Homology Modeling and Molecular Dynamics Study of NAD-Dependent Glycerol-3-Phosphate Dehydrogenase from *Trypanosoma brucei rhodesiense*, a Potential Target Enzyme for Anti-Sleeping Sickness Drug Development

Igor Z. Zubrzycki
Department of Biochemistry and Microbiology, Rhodes University, Grahamstown 6140, South Africa

ABSTRACT Sleeping sickness and Chagas disease are among the most severe diseases in Africa as well as Latin America. These two diseases are caused by *Trypanosoma* spp. Recently, an enzyme of a glycolytic pathway, NAD-dependent glycerol-3-phosphate dehydrogenase, of *Leishmania mexicana* was crystallized and its structure determined by x-ray crystallography. This structure has offered an excellent template for modeling of the homologous enzymes from another *Trypanosoma* species. Here, a homology model of the *T. brucei* enzyme based on the x-ray structure of *Lm*GPDH has been generated. This model was used as the starting point for molecular dynamics simulation in a water box. The analysis of the molecular dynamics trajectory indicates that the functionally important motifs have both a very stable secondary structure and tertiary arrangement.

INTRODUCTION

Trypanosomatids are the cause of two diseases, namely sleeping sickness in Africa and Chagas disease in Latin America. Official figures from the World Health Organization state that between 10 and 20,000 new cases of sleeping sickness occur annually, and that the total number of infected people is in the order of 500,000 (Anker and Schaaf, 2000; Moore et al., 1999). In Africa two species of trypanosomes are responsible for sleeping sickness, *Trypanosoma brucei gambiense* in West and Central Africa, and *Trypanosoma brucei rhodesiense* in Eastern Africa. The disease caused by the former pathogen, if untreated, has a chronic and protracted course lasting several years, whereas the latter disease is acute and may lead to death in a few weeks. Currently, there is no vaccine available against this disease and the drugs in use are highly toxic and often not effective because of drug resistance (Barrett, 1999). There is an urgent need for new and better drugs. Recently several different targets have been studied and suggested with the reference to the metabolism of the parasite (Barrett et al., 1999). Thus, the structures of a number of possible target points, enzymes of metabolic pathways, have been elucidated: triosephosphate isomerase for *T. brucei* (Wierenga et al., 1991); glyceraldehyde-3-phosphate dehydrogenase for *T. brucei* and *Leishmania mexicana* (Kim et al., 1995; Vellieux et al., 1993); phosphoglycerate kinase for *T. brucei* (Bernstein et al., 1997, 1998); glycerol-3-phosphate dehydrogenase (GPDH) for *L. mexicana* (Suresh et al., 2000); and aldolase for *T. brucei* (Chudzik et al., 2000). One of the most recently elucidated target points is an enzyme of the glycolytic metabolic pathway of the parasite, GPDH. Targeting the enzymes of this physiologically important metabolic pathway should deprive the parasite of the energy necessary for survival. *T. brucei* GPDH, the focus of the present study, is a polypeptide consisting of 354 amino acids forming a homodimer. It is expressed in the cytosol. From the cytosol it is targeted to the glycosome, a specialized organelle encompassing nine glycolytic enzymes (Opperdoes and Borst, 1977). The targeting process occurs with the aid of the peroxisomal targeting signal type 1 which is an obligately C-terminal tripeptide of consensus sequence SKL in *Lm*GPDH and SKM in *Tb*GPDH. *Tb*GPDH shares 63.9% identity (346 residues overlap) with the corresponding GPDH isolated from *L. mexicana*. The comparison of the *L. mexicana* sequence with the other species such as *T. brucei brucei*, *Homo sapiens*, and *Mus musculus* shows that *L. mexicana*/*T. brucei brucei* shares 63.9% identity with 346 residues overlap; *L. mexicana*/*M. musculus*, 31.6% identity with 285 residues overlap; *L. mexicana*/*H. sapiens*, 31.6% identity with 285 residues overlap; and *H. sapiens*/*M. musculus*, 94% identity with 348 residues overlap. The physiological function of the GPDH is the catalysis of the interconversion of dihydroxyacetone to L-glycerol-3-phosphate using NAD molecule as a cofactor. GPDH maintains the NAD/NADH balance in the glycosome by the reoxidation of the NADH produced by glyceraldehyde-3-phosphate dehydrogenase during glycolysis. The significant structural difference between the human and the trypanosomal system is in the binding site of adenosine portion of NAD (Fig. 1). These differences should allow to develop adenine analogs that selectively inhibit this trypanosomal enzyme (Aronov et al., 1999), rendering it inactive. The inhibition of GPDH will lead to the accumulation of dihydroxyacetone phosphate in the glycosome, which could be very harmful to the parasite. The dihydroxyacetone phos-

Submitted 7 August 2001, and accepted for publication 7 March 2002.

Address reprint requests to Dr. Igor Z. Zubrzycki, Department of Biochemistry and Microbiology, Rhodes University, Grahamstown 6140, South Africa. Tel.: 27-46-603-8081; Fax: 27-46-622-3984; E-mail: i.zubrzycki@ru.ac.za.

0006-3495/02/06/2906/10 $2.00

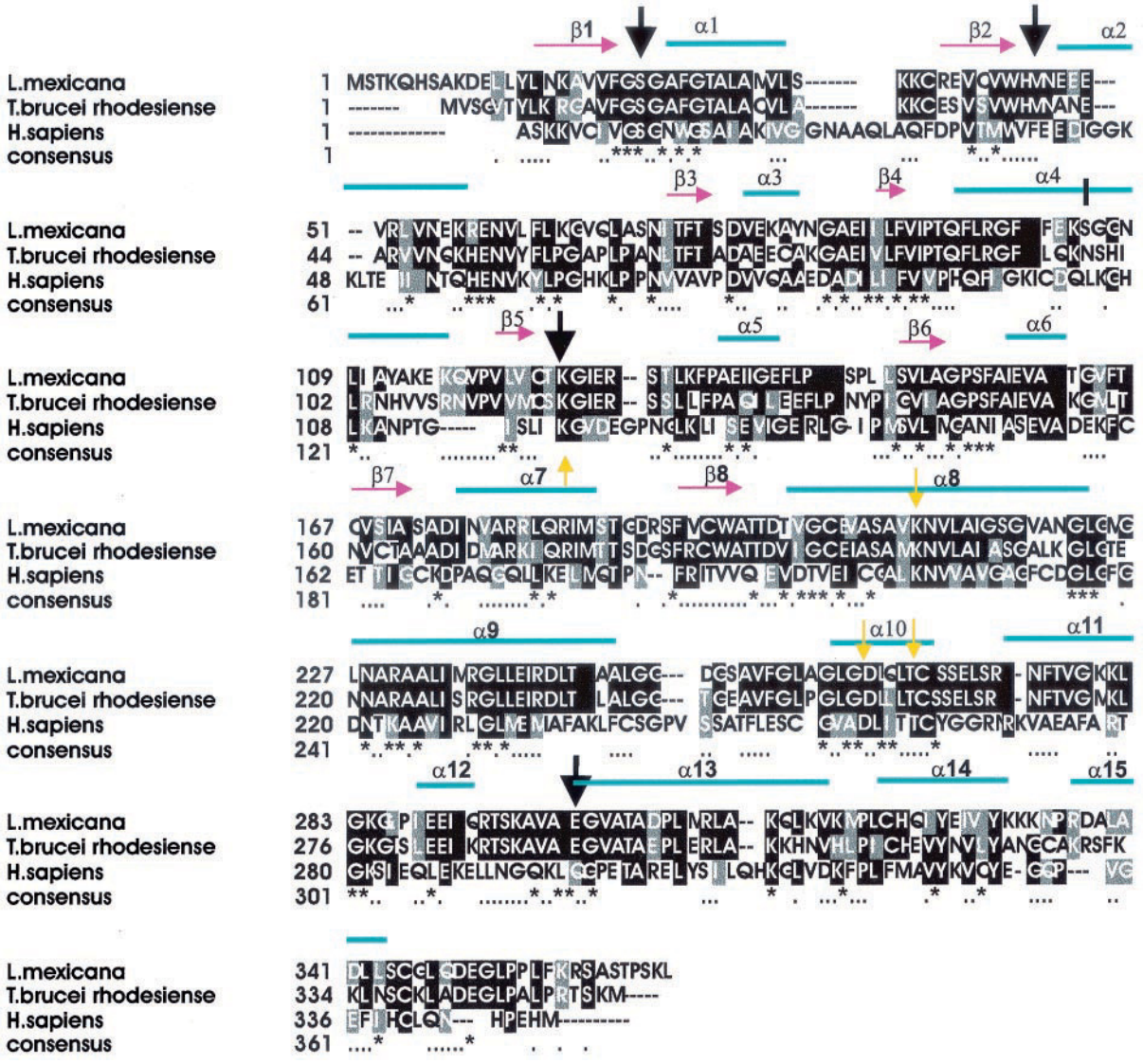

FIGURE 1 Amino acid-sequence alignment and secondary structure assignment, using DSSP program, of *L. mexicana*, *T. brucei*, and *H. sapiens* GPDH. The α-helices are shown as blue rectangles, β-sheets as red arrows, NAD-binding residues as black arrows, and active center residues as yellow arrows; black bar on helix 4 indicates the kink in the helical structure. The assignment of the secondary structure was performed using DSSP program (Kabsch and Sander, 1983).

phate molecule spontaneously converts to methylglyoxylate, which is a toxic compound reacting with proteins (Denise et al., 1999; Lo et al., 1994). Here homology modeling and molecular dynamics (MD) studies of *T. brucei* GPDH enzyme are presented. The structural features and the structural stability of this homology model are exploited for future possibility of rational in silico drug design.

METHODS

The protocol used to derive the *T. brucei rhodesiense* GPDH model is divided into three phases: sequence analysis, model building, and model evaluation. These three phases are reported and discussed in detail.

Homology modeling

The amino acid sequence (Q26756) of *T. brucei rhodesiense* was obtained from Swiss-Prot protein sequence database (http://www.expasy.ch/sprot) (Bairoch and Apweiler, 2000). A homology model was generated using Swiss-Model, an automated protein modeling server (GlaxoSmithKline, Research Triangle Park, NC) (Guex et al., 1999; Guex and Peitsch, 1997; Schwede et al., 2000). *T. brucei* modeling procedure consisted of the two steps: the first approach mode and the project mode modeling of a monomeric and homodimeric structure. The monomer homology model was built on the two templates 1EVY, and 1EVZ from the Protein Data Bank (PDB) (Berman et al., 2000) and the homodimer homology model was built on a 1EVY template. The homology-modeling phase was followed by the model evaluation phase. The structural stability of the models was tested by means of MD simulations. Using a Gromacs force field (van der

Spoel et al., 1996; van Gunsteren and Berendsen, 1987) that describes the interatomic interactions, it was possible to solve the Newton motion equations for the investigated system and to calculate how each atom's coordinates vary as a function of time.

MD simulations

The models obtained from the homology modeling were embedded in a solvent box consisting of 8211 and 15,998 simple point charge water molecules (Berendsen et al., 1981) for monomer and homodimer model, respectively, giving the total number of atoms in the simulation of 27,940 and 54,608, respectively. A twin-range cutoff was used for long-range interactions: 1.5 nm for electrostatic interactions and 1.2 nm for van der Waals interactions. The Shake algorithm was used to constrain hydrogen bond lengths (Ryckaert et al., 1977). The simulation was performed under the normal pressure and temperature conditions. Thus, a constant pressure of 1 bar applied independently in all three directions was used with a coupling constant of $\tau_p = 0.5$ ps and compressibility of 4.5 e^{-5} bar^{-1}. Water and protein molecules were coupled separately to the thermal bath at 300 K, using a coupling constant $\tau_T = 0.1$ ps (Berendsen et al., 1984). MD simulations were performed on a 2 × 733 MHz Pentium III, Dell 420 Workstation (Intel, Santa Clara, CA), taking ~79 and ~120 h of CPU time per nanosecond of simulation for monomeric and homodimeric structures. All simulations were performed using Gromacs v. 3.0 software (Berendsen et al., 1995; Lindahl et al., 2001; van der Spoel et al., 2001). Both simulations consisted of two phases: a short 20-ps canonical ensamble simulation allowing for randomization of water molecules surrounding the protein molecule and a 500-ps isobaric-isothermal ensamble simulation. The protein as well as water parameters were those of the Gromacs force field (van der Spoel et al., 1996; van Gunsteren and Berendsen, 1987).

Structural diagrams were prepared using the programs VMD (Humphrey et al., 1996), Swiss-Model (Guex and Peitsch, 1997), and Pov-Ray (http://www.povray.org). Secondary structure analysis was performed using the program DSSP (Kabsch and Sander, 1983). Other analyses were performed using scripts included with the Gromacs distribution.

RESULTS AND DISCUSSION

Sequence alignment analysis

The sequence alignment was performed using SIM–a conventional alignment tool for protein sequences (Huang and Miller, 1991). The *L. mexicana* (P90551) and *T. brucei rhodesiense* (Q26756) GPDH amino acid-sequence alignment shows a high percentage of identity between these two sequences (63.9% identity was calculated using a Block Substitution Matrix (BLOSUM62) (Henikoff and Henikoff, 1992)). The stretches of highly conserved residues within the sequence of *L. mexicana*-*T. brucei rhodesiense* have been located and are presented in Fig. 1. The residues involved in NAD binding as well as substrate binding are fully conserved (NAD-binding residues, *black arrows*; active center, *yellow arrows* (Fig. 1)). The analysis of the placing NAD-binding and active site residues versus secondary structure pattern indicates that they reside within the regions of well defined secondary structures. In contrast, quite low sequence identity among *L. mexicana*, *T. brucei rhodesiense*, and *H. sapiens* is observed (31.6% identity with 285 residues overlap). The analysis of the conservation of the residues involved in NAD-binding shows that Ser23 and Lys125 are fully conserved in *L. mexicana*, *T. brucei*, and *H. sapiens* species. However, Met46 is replaced by phenylalanine residue and E300 by glutamine in *H. sapiens*. As one could expect, the residues involved in substrate specificity (K125, K210, D263, and T 267) are fully conserved among the three species.

The analysis of the homologous models and general structural features

The quality of both homology models (monomer and homodimer) (Fig. 2, *A* and *B*), returned from SWISS-MODEL was evaluated by analysis of WhatCheck program report (Hooft et al., 1996; Rodriguez et al., 1998). The accuracy of a model was evaluated by analysis of root mean square (rms) deviation of a model from its template. Thus, the rms of the backbone atoms between the monomeric model and the two templates 1EVA and 1EVZ is equal (0.29 Å and 0.23 Å) and the rms for the homodimeric model and 1EVA template structure is equal (0.17 Å). The inspection of B-factors shows very low confidence for the loop consisting of residue 285–288 in *T. brucei* sequence. It is the sequence that corresponds to the missing residues 294–296 of the crystal structure of *L. mexicana*. The rms *z*-score for bond angles and bond lengths of a monomer are equal to 1.140 and 0.669, respectively (*z*-score is a transformation of raw scores into a standard form and rms *z*-score is used as a measure of a drift from the expected behavior in a given set of values). There are also 37 residues with abnormal bond angles. Abnormal packing environment for at least three sequential residues with a questionable packing environment was found. This indicates that these residues are part of a strange loop. Nine residues with forbidden *phi*/*psi* combinations are also present in the monomeric homology model. The quality of the homodimer homology model is as follows. The rms *z*-score for bond angles is equal (1.037), which is close to the data obtainable for high resolution of x-ray structures. There are four residues with abnormal torsion angles; these are Ile168(A), Pro312(A), Ile168(B), Pro313(B), and Gly15(B) (A and B stands for the subunits of the homodimeric structure). The rms *z*-score for bond length is 0.622. There are also 20 residues with disallowed *phi*/*psi* combinations. There is a large number of pairs with abnormally short interactomic distances. A large B-factor of 50 is observed for buried atoms (the average value for a room temperature x-ray study lies between 10 and 20). The analysis of these results indicates good quality of the homologous models. However, the further refining of the model is desirable. The tertiary structure of *Tb*GPDH consists of the two well defined domains, the N-terminal and the C-terminal domain. The DSSP (Kabsch and Sander, 1983) secondary structure analysis shows that the N-terminal NAD-binding domain consists of residues 1 to 189, which are arranged into eight β-sheets and seven α-helices.

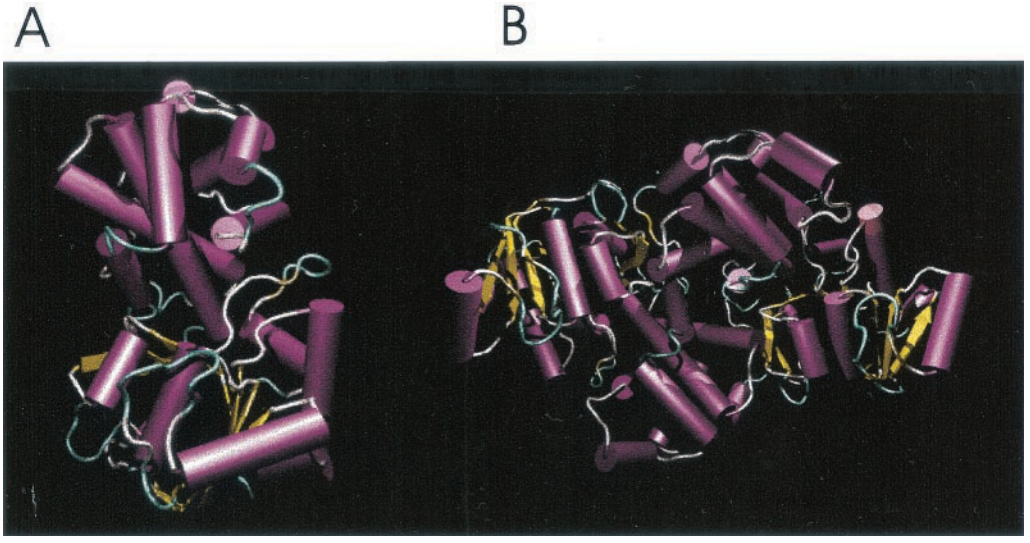

FIGURE 2 The tertiary structure of the *Tb*GPDH monomer (*A*) and homodimer (*B*). The α-helices are shown as violet cylinders, β-sheets as yellow arrows, the turns are colored blue, and the random coil structures in white. Color and the secondary structure assignment are based on the output from the DSSP program (Kabsch and Sander, 1983).

The β-sheet 7 consists of residues 159–165 and is antiparallel to the β-sheet 6 (residues 141–145). The β-sheet number 8 is parallel to the β-sheet seven. All β-sheets are flanked by the two α-helical structures, creating long flanking helices (helix 4 has a "kink" at Ser99 residue) and a number of shorter helices at the bottom of the structure. The β-sheets 7 and 8 are solvent exposed. The first β-α unit contains the highly conserved GxGxxG NAD-binding motif (residues 15–20). The N-terminal and the C-terminal domains are connected by the short loop comprising three residues 191 to 193. The C-terminal substrate-binding domain consists of residues 193 to 354. The residues 194 to 336 are organized into eight α-helical structures. Helices 1 and 2 are particularly long and consist of 23 and 21 residues, respectively. They are antiparallel with an angle of 151.95°. The shortest distance between the straight lines connecting the endpoints of helices is equal to 0.585 nm. Residues 336 to 349 adopt a random coil structure. The subunit interactions are of hydrophobic origin with small contribution of hydrogen bonding (Fig. 3, *A* and *B*; Fig. 10). The visual inspection of the interaction surfaces between the N- and the C-terminal domain show that the hydrophobic surface of the

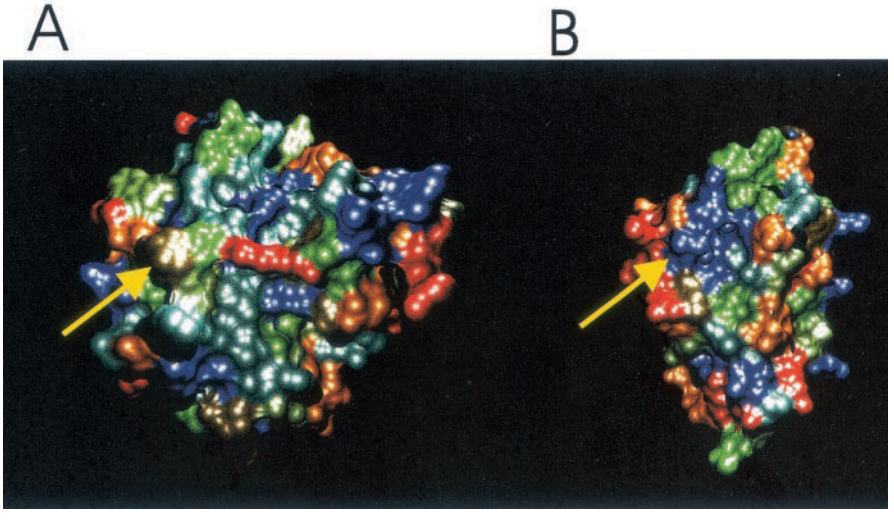

FIGURE 3 (*A*) The surface of the N-terminal domain is colored according to the hydrophobicity scale (Kyte and Doolittle, 1982); the arrow shows a contact area between the N- and C-terminal domains. (*B*) The surface of the C-terminal domain is colored according to the hydrophobicity scale (Kyte and Doolittle, 1982); the arrow shows a contact area between the N- and C-terminal domains. Color code: blue (hydrophobic) <> red (hydrophilic).

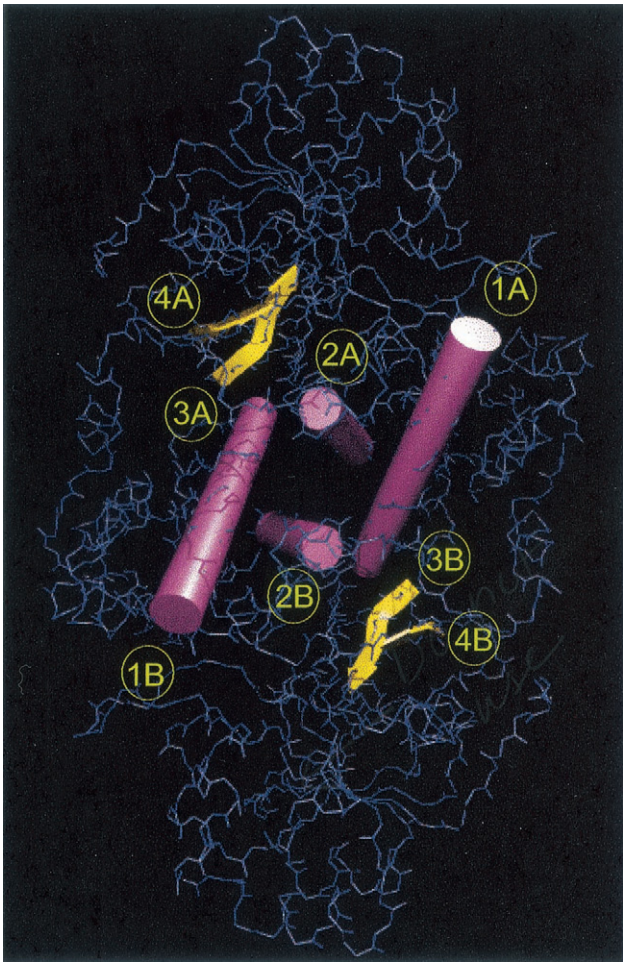

FIGURE 4 The "zipping" motif of the homodimeric structure. 1*A*, B helix 2 of the subunit A and B; 2*A*, B helix 3 of subunit A and B of the C-terminal domain; 3*A*, B β-sheets 7 of subunit A and B; and 4*A*, B β-sheet 8 of subunit A and B of the N-terminal domain.

N-terminal domain is much larger than the corresponding surface of the C-terminal domain (Fig. 3, *A* and *B*). The analysis of the number of hydrogen bonds and the quality of the interacting surfaces between the domains of the monodimer indicates that domains are bound together by means of hydrophobic interactions within the monomeric structure. The interactions between subunits of the homodimeric structure are also of hydrophobic origin. There is, however, a specific motif "zipping" together the subunits. It consists of the helices 2 and 3 of the C-terminal domain (Fig. 4). This zipping structure is shielded from the solvent environment by the two β-sheets at each side, sheets 7 and 8 of the N-terminal domain. The analysis of the accessibility of NAD-binding site and the putative substrate binding site shows that the accessibility is not affected by the aggregation process and creation of a homodimeric structure.

NAD-binding site

The *Lm*GPDH crystal structure in complex with NADH (Suresh et al., 2000) allowed for identification of the residues directly involved in NAD-binding. Within the *Tb*GPDH sequence these residues, Glu293, Lys118, Ser16, and Met39, are fully conserved. The NAD-binding domain consists of four residues. Three residues belong to the N-terminal domain and the one residue, the Glu293, is the first residue of the sixth helix of the C-terminal domain (α13, Fig. 1). The striking feature is that most of the residues involved in NAD-binding are located within the well defined secondary structures. Only Lys118 is located in the middle of the disordered loop connecting helices 5 and 6 of the N-terminal domain. The net of distances and their standard deviations during the simulation that characterize the NAD-binding site of the *Tb*GPDH are summarized in Table 1.

Structural fluctuations

Time-dependent Cα rms deviation (RMSD) has been used to provide a picture of the global drift of the homology model during the simulation period. Thus, during the simulation, the RMSD drift of Cα atoms (for both the monomer and homodimer structures) from the initial protein structure was determined (Fig. 5). The drift observed for the monomeric structure reaches a plateau after ~0.35 ns of simulation and is equal to ~0.4 nm. A similar behavior can also be observed for the homodimeric structure. The initial rise of

Table 1. The net of distances and their standard deviations during the simulation that characterize the NAD-binding site of the *Tb*GPDH

Residues	Homologous monomer (nm)	Monomer NPT-simulation (nm)	Homodimer chain A NPT-simulation (nm)	Homodimer chain B NPT-simulation (nm)
E293OE1-K118N	0.842	0.660 ± 0.093	0.817 ± 0.108	0.842 ± 0.078
E293OE1-S16OG	1.433	1.104 ± 0.119	1.316 ± 0.215	1.353 ± 0.106
K118N-S16OG	1.425	1.3106 ± 0.082	1.322 ± 0.175	1.270 ± 0.101
S16OG-M39N	0.494	0.5028 ± 0.034	0.526 ± 0.058	0.797 ± 0.112
E293E1-M39N	1.759	1.42648 ± 0.146	1.629 ± 0.079	1.96 ± 0.131
K118N-M39N	1.731	1.49967 ± 0.160	1.626 ± 0.069	1.580 ± 0.144

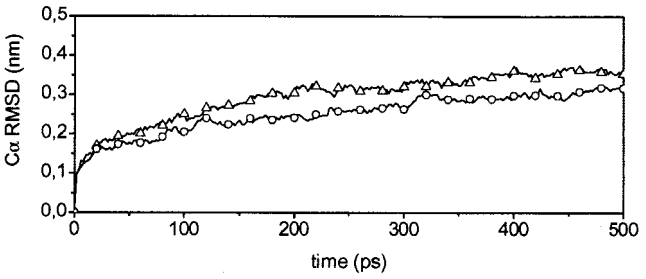

FIGURE 5 Drift of protein structure from the initial model. The RMSD of all Cα atoms from the starting structure is shown as a function of time. (–△—monomer, –○– homodimer)

the Cα RMSD over the first 300 ps is possibly attributable to the relaxation motion of the protein or inaccuracies in the force field. To examine the fluctuation of the structure on a residue-by-residue basis, the time-averaged (last 200 ps of simulation) rms fluctuation (RMSF) of Cα atoms (Fig. 6) has been analyzed. The averaging time period, 300 ps to 500 ps, was determined based on the observation of the Cα RMSD drift. The analysis of Fig. 6, *A* and *B* shows that RMSFs for a monomeric structure adopt the largest values at the C terminus where RMSF is equal ~0.5 nm. Such large fluctuation value observed for the C terminus is attributable to the presence of unstructured free end consisting of residues 335 to 350. The central region consisting of a set of β-sheets and α-helices undergoes rather small fluctuations with the maximum in the order of ~0.15 nm. This stable behavior of the central region is attributable to the network of hydrogen bonds stabilizing these secondary structures. The RMSF distribution observed for homodimeric structure has a much lower (~0.18 nm) and

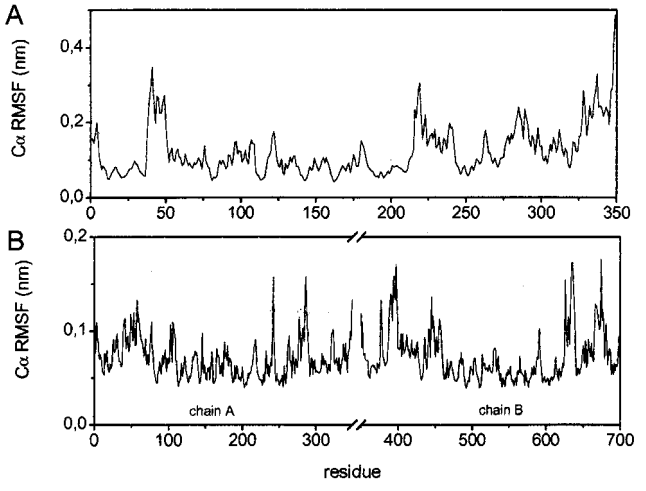

FIGURE 6 Fluctuation of protein coordinates. The RMSFs of the Cα coordinates from their time-averaged values (last 200-ps simulation) are shown as a function of residue number for the monomeric (*A*) and homodimeric (*B*) structures.

more uniformly distributed fluctuation than those observed for the monomeric structure. The more stable behavior of the homodimeric structure can be explained by the nature of the interactions between the units of the homodimer and by the wrapping of the C-terminal ends of each monomer around one another. Additional information on the structural flexibility is offered by the analysis of time-dependent secondary structure fluctuations (Fig. 7). Analysis of Fig. 7 reveals very high stability of the motifs with the well defined secondary structure. Thus, β-sheets and α-helices observed within the *Tb*GPDH structure are very stable during the whole simulation period, whereas turns become bends, and vice versa. The lowest time-dependent stability and lowest structure conservation is observed for the C-terminal end. This observation is in agreement with the results of RMSF analysis. The comparison of the stability patterns for the secondary structures within the monomeric and the homodimeric structure indicates slightly greater stability for a dimeric structure. This observation is in agreement with a biologically active structure of a GPDH enzyme.

Distance fluctuations of NAD-binding residues

The analysis of the time-dependent fluctuation of the distances between the atoms directly involved in creation of NAD-binding site, for both monomer and homodimer, shows that the distances are fairly stable during the simulation time (Table 1). The analysis of Table 1 and the comparison with the results of the x-ray studies on *Lm*GPDH shows that distance differences are within the range of ~0.1 nm. Furthermore, one can observe that the distances derived for the homodimeric structures are slightly shorter than those observed for monomeric structures. This tighter packing could be explained by increased stability of the homodimeric structure versus the monomeric structure. The results obtained also show that the shape of the NAD-binding site stays practically constant during the simulation time.

GxGxxG NAD-binding motif.

All NAD-dependent GPDHs studied to date have the highly conserved motif, GxGxxG (Wierenga et al., 1986). This motif is also present within the sequence of *Tb*GPDH and comprises the following amino acids GSGAFG (residue 15–20, Fig. 1). It has also been shown that the core topologies of the classical nicotinamide-binding proteins overlap very well. The RMSD of backbone atoms superposition ranges from 0.07 to 0.47 nm (Bellamacina, 1996). To estimate the time-dependent stability of the GxGxxG motif, the RMSD fluctuation of the six residues building the NAD-binding motif was analyzed. The result is presented in Fig. 8, *A* and *B*. The mean fluctuation versus simulation time is

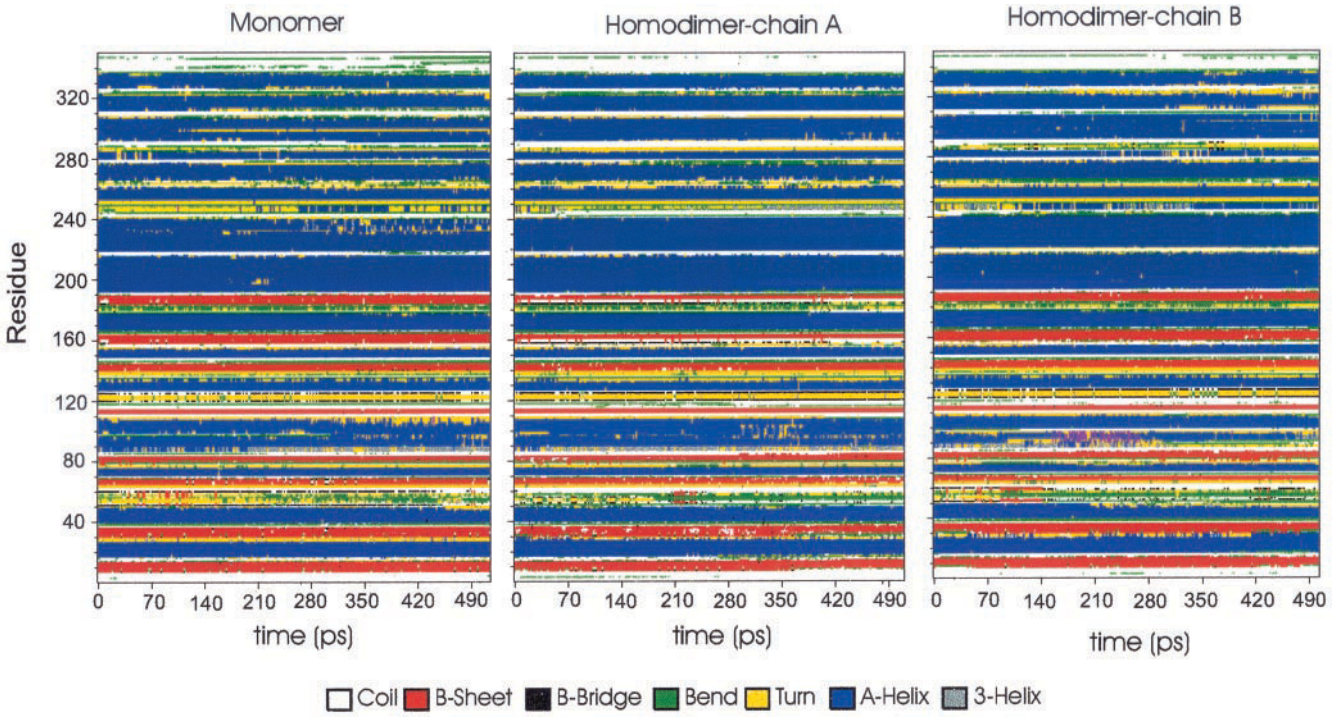

FIGURE 7 Secondary structure fluctuations (analyzed using the DSSP program (Kabsch and Sander, 1983)) as a function of time for the simulation of the GPDH enzyme.

in the order of 0.066 nm, with the standard deviation of ± 0.015 nm and 0.06 ± 0.014, 0.115 ± 0.015 for monomeric and homodimeric chains A and B, respectively. These low values of fluctuations indicate that the structure of this functionally important unit is very stable. However, a slight increase in flexibility is observed for the GxGxxG motif residing within the chain B of the homodimeric structure.

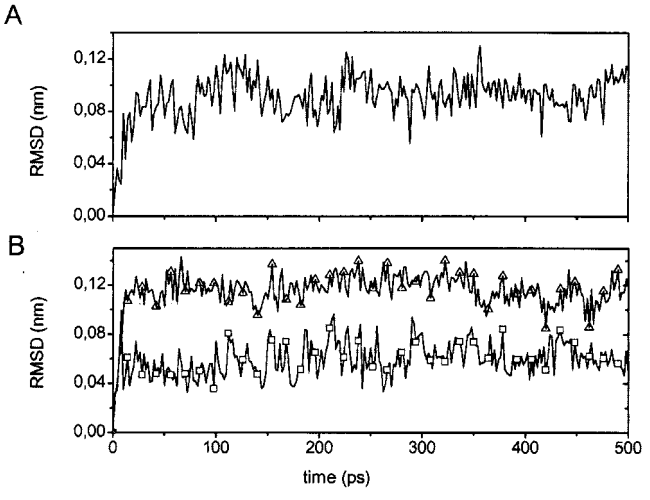

FIGURE 8 GxGxxG pattern Cα-atom RMSD fluctuation as a function of time. (*A*) Monomeric structure; (*B*) Homodimeric structure (–○– chain A, –△– chain B).

The overall structural stability could be attributable to the very stable secondary structures created by residues 9–14, a β-structure and 21–28, an α-helix.

The x-ray analysis of the *Lm*GPDH in complex with NADH allowed delineation of a putative active site of this enzyme. The residues Lys125, Lys210, Asp263, and Thr267 were designated to be involved in substrate binding or catalysis. The analysis of the sequence alignment (Fig. 1) shows that these residues are fully conserved within the *T. brucei* sequence. Thus, the analysis of the active site of *Lm*GPDH can be extended on the active site of *Tb*GPDH. Additionally, the structure-structure similarity search using VAST (The National Center for Biotechnology Information) service allowed for the mapping of the sequence conservation onto the secondary structure of the protein (Fig. 9). The analysis of the color map indicates that by approaching the core of the protein, the higher structural and sequential conservation is observed. Additionally a structural similarity was discovered between the crystal structure of the *N*-(1-D-carboxylethyl)-L-norvaline dehydrogenase from *Arthrobacter* species (PDB code: 1BG6).

Interdomain interactions and structure stability

The *Tb*GPDH molecule consists of two well defined domains. The N-terminal domain consists of residues 1–189 and the C-terminal domain of 192–349. These two domains

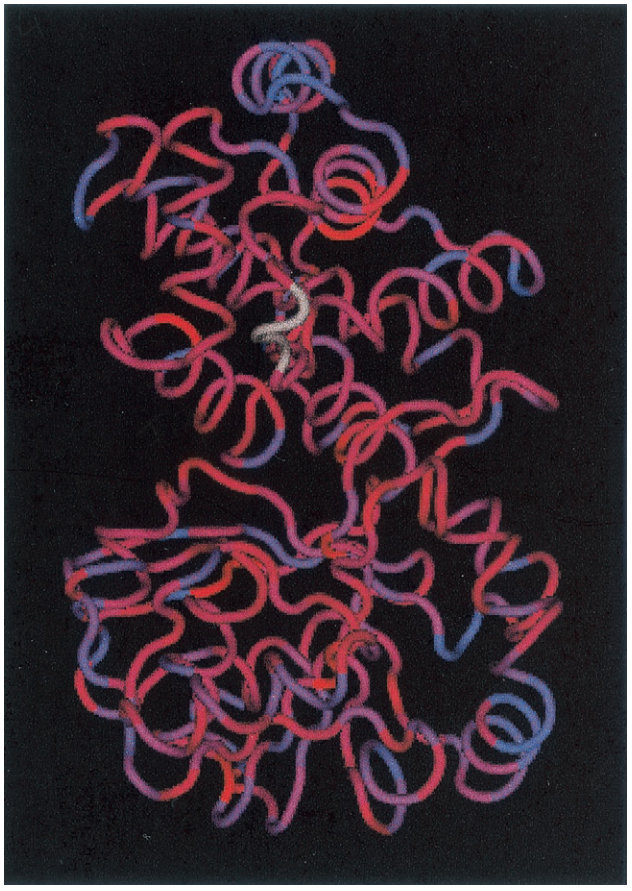

FIGURE 9 Sequence conservation between *Lm*GPDH and *Tb*GPDH mapped onto the secondary structure of the protein. The color code is scaled from red to blue depending on the number of different residues in the column. The variety is weighted by a Block Substitution Matrix (BLOSUM62).

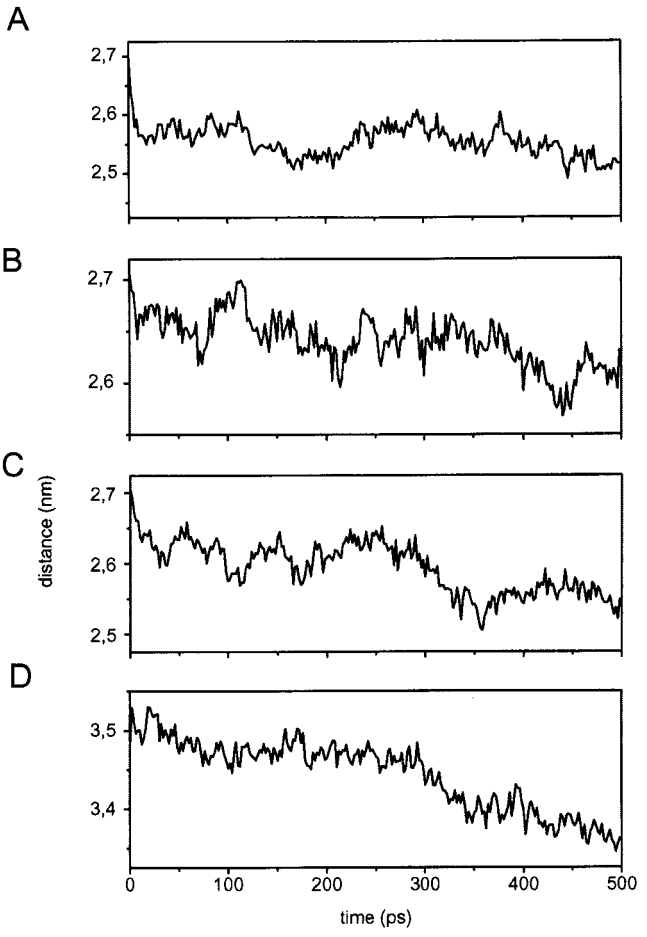

FIGURE 10 The distances of the centers of mass of the N- and the C-terminal domain versus simulation time for the monomeric structure (*A*) and the homodimeric structure (*B*, *C*) and between the subunits of the homodimeric structure (*D*).

are linked by a short three-residue loop. The analysis of the fluctuation of the centers of mass of these two domains as a function of time (Fig. 10, *A–D*) and the calculation of the changes of the number of hydrogen bonds (Fig. 11, *A–D*) between these two domains reveals the forces stabilizing the molecule and the overall shape of the molecule. The analysis of the distances between the centers of mass between these two domains shows that these distances are in the range of 0.255–0.344 nm with the range of fluctuations between 0.013 and 0.020 nm. It is noticeable that distances between the centers of mass of domains as well as subunits decrease during the simulation time. However, these fluctuations are effectively negligible when the size of the protein is taken into account. The analysis of the hydrogen bonding profile shows that after 300 ps of the simulation, the number of hydrogen bonds between domains differs for the monomeric and homodimeric structures. Thus, one can observe that the average number of hydrogen bonds after 300 ps of simulation for monomeric structure is seven and eight for the chain A and B of homodimer. There is, however, a significant increase in the number of hydrogen-bonds (four to eight hydrogen bonds increase after 300 ps of simulation) within the unit B of the homodimeric structure. The number of hydrogen bondings between subunits of the homodimeric structure is slightly larger and is equal to 15. The total number of hydrogen bonds observed indicates that hydrophobic forces are responsible for interactions between domains and between subunits. The comparison of Figs. 10 and 11 does not reveal any specific correlation between the distances of centers of mass and the number of hydrogen bonds. To validate the overall shape stability, the RMSD fluctuations of all atoms within the subdomains of both the monomeric structure and the subunits of the homodimeric structure were analyzed (Fig. 12). The analysis of Fig. 12 shows that the RMSD of all atoms for the N-terminal domain of the monomeric structures oscillates ~0.2 nm and for the C-terminal ~0.4 nm. Similar analyses performed on the homodimeric structures show that the RMSD of subunit A oscillates ~0.27 nm and for subunit B ~0.32 nm. Con-

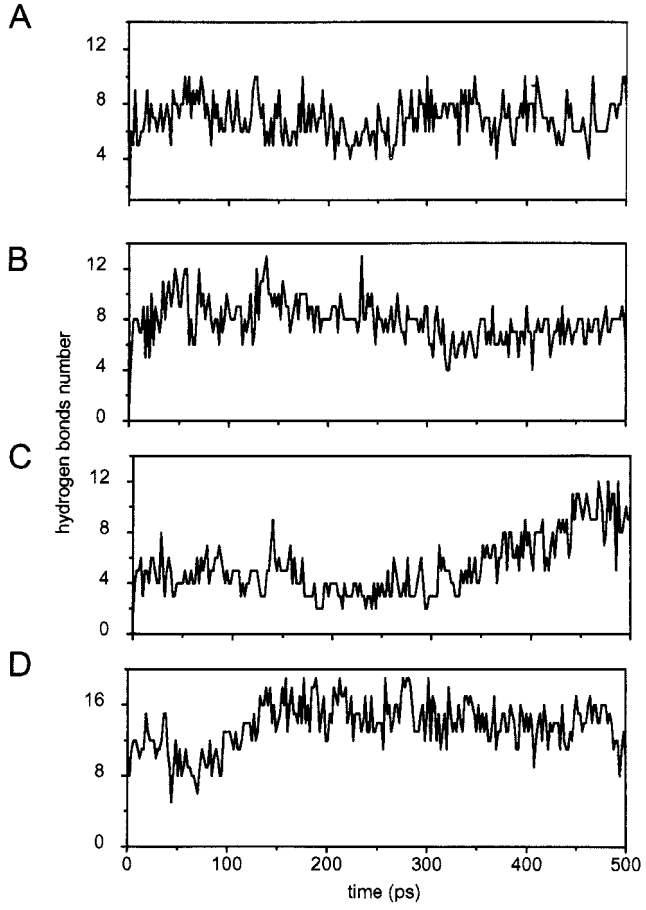

FIGURE 11 Hydrogen bonding number versus simulation time (*A*) between the N- and C-terminal domains of a monomer; (*B*) N- and C-terminal domains of the homodimer chain A; (*C*) N- and C-terminal domains of the homodimer chain B; and (*D*) between chain A and B of the homodimer.

sidering the size of the protein, it is possible to state that the overall shape of the molecule is remarkably stable and after 300 ps of simulation does not undergo any significant changes.

CONCLUSION

The *Tb*GPDH homology model built on the template of the *Lm*GPDH enzyme may be used as a valid target molecule for designing antitrypanosomal drugs. The MD simulation data analysis reveals that the *Tb*GPDH enzyme consists of the two very stable domains connected by a short three-residue loop. The interdomain interactions are based on hydrophobic interactions with some assistance from hydrogen bonding. The analysis also shows that the main structural features of the NAD-binding domain are very stable. The comparison of the dynamic trajectories of the monomeric and homodimeric structures shows insignificant differences, indicating that the solvent-exposed large hydrophobic area responsible for monomer-monomer interactions

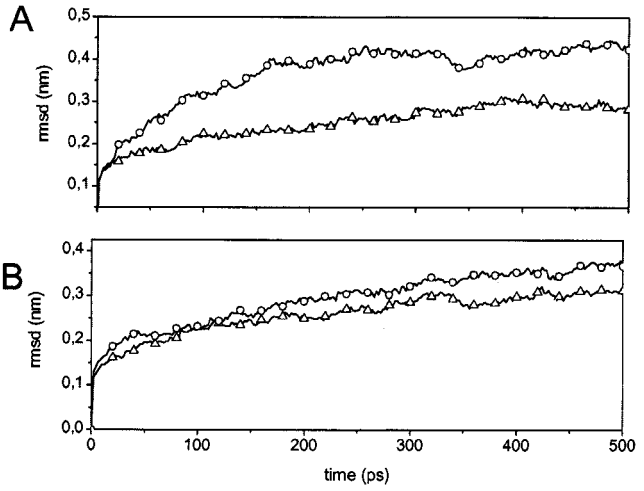

FIGURE 12 The RMSD fluctuations of all atoms within the domains and subunits of (*A*) the monomer (–△– N-terminal domain, –○– C-terminal domain); (*B*) the homodimer (–△– chain A, –○— chain B).

does not undergo any structural changes during the MD simulation of a monomeric structure. This behavior is most probably owed to hydrogen bonding stabilizing the secondary structure motifs (helices, β-sheets) within these regions. The energy of these bonds is high enough to compensate for the unfavorable solvent-hydrophobic surface interactions. Considering the stability differences between monomer and homodimer, the monomeric structure may be used as a template for drug-designing studies. Thus, selective inhibitors that have a similar shape/size to the NAD molecule can be designed. A similar approach has been used in the designing of the glyceraldehydes-3-phosphate dehydrogenase inhibitors (Aronov et al., 1999, 1998; Bressi et al., 2000; Kennedy et al., 2001). The inhibitor designing should be based on the exploitation of the structural differences between the human and the trypanosomal β-α-β motif encompassing the GxGxxG sequence. A very good RMS fitting between the *Lm*GPDH and *Tb*GPDH homology model provides the unique chance to exploit the information concerning the NAD binding as a template for the future docking experiment using NAD analogs.

The author thanks Prof. H. Klump for his comments to the manuscript and acknowledge the financial support from the Joint Research Committee, Rhodes University, South Africa.

REFERENCES

Anker, M., and D. Schaaf. 2000. WHO Report on Global Surveillance of Epidemic-Prone Infectious Diseases: African Trypanosomiasis. World Health Organization, Geneva, Switzerland.

Aronov, A. M., S. Suresh, F. S. Buckner, W. C. Van Voorhis, C. L. Verlinde, F. R. Opperdoes, W. G. Hol, and M. H. Gelb. 1999. Structure-based design of submicromolar, biologically active inhibitors of trypanosomatid glyceraldehyde-3-phosphate dehydrogenase. *Proc. Natl. Acad. Sci. U.S.A.* 96:4273–4278.

Aronov, A. M., C. L. Verlinde, W. G. Hol, and M. H. Gelb. 1998. Selective tight binding inhibitors of trypanosomal glyceraldehyde-3-phosphate dehydrogenase via structure-based drug design. *J. Med. Chem.* 41: 4790–4799.

Bairoch, A., and R. Apweiler. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* 28:45–48.

Barrett, M. P. 1999. The fall and rise of sleeping sickness. *Lancet.* 353: 1113–1114.

Barrett, M. P., J. C. Mottram, and G. H. Coombs. 1999. Recent advances in identifying and validating drug targets in trypanosomes and leishmanias. *Trends Microbiol.* 7:82–88.

Bellamacina, C. R. 1996. The nicotinamide dinucleotide binding motif: a comparison of nucleotide binding proteins. *FASEB J.* 10:1257–1269.

Berendsen, H. J., J. P. Postma, A. DiNola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81: 3684–3690.

Berendsen, H. J., J. P. Postma, W. F. van Gunsteren, and J. Hermans. 1981. Intramolecular Forces. D. Reidel Publishing Company, Dordrecht. 331–342.

Berendsen, H. J., D. van der Spoel, and R. van Drunen. 1995. GROMACS: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.* 91:43–56.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.

Bernstein, B. E., P. A. Michels, and W. G. Hol. 1997. Synergistic effects of substrate-induced conformational changes in phosphoglycerate kinase activation. *Nature.* 385:275–278.

Bernstein, B. E., D. M. Williams, J. C. Bressi, P. Kuhn, M. H. Gelb, G. M. Blackburn, and W. G. Hol. 1998. A bisubstrate analog induces unexpected conformational changes in phosphoglycerate kinase from *Trypanosoma brucei*. *J. Mol. Biol.* 279:1137–1148.

Bressi, J. C., J. Choe, M. T. Hough, F. S. Buckner, W. C. Van Voorhis, C. L. Verlinde, W. G. Hol, and M. H. Gelb. 2000. Adenosine analogues as inhibitors of *Trypanosoma brucei* phosphoglycerater kinase: elucidation of a novel binding mode for a 2-amino-N(6)-substituted adenosine. *J. Med. Chem.* 43:4135–4150.

Chudzik, D. M., P. A. Michels, S. de Walque, and W. G. Hol. 2000. Structures of type 2 peroxisomal targeting signals in two trypanosomatid aldolases. *J. Mol. Biol.* 300:697–707.

Denise, H., C. Giroud, M. P. Barrett, and T. Baltz. 1999. Affinity chromatography using trypanocidal arsenical drugs identifies a specific interaction between glycerol-3-phosphate dehydrogenase from *Trypanosoma brucei* and cymelarsan. *Eur. J. Biochem.* 259:339–346.

Guex, N., A. Diemand, and M. C. Peitsch. 1999. Protein modelling for all. *Trends Biochem. Sci.* 24:364–367.

Guex, N., and M. C. Peitsch. 1997. SWISS-MODEL and Swiss-Pdb Viewer: an environment for comparative protein modelling. *Electrophoresis.* 18:2714–2723.

Henikoff, S., and J. G. Henikoff. 1992. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U.S.A.* 89:10915–10919.

Hooft, R. W. W., G. Vriend, C. Sander, and C. C. Abola. 1996. Errors in protein structures. *Nature.* 381:272–272.

Huang, X., and W. Miller. 1991. A time-efficient, linear-space local similarity algorithm. *Adv. Appl. Math.* 12:337–357.

Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38.

Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 22:2577–2637.

Kennedy, K. J., J. C. Bressi, and M. H. Gelb. 2001. A disubstituted NAD^+ analogue is a nanomolar inhibitor of trypanosomal glyceraldehyde-3-phosphate dehydrogenase. *Bioorg. Med. Chem. Lett.* 11:95–98.

Kim, H., I. K. Feil, C. L. Verlinde, P. H. Petra, and W. G. Hol. 1995. Crystal structure of glycosomal glyceraldehyde-3-phosphate dehydrogenase from *Leishmania mexicana*: implications for structure-based drug design and a new position for the inorganic phosphate binding site. *Biochemistry.* 34:14975–14986.

Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157:105–132.

Lindahl, E., B. Hess, and D. van der Spoel. 2001. Gromacs 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Med.* 7:306–317.

Lo, T. W., M. E. Westwood, A. C. McLellan, T. Selwood, and P. J. Thornalley. 1994. Binding and modification of proteins by methylglyoxal under physiological conditions. A kinetic and mechanistic study with N α-acetylarginine, N α-acetylcysteine, and N α-acetyllysine, and bovine serum albumin. *J. Biol. Chem.* 269:32299–32305.

Moore, A., M. Richer, M. Enrile, E. Losio, J. Roberts, and D. Levy. 1999. Resurgence of sleeping sickness in Tambura County, Sudan. *Am. J. Trop. Med. Hyg.* 61:315–318.

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

Rodriguez, R., G. Chinea, N. Lopez, T. Pons, and G. Vriend. 1998. Homology modeling, model and software evaluation: three related resources. *Bioinformatics.* 14:523–528.

Ryckaert, J. P., G. Ciccotti, and H. J. Berendsen. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Phys.* 23: 327–341.

Schwede, T., A. Diemand, N. Guex, and M. C. Peitsch. 2000. Protein structure computing in the genomic era. *Res. Microbiol.* 15:107–112.

Suresh, S., S. Turley, F. R. Opperdoes, P. A. Michels, and W. G. Hol. 2000. A potential target enzyme for trypanocidal drugs revealed by the crystal structure of NAD-dependent glycerol-3-phosphate dehydrogenase from *Leishmania mexicana*. *Structure.* 8:541–552.

van der Spoel, D., A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. Sijbers, B. Hess, K. A. Feenstra, E. Lindahl, R. van Drunen, and H. J. C. Berendsen. 2001. Gromacs User Manual, Version 3.0. Nijenborgh, AG Groningen, The Netherlands. Internet: www.gromacs.org.

van der Spoel, D., A. R. van Buuren, D. P. Tieleman, H. J. Berendsen. 1996. Molecular dynamics simulations of peptides from BPTI: a closer look at amide-aromatic interactions. *J. Biomol. NMR.* 8:229–238.

van Gunsteren, W. F., and H. J. Berendsen. 1987. Gromos-87 manual. Biomos BV, Nijenborgh, Groningen, The Netherlands.

VAST: Vector alignment search tool. The National Center for Biotechnology Information.

Vellieux, F. M., J. Hajdu, C. L. Verlinde, H. Groendijk, R. J. Read, T. J. Greenhough, J. W. Campbell, K. H. Kalk, J. A. Littlechild, H. C. Watson, and W. G. Hol. 1993. Structure of glycosomal glyceraldehyde-3-phosphate dehydrogenase from *Trypanosoma brucei* determined from Laue data. *Proc. Natl. Acad. Sci. U.S.A.* 90:2355–2359.

Wierenga, R. K., M. E. Noble, G. Vriend, S. Nauche, and W. G. Hol. 1991. Refined 1.83 A structure of trypanosomal triosephosphate isomerase crystallized in the presence of 2.4 M-ammonium sulphate. A comparison with the structure of the trypanosomal triosephosphate isomerase-glycerol-3-phosphate complex. *J. Mol. Biol.* 220:995–1015.

Wierenga, R. K., P. Terpstra, and W. G. Hol. 1986. Prediction of the occurrence of the ADP-binding $\beta\alpha$ β-fold in proteins, using an amino acid sequence fingerprint. *J. Mol. Biol.* 187:101–107.