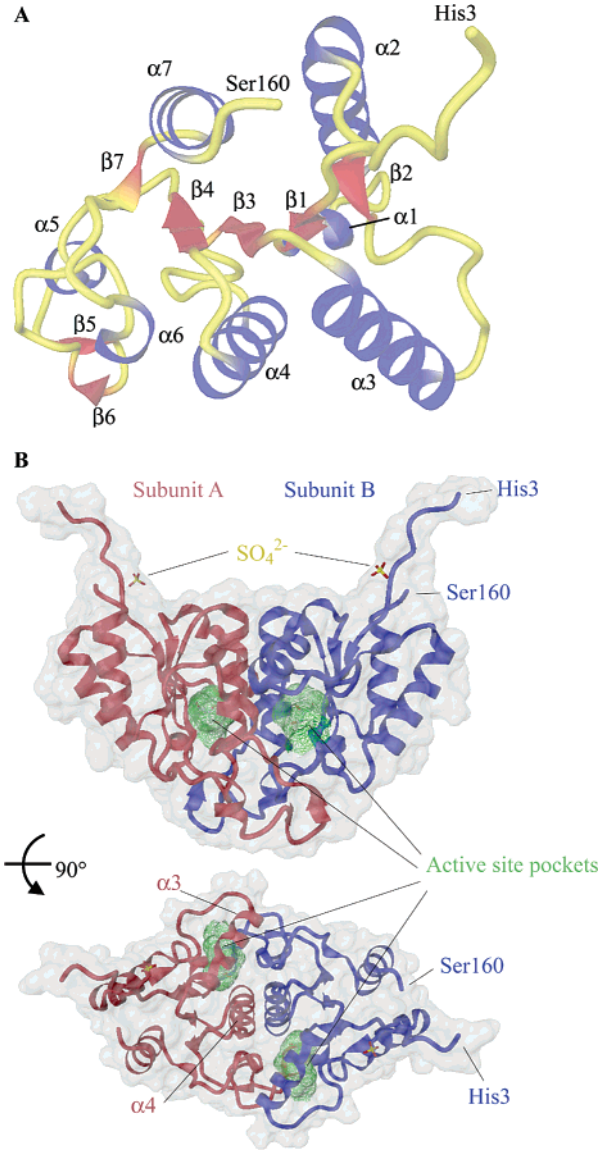
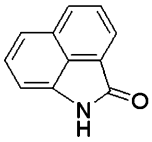
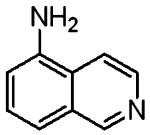
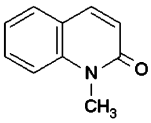
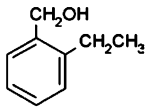


Figure 1. Structural overview of *T. brucei* nucleoside 2-deoxyribosyltransferase (TbNDRT): (A) TbNDRT subunit representation, where α-helices are shown in blue and *β*-strands in red; (B) transparent surface representation of the TbDNRT dimer showing subunit A in red and subunit B in blue, where the active site pocket is highlighted in green and the SO_4^{2-} ion is shown as a stick model.

Table 2. In Vitro Inhibition Assay Results

Ligand structure	Name (Abbreviation)	ED50 (mM)	PDB code
	Benzo[cd]indol-2(1H)-one (12B)	0.120	2F67
	5-aminoisoquinoline (5IQ)	0.252	2F2T
	1-methyl-2(1H)-quinolone (12Q)	0.413	2F64
	2-ethylbenzylalcohol (12M)	1.34	2F62

fragment cocktails had a detrimental effect on the resolution for these particular protein crystals. The fragment cocktail soaks that resulted in diffraction worse than 3 Å were rescreened with shorter incubation times in a second experiment and then diffracted better than 2.5 Å. Soaking protein crystals in 10% DMSO enriched mother liquor without any cocktail compound also showed no significant effect on the resolution limit even after 1 week of incubation. A typical example of a cocktail composition is represented in Figure 2B. The full set of cocktails will be published elsewhere (Verlinde et al., in preparation).

X-ray data sets of fragment cocktail soaked crystals were analyzed by molecular replacement using the unliganded TbNDRT coordinates 2A0K. We were able to identify four ligands in the active site (Table 2, Figure 3). Two data sets showed additional density in the active site, but the density could not be identified as belonging to a specific compound yet. These cocktail soaks need to be further investigated by follow-ups with single-compound soaks.

Time Course Experiment. After identification of the first compound benzo[*cd*]indol-2(1*H*)-one in the active site by a 12 h soaking experiment, a time course experiment was performed in order to evaluate and establish requirements for future successful binding of ligands. Crystals were first isolated and equilibrated in artificial mother liquor prior to fragment soaking.

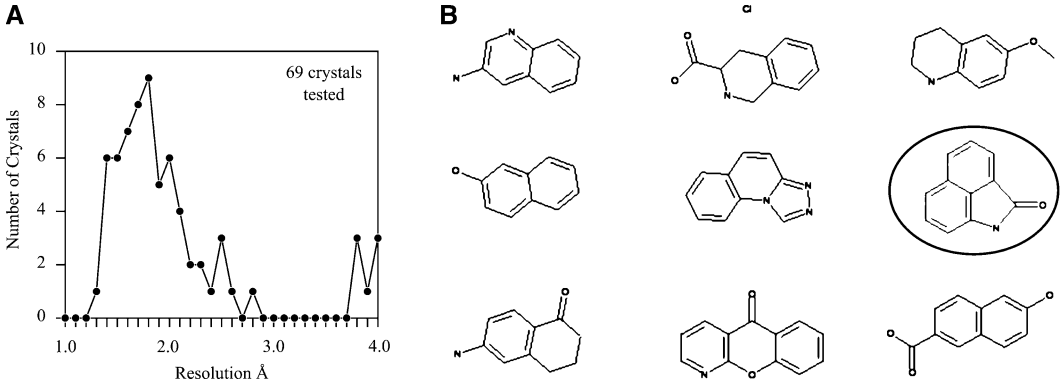


Figure 2. Fragment cocktail crystallography of TbNDRT. (A) Shown is the diffraction limit distribution plot. The graph shows a plot of number of crystals (vertical) versus their resolution limit after fragment soaking experiments for a total of 69 crystals. A cutoff value of better than 2.5 Å was used as useful limit for data acquisition. Fragment soaks with crystals diffracting worse than 2.5 Å were rescreened with shorter incubation times. (B) Shown is a schematic drawing of the compounds present in SGPP fragment cocktail no. 4. Encircled is the identified compound benzo-[*cd*]indol-2(1*H*)-one.

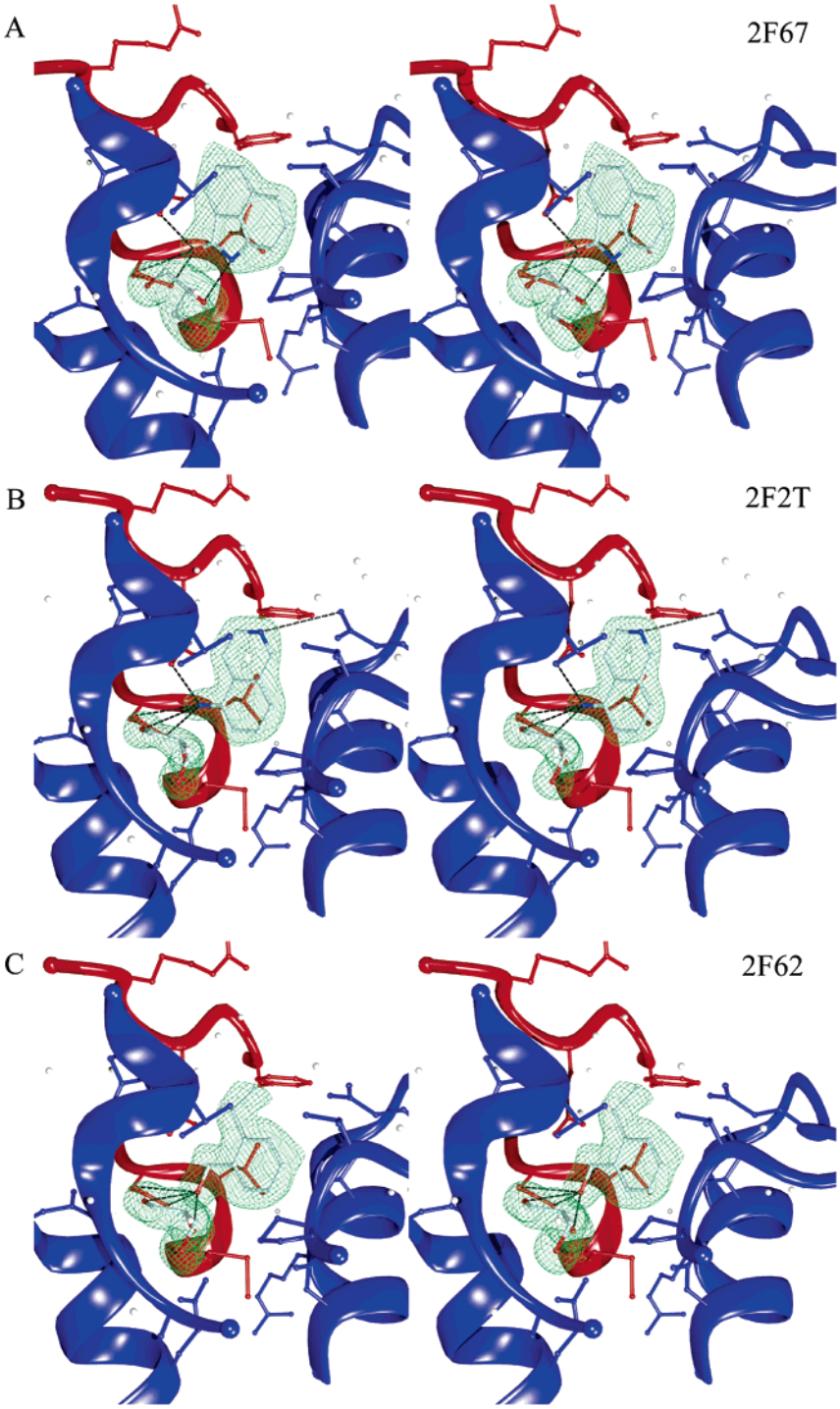


Figure 3. (Continued on next page)

Fragment cocktail no. 4 (Figure 2B) was 10-fold diluted into artificial mother liquor and thoroughly mixed. Crystals were then transferred individually and incubated for 30 min, 10 min, 3 min, 1 min, and 10 s, then flash-frozen in liquid nitrogen and examined at a synchrotron radiation source.

The ligand benzo[*cd*]indol-2(1*H*)-one could be identified unambiguously from the difference density map contoured at 2.5σ even in the shortest soaking experiment of only 10 s (Figure 3A). Furthermore, specific soaks with only benzo[*cd*]indol-2(1*H*)-one dissolved in DMSO were carried out, confirming the previous results obtained from cocktail soaking experiments (data not shown).

These results show that very short incubation times are sufficient to unambiguously reveal specific binding of small molecules in appropriate pockets, suggesting that short quick-dip experiments might be worthwhile to evaluate in cases where protein crystals are less robust to cocktail crystallography procedures.

Identification of the Bound Ligands. A total of 4 ligands from 31 cocktails containing 304 fragments in total were identified in the active site. The ligands in their observed positions are shown in Figure 4A after superimposing the 316 C^{α} atoms of the ligand-bound dimers onto the unliganded structure. The rmsd values for the C^{α} atoms ranges from 0.11 to 0.24 Å (2F2T, 0.18 Å; 2F62, 0.24 Å; 2F64, 0.11 Å; 2F67, 0.17 Å) when compared to the unliganded state. The residues representing the walls of the active site pocket show a rmsd of 0.09−0.11 Å over 19 C^{α} atoms and 0.59−0.75 Å over 153

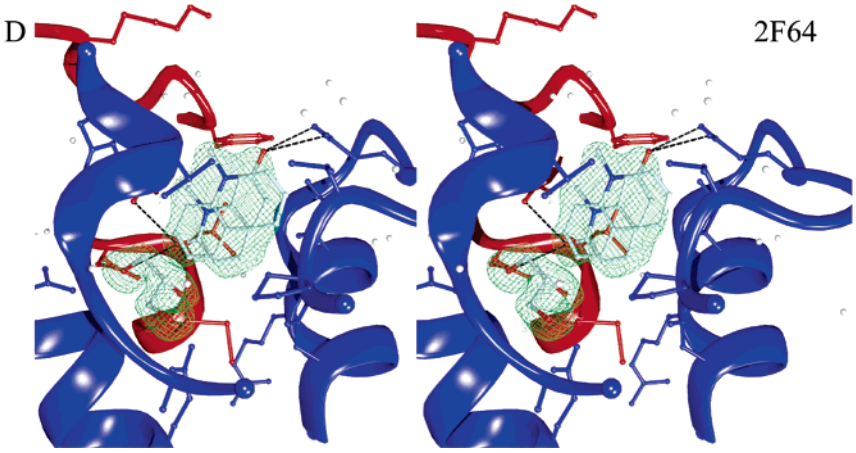


Figure 3. Stereoviews of the active site of TbNDRT with bound ligands. Only part of the C^{α} backbone (blue) is shown for clarity. Residues forming the active site pocket are depicted as ball-and-stick in colors according to the subunit. The σA weighted $2mF_O - DF_C$ maps are shown at the 1.5σ contour level in green, with the ligands omitted during structure factor calculation: (A) benzo[*cd*]indol-2(1*H*)-one (2F67); (B) 5-aminoisoquinoline (2F2T); (C) 2-ethylbenzyl alcohol (2F62); (D) 1-methylquinoline-2(1*H*)-one (2F64) in two orientations (see also text).

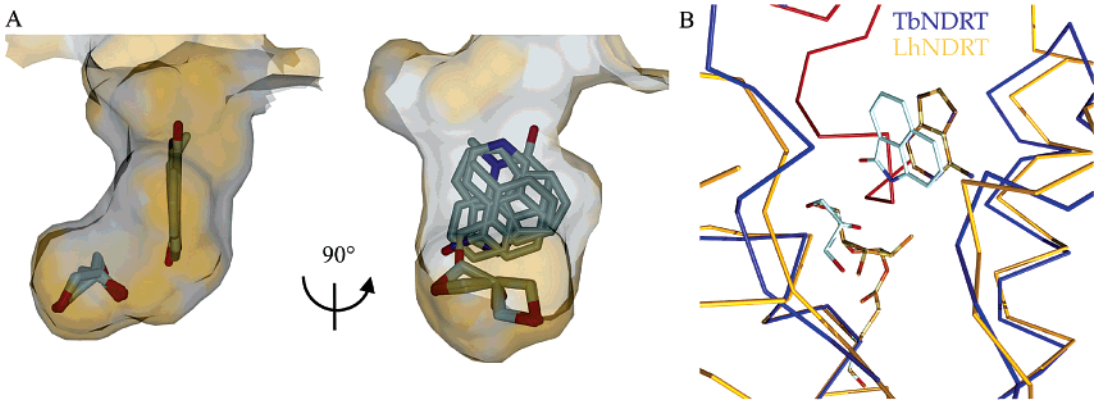


Figure 4. Comparison of ligand binding to TbNDRT. (A) Superposition of identified ligands. The ligands and the glycerol molecule are drawn as sticks. The transparent surface of the L-shaped pocket is approximately 14 Å deep and 7 Å wide at the entrance. Hydrophobic areas on the surface of TbNDRT are in gray and polar regions are in orange. (B) Superposition of *Lactobacillus helveticus* (in a ribosylated ester intermediate state with the ribose moiety covalently bound to Glu101[5] plus free adenine (orange, 1S2D)) onto ligand-bound TbNDRT (blue and red, 2F67). Benzo[*cd*]indol-2(1*H*)-one occupies almost the same position as the adenine ring, and the glycerol molecule mimics part of the ribose.

atoms when including the side chains. In all five structures a glycerol molecule was identified buried deep inside the active site pocket.

Comparing the unliganded with the four ligand-bound structures indicates a small movement upon ligand binding. Helices α3, α4, α6 and most of the β-sheet core stay rigid, while α2 and α7 move slightly inward toward the dimer interface by approximately 0.5 Å. The identified ligands are shown in more detail in Figure 3. The active site pocket shows a longitudinal distribution of polar patches and larger hydrophobic patches, as can be seen in Figure 4A. This architecture arrests the compounds in similar orientations, which vary only in rotation about an axis perpendicular to the planar face of the compounds. This binding mode has also been observed in *L. helveticus* NDRT.[5]

Superimposing structures of *Lactobacillus helveticus* NDRT (LhNDRT) with substrate analogues bound onto the present TbNDRT structures indicates the functional importance of components of the glycerol molecule at this specific position within the active site pocket in our structure. The oxygens of the glycerol molecule lie within 0.4 and 0.8 Å of the corresponding O3 and O5 atom positions observed for the bound substrate in LhNDRT. This is by itself of potential interest from a drug design point of view because such groups might be incorporated into future inhibitors (Figure 4B). This glycerol molecule originates from the 5% glycerol present in the standard SGPP protein buffer.

Ligand Interactions. The experimental electron density observed for four chemical fragments is shown in Figure 3.

Benzo[*cd*]indol-2(1*H*)-one (12B). As mentioned earlier, time course experiments were carried out with this compound as part of cocktail mixture no. 4 (Figure 2B). We also prepared single compound soaks of benzo[*cd*]indol-2(1*H*)-one (2F67) for 15 min at 4 °C, varying the concentration from 0.001 to 0.1 mM. In the case of the highest concentration soak only, the bound ligand was identified (data not shown).

The ligand is burying a 405 $Å^2$ solvent accessible surface area and is mainly stabilized by numerous hydrophobic contacts with residues Val19, Phe20, Pro46, Thr47, Glu50, Ile57, Asn61, Glu90 from chain A and Leu135 and Met136 from chain B. Additionally, the oxygen (OAL) forms a hydrogen bond to Asn134(B) OD1 and ND2 with 2.7 and 3.2 Å distance as well as to Glu127(B) OE2 at 2.8 Å. The nitrogen atom (NAE) makes a 2.9 Å hydrogen bond to the O2 atom of the neighboring glycerol from the cryoprotectant, which by itself is stabilized through contacts of its O1-hydroxyl to Asn134(B) ND2 at 3.0 Å and to Asp84(A) OD2 at 2.5 Å. Finally, the O3-hydroxyl makes contact with the backbone nitrogen of Gly16(A) at 2.8 Å (Figure 3A).

5-Aminoisoquinoline (5IQ). Here, the ligand is stabilized by hydrogen bonds from its NAJ nitrogen to Glu127(B) OE2 and Asn134(B) OD1 and O1 of the glycerol with distances of 2.9, 3.0, and 3.6 Å, respectively. All other contacts are hydrophobic in nature (Figure 3B) and, like in the case of benzo[*cd*]indol-2(1*H*)-one, involve side chains Val19, Phe20, Pro46, Ile57, Asn61 from chain A and Leu135, Met136 from chain B of TbNDRT.

2-Ethylbenzyl Alcohol (12M). The density for this compound is well-defined, as can be seen in Figure 3C. The ligand oxygen OAJ makes contacts at 2.7 and 3.3 Å to Asn134 (B) OD1 and ND2. The nearby glycerol molecule is contacted via its O1 and O2 atoms at 3.2 and 3.0 Å distances to OAJ of the ligand; all other contacts are hydrophobic (Figure 3C).

1-Methylquinoline-2(1*H*)-one (12Q). The ligand 12Q could be placed in two equally occupied conformations, as judged by the electron density (Figure 3D). In one conformation, the 12Q (OAL) hydroxyl forms hydrogen bonds to Glu127 (B) OE2 and O1 of the glycerol molecule at 3.0 Å. The second conformation makes hydrogen bonds to Glu50 (A) OE1 and OE2 at 3.1 and 2.9 Å, respectively. All other contacts are mediated by hydrophobic contacts, similar to those seen in the three previous cases.

In Vitro Screening for Effect on Growth of *Trypanosoma brucei brucei*. The four identified ligands were assayed in triplicate for their growth inhibition effect against bloodstream form *T. brucei* cell cultures. Solubility tests showed that three of the four compounds would stay in solution up to 5 mM in HMI-9 medium containing 0.5% DMSO. However, in this medium benzo[*cd*]indol-2(1*H*)-one forms a fine suspended precipitate when the concentration exceeds 1 mM. Table 2 shows that these four compounds have activity in the high-micromolar range against the parasites; benzo[*cd*]indol-2(1*H*)-one was the most active growth inhibitor with an ED_{50} of 120 μM.

Discussion

By employing X-ray crystallography and fragment cocktail crystal soaking procedures, we successfully identified four ligands in the active site of TbNDRT at high resolution. This enabled us to probe the chemical environment of the active site, which might be beneficial for future development of compounds with higher affinity. We also discovered adventitiously a glycerol bound to the active site. Creating compounds that combine the hydrophobic characteristics of our fragments with, for example, mimics of glycerol in the hydrophilic "tip" of the cavity (Figures 3 and 4) could result in molecules that may outcompete natural substrates of *Trypanosoma*. It is worth mentioning that 84.2% (16 out of 19) of the pocket-forming residues in the *T. cruzi* homologue and 89.5% (17 out of 19) in *L. major* and *L. infantum* are identical, potentially allowing drug design for all of these species simultaneously (Figure 5). Observations made with substrate analogues and inhibitors in *L. helveticus* are unlikely to help significantly because only 21.0% (4 out of 19) of the pocket-forming residues are identical with *T. brucei*.

Purine salvage/recycling in trypanosomatid parasites is clearly essential because of the absence of enzymes for *de novo* purine biosynthesis. One of the processes of purine salvage/recycling entails the removal of the ribose or deoxyribose moieties from purine nucleosides. This process is accomplished by at least three types of enzymes: either purine nucleoside phosphorylases (EC 2.4.2.1), purine nucleoside hydrolases (EC 3.2.2.1), or nucleoside deoxyribosyltransferases (EC 2.4.2.6). The latter enzyme (the topic of this report) removes a deoxyribose from

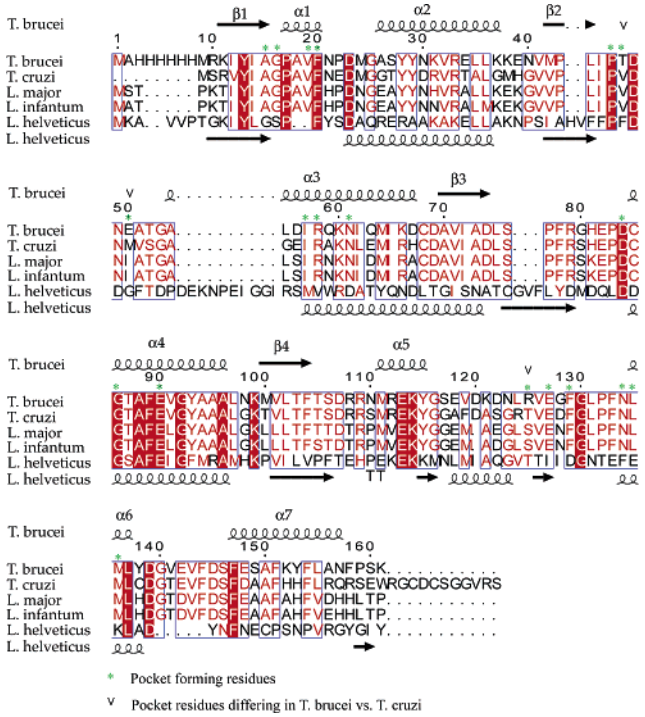


Figure 5. Structure-based sequence alignment. The secondary structure elements of TbNDRT and LhNDRT are depicted in the top and bottom rows, respectively. Conserved residues are shown in red, and pocket-forming residues of the active site groove are indicated with green asterisks. Pocket amino acid differences between *T. brucei*, *T. cruzi*, *L. major*, and *L. infantum* are indicated with arrows.

a nucleoside and transfers it to a nucleobase. In the absence of an acceptor nucleobase, NDRT displays nucleoside hydrolase activity.[5] Enzymes of these three categories are present in the genomes in *Trypanosoma brucei*, *Trypanosoma cruzi*, and *Leishmania major*. The overlapping functions of these enzymes and those of other enzymes of purine salvage have made it difficult to identify, by metabolic predictions, specific purine salvage enzymes that are essential.[16] The validation of *T. brucei* NDRT as a drug target will need to be accomplished by gene interruption or gene knockdown methods in future studies. It should be kept in mind that it may be possible to exploit the *T. brucei* NDRT by a strategy in which multiple specific components inhibit several enzymes to block purine salvage in *T. brucei*.[16] The fact that there is no NDRT homologue in humans makes *T. brucei* NDRT a particularly attractive target within the purine salvage/recycling pathway for drug development.

The crystal structures described here with bound small-molecule ligands provide the basis for developing inhibitors of the TbNDRT for potential therapeutic purposes. It is remarkable that they all bind in the active site with their ring systems essentially in the same place, relatively close to a second ligand, glycerol bound even deeper in the active site pocket. Hence, our studies not only have provided the three-dimensional structure of a potential sleeping sickness drug target but by application of a fragment cocktail crystallography strategy also show a collection of ligands bound to a critical region of TbNDRT. This combination of results is a promising platform to arrive at improved inhibitors and possibly new pharmaceuticals.

In vitro assays demonstrated that these compounds had rather weak activity against bloodstream form *T. brucei* (i.e., the most potent compound showed an IC_{50} of 120 μM). Clearly, these initial compounds that are small in size need substantial

optimization to achieve the necessary binding affinities to effectively inhibit *T. brucei* NDRT.

Materials and Methods

Protein Expression and Purification. cDNA was reverse-transcribed from *T. brucei* TREU927/4 GUTat10.1 using standard methods and subcloned in the pET14b expression vector using ligation independent cloning (LIC) techniques. BL21 Star cells were transformed using this plasmid and grown in 1 L of ZYP-5052 autoinduction media overnight at 37 °C and then further induced overnight at 18 °C. Cells were harvested and lysed by sonication in a standard buffer containing 25 mM HEPES, pH 7.5, 500 mM NaCl, 5% (v/v) glycerol, and 0.025% (w/v) sodium azide, supplemented with 1 mg/mL lysozyme and 0.2% (w/v) sodium cholate. After centrifugation, the protein was purified from clarified supernatant through application to nickel-NTA resin, washed twice with the standard buffer used for sonication plus 10 and 20 mM imidazole, respectively, and eluted with 250 mM imidazole in standard buffer.

The protein was then concentrated by centrifugation, dialyzed against SGPP standard buffer (25 mM HEPES, pH 7.5, 500 mM NaCl, 5% (v/v) glycerol, 2 mM β-mercaptoethanol, 0.025% (w/v) sodium azide), and further purified by size exclusion chromatography on a Superdex 75 column at 4 °C. The molecular weight for TbNDRT was determined through analytical size exclusion chromatography to be 35.5 kDa, which is consistent with a dimer. A final yield of 31.3 mg/L was concentrated to 27 mg/mL, flash-frozen, and stored at −80 °C for crystallization experiments.[17]

Initial Crystal Screening, Optimization, and Dynamic Light Scattering. TbNDRT was shipped on dry ice for high-throughput crystallization screening at the Hauptman-Woodward Medical Research Institute.[18] After thawing at 23 °C, crystallization experiments were set up using the microbatch under oil technique[19] with a Robbins Scientific Tango liquid handling system. Each of the 1536 experiments contained 200 nL of crystallization cocktail solution combined with 200 nL of protein solution under paraffin oil (Fluka catalog no. 76235) contained in a 1536-well plate (Greiner BioOne catalog no. 79101). The experiment plate was stored at 4 °C for 1 week and imaged at 23 °C. Images were manually reviewed; 80 of the 1536 crystallization conditions produced outcomes considered to be suitable for further optimization trials.

A subset of the initial crystallization conditions found during high-throughput screening were optimized using the vapor diffusion sitting drop method to obtain the crystals used in the current study. An amount of 400 nL of 27 mg/mL TbNDRT was mixed with 400 nL of the reservoir solution in a 96-well Intelli-plate (Hampton Research) and incubated at 4 °C. Plates were set up with a Hydra liquid handling robot, and the final optimized crystal growth conditions were as follows: 30% PEG MME 2000, 0.2 M $(NH_4)_2SO_4$, 0.1 M sodium acetate trihydrate, pH 4.6. Dynamic light scattering measurements were made using a DynaPro light-scattering instrument (Wyatt).

Fragment Cocktail Soaking. Fragment cocktails were prepared by mixing 7−10 compounds with mutually distinct shapes and chemical properties in 100% DMSO at concentrations ranging from 10 to 100 mM (Verlinde et al., manuscript in preparation).

To evaluate the effects on the crystals by addition of DMSO alone, a control experiment was carried out with a subset of crystals using a final concentration of 10% DMSO (without fragments) before attempting fragment cocktail soaks. Three time points where selected for this procedure: 10 min, 1 h, and 24 h of incubation in mother liquor containing 10% DMSO. Before cocktail soaks with a new batch of crystals were attempted, a subset was tested for good diffraction quality beforehand, as an aging effect could be observed after 5−6 weeks. Preferably batches that diffracted better than 2.5 Å on average where selected for further fragment cocktail soaking experiments.

Crystals of TbNDRT were first isolated and allowed to equilibrate for 1−5 min in cryoprotected mother liquor consisting of 30% PEG MME 2000, 10% DMSO, 0.2 M $(NH_4)_2SO_4$, 0.1 M sodium acetate trihydrate, pH 4.6, before addition of a defined amount of fragment cocktail solution dissolved in 100% DMSO to the drop. All incubation steps were carried out at 4 °C. At various time points, crystals were harvested from the drop and flash-frozen in liquid nitrogen for further examinations at synchrotron facilities. Specific soak experiments containing only the previously identified fragment were carried out with 0.1−100 mM fragment concentration and an incubation time of 15 min.

Data Collection and Structure Determination. Crystals were collected at beamlines ALS 5.0.3, ALS 8.2.1, ALS 8.2.2., SSRL 9-1, SSRL 9-2, and SSRL 1-5 at the Se K absorption edge with 1−2° rotation images and 1−3 s of exposure time depending on the mosaicity of the crystals as determined by MOSFLM.[20]

Data reduction was carried out with Wedger ELVES[21] as a front end for MOSFLM/SCALA[20] from the CCP4 package.[14] The crystals belonged either to the monoclinic space group $C2$ with $a = 60.00$ Å, $b = 75.16$ Å, $c = 87.37$ Å, and $\gamma = 90.03°$ or to space group $P1$ with $a = 41.20$ Å, $b = 41.50$ Å, $c = 42.90$ Å, $\alpha = 95.5°$,$\beta = 112.2°$, and $\gamma = 91.4°$, in both cases with two subunits per asymmetric unit resulting in a solvent content of 54.2% and 40.1%, respectively. A total of 14 Se sites were located by the program SHELXD.[22] Subsequent density modification carried out with SHELXE[23] resulted in an interpretable electron density map, which was submitted to ARP/wARP[24] for automatic model building. A complete sequence-assigned trace was obtained, and minor modifications were carried out using the program Coot.[25] Each subunit contained a SO_4^{2-} ion bound to the His_6 tag and a glycerol molecule in the active site groove. TLS refinement was performed with REFMAC5[26] using the TLS groups determined by the TLSMD Web server.[27] The TbNDRT structure without a bound ligand was refined to a crystallographic R_{work} of 17.5% ($R_{free} = 20.6\%$).[28] The rms deviations of the final model from ideal geometry are 0.008 Å for bond lengths and 0.953° for bond angles (see Table 1).[29] The final models were analyzed with validation tools in Coot[25] as well as MOLPROBITY[30] and SFCHECK.[14]

Fragment soaked TbNDRT crystal structures were solved by molecular replacement using the unliganded TbNDRT structure (2A0K) and MOLREP.[31] The anomalous signal from the Se atoms in resultant anomalous difference maps with calculated phases was a useful confirmation of the correct solution. Extra electron density was determined by applying the Coot[25] routine "find unmodeled blobs" in a σA weighted $2mF_O - DF_C$ difference density map at the 2.5σ level. After visual inspection of these blobs, the coordinates of the expected ligands were loaded into Coot[25] and an attempt was made to automatically place those ligands in the extra electron density map. Correctly placed ligands were examined manually, and only those that fitted the difference density were used for further refinement with Refmac5[26] until no further improvements were achieved. Coordinates and cif libraries for the individual ligands were generated with the PRODRG server.[32]

Parasite Cultures and in Vitro Compound Screening Assays. *T. brucei brucei* (bloodstream-form strain 427 received from K. Stuart, Seattle Biomedical Research Institute, Seattle, WA) was cultured in HMI-9 medium containing 10% heat-deactivated fetal bovine serum and 1% penicillin and streptomycin at 37 °C with 5% CO_2.[33] Compound sensitivity of the *T. brucei* strain was determined in 96-well microtiter plates in triplicate with an initial inoculum of 1×10^4 trypomastigotes per well. Compound stock solutions were prepared in DMSO at 1.0 M and diluted in HMI-9 medium to 5 mM, 0.5% DMSO, then added in serial dilutions for a final volume of 200 μL/well. Parasite growth was quantified at 48 h by the addition of Alamar Blue (Alamar Biosciences, Sacramento, CA).[34] Pentamidine isethionate (Aventis, Dagenham, U.K.) was included in each assay as a positive control. The ED_{50} values (effective dose that causes 50% growth inhibition) were calculated using Microsoft Excel and nonlinear regression in Prism (Graphpad Software, Inc.).

Acknowledgment. We thank Margaret Holmes, Jonathan Caruthers, Tracy Arakaki, Isolde Le Trong, Linda Mayhugh, Larry DeSoto, Martin Criminale, J.T. Reddy, Keith O. Hodgson,

and other members of the SGPP consortium (www.sgpp.org) for assistance and stimulating discussions. We are grateful for the *T. brucei* genomic DNA provided by Ken Stuart. We thank the staff at the ALS 5.0.3 and HHMI 8.2.1 and 8.2.2 beamlines at ALS and at the SSRL 1-5, 9-1 and 9-2 beamlines for their assistance. This project was supported in part by financial support from NIH to W.G.J.H. (NIGMS Grant No. 1P50 GM64655-01, Structural Genomics of Pathogenic Protozoa (SGPP)).

References

(1) Barrett, M. P.; Burchmore, R. J. S.; Stich, A.; Lazzari, J. O.; Frasch, A. C.; Cazzulo, J. J.; Krishna, S. The trypanosomiases (review). *Lancet* **2003**, *362*, 1469−1480.

(2) Fairlamb, A. H. Chemotherapy of human African trypanosomiasis: current and future prospects (review). *Trends Parasitol.* **2003**, *19*, 488−494.

(3) Armstrong, S. R.; Cook, W. J.; Short, S. A.; Ealick, S. E. Crystal structures of nucleoside 2-deoxyribosyltransferase in native and ligand-bound forms reveal architecture of the active site. *Structure* **1996**, *4*, 97−107.

(4) Steenkamp, D. J. The purine-2-deoxyribonucleosidase from *Crithidia luciliae*. Purification and *trans-N*-deoxyribosylase activity. *Eur. J. Biochem.* **1991**, *197*, 431−439.

(5) Anand, R.; Kaminski, P. A.; Ealick, S. E. Structures of purine 2′-deoxyribosyltransferase, substrate complexes, and the ribosylated enzyme intermediate at 2.0 Å resolution. *Biochemistry* **2004**, *43*, 2384−2393.

(6) Short, S. A.; Armstrong, S. R.; Ealick, S. E.; Porter, D. J. Active site amino acids that participate in the catalytic mechanism of nucleoside 2′-deoxyribosyltransferase. *J. Biol. Chem.* **1996**, *271*, 4978−4987.

(7) Verlinde, C. L. M. J.; Kim, H.; Bernstein, B. E.; Mande, S. C.; Hol, W. G. Antitrypanosomiasis Drug Development Based on Structures of Glycolytic Enzymes. In *Structure-Based Drug Design*; Veerapandian, P., Ed.; Marcel Dekker: New York, 1997; pp 365−394.

(8) Nienaber, V. L.; Richardson, P. L.; Klighofer, V.; Bouska, J. J.; Giranda, V. L.; Greer, J. Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nat. Biotechnol.* **2000**, *18*, 1105−1108.

(9) Carr, R.; Jhoti, H. Structure-based screening of low-affinity compounds (review). *Drug Discovery Today* **2002**, *7*, 522−527.

(10) Rees, D. C.; Congreve, M.; Murray, C. W.; Carr, R. Fragment-based lead discovery. *Nat. Rev. Drug Discovery* **2004**, *3*, 660−672.

(11) Hartshorn, M. J.; Murray, C. W.; Cleasby, A.; Frederickson, M.; Tickle, I. J.; Jhoti, H. Fragment-based lead discovery using X-ray crystallography. *J. Med. Chem.* **2005**, *48*, 403−413.

(12) Lundqvist, T. The devil is still in the details—driving early drug discovery forward with biophysical. *Curr. Opin. Drug Discovery Dev.* **2005**, *8*, 513−519.

(13) Laskowski, R. A.; Watson, J. D.; Thornton, J. M. ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res.* **2005**, *33*, W89−W93.

(14) Bailey, S. The CCP4 suite. Programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1994**, *50*, 760−763.

(15) Krissinel, E.; Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2256−2268.

(16) Craig, S. P.; Eakin, A. E. Purine salvage enzymes of parasites as targets for structure-based inhibitor design. *Parasitol. Today* **1997**, *13*, 238−241.

(17) Deng, J. P.; Davies, D. R.; Wisedchaisri, G.; Wu, M. T.; Hol, W. G. J.; Mehlin, C. An improved protocol for rapid freezing of protein samples for long-term storage. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 203−204.

(18) Luft, J. R.; Collins, R. J.; Fehrman, N. A.; Lauricella, A. M.; Veatch, C. K.; DeTitta, G. T. A deliberate approach to screening for initial crystallization conditions of biological macromolecules. *J. Struct. Biol.* **2003**, *142*, 170−179.

(19) Chayen, N. E.; Shaw Stewart, P. D.; Blow, D. M. Microbatch crystallization under oil. A new technique allowing many small-volume crystallization trials. *J. Cryst. Growth* **1992**, *122*, 176−180.

(20) Leslie, A. G. W. The integration of macromolecular diffraction data. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2006**, *62*, 48−57.

(21) Holton, J.; Alber, T. Automated protein crystal structure determination using ELVES. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 1537−1542.

(22) Schneider, T. R.; Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, *58*, 1772−1779.

(23) Sheldrick, G. M. Macromolecular phasing with SHELXE. *Z. Kristallogr.* **2002**, *217*, 644−650.

(24) Cohen, S. X.; Morris, R. J.; Fernandez, F. J.; Ben Jelloul, M.; Kakaris, M.; Parthasarathy, V.; Lamzin, V. S.; Kleywegt, G. J.; Perrakis, A. Towards complete validated models in the next generation of ARP/wARP. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2222−2229.

(25) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126−2132.

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

(27) Painter, J.; Merritt, E. A. A molecular viewer for the analysis of TLS rigid-body motion in macromolecules. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2005**, *61*, 465−471.

(28) Brünger, A. T. Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **1992**, *355*, 472−475.

(29) Engh, R. A.; Huber, R. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1991**, *47*, 392−400.

(30) Lovell, S. C.; Davis, I. W.; Adrendall, W. B.; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. Structure validation by C alpha geometry: phi,psi and C beta deviation. *Proteins: Struct., Funct., Genet.* **2003**, *50*, 437−450.

(31) Vagin, A.; Teplyakov, A. MOLREP, an automated program for molecular replacement. *J. Appl. Crystallogr.* **1997**, *30*, 1022−1025.

(32) Van Aalten, D. M. F.; Bywater, R.; Findlay, J. B. C.; Hendlich, M.; Hooft, R. W. W.; Vriend, G. PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 255−262.

(33) Hirumi, H.; Hirumi, K. Continuous cultivation of *Trypanosoma brucei* blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers. *J. Parasitol.* **1989**, *75*, 985−989.

(34) Raz, B.; Iten, M.; Grether-Buhler, Y.; Kaminsky, R.; Brun, R. The Alamar Blue assay to determine drug sensitivity of African trypanosomes (T.b. rhodesiense and T.b. gambiense) in vitro. *Acta Trop.* **1997**, *68*, 139−147.

JM060429M