Genomics 103 (2014) 107–113

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

3D molecular modeling and evolutionary study of the *Trypanosoma brucei* DNA Topoisomerase IB, as a new emerging pharmacological target

Dimitrios Vlachakis [a], Athanasia Pavlopoulou [a], Maria G. Roubelakis [b,c], Christos Feidakis [a], Nikolaos P. Anagnou [b,c], Sophia Kossida [a,*]

[a] Bioinformatics & Medical Informatics Team, Biomedical Research Foundation, Academy of Athens, Soranou Efessiou 4, Athens 11527, Greece
[b] Laboratory of Biology, University of Athens School of Medicine, Greece
[c] Cell and Gene Therapy Laboratory, Biomedical Research Foundation, Academy of Athens, Soranou Efessiou 4, Athens 11527, Greece

ARTICLE INFO

Article history:
Received 4 June 2013
Accepted 29 November 2013
Available online 5 December 2013

Keywords:
3D molecular modeling
Evolutionary study
Trypanosoma brucei
DNA Topoisomerase IB

ABSTRACT

In the present study, an outline is proposed that may lead to specific drug design targeting of the *Trypanosoma brucei* DNA Topoisomerase IB. In this direction, an unequivocally specific platform was designed for the development of selective modulators. The designed platform is focused on the unique structural and catalytic features of the enzyme. Extensive phylogenetic analysis based on all available published genomes indicated a broad distribution of DNA topoisomerases across eukaryotic species and revealed structurally important amino acids which could be assigned as potentially strong contributors to the regulation of the mechanism of the *T. brucei* DNA Topoisomerase IB. Based on the above, we propose a comprehensive *in silico* 3D model for the structure of the *T. brucei* DNA Topoisomerase IB. Our approach provides an efficient intergraded platform with both evolutionary and structural insights for the rational design of pharmacophore models as well as novel modulators as the anti-*T. brucei* DNA Topoisomerase IB agents with therapeutic potential.

1. Introduction

DNA topoisomerases are essential enzymes that control DNA supercoiling and relieve superhelical tension during transcription, replication and chromatin assembly by transient DNA strand cleavage and re-ligation. DNA topoisomerases are broadly divided into two major families based on the number of DNA strands they cleave, type I (single-strand) and type II (double-strand), respectively. Type I enzymes, which introduce a single-stand break, are divided further into the TopIA and TopIB classes on the basis of their protein architecture (monomer versus dimer), DNA substrate specificity (single-strand versus duplex) and catalytic mechanism. The TopIB enzymes cleave single-strand DNA and relax both positive and negative supercoiled DNA [1–4].

TopIB members have been identified in eukaryotes, prokaryotes and viruses [5–10]. TopIB are monomeric enzymes, consisting of an amino-terminal DNA binding domain, exhibiting variability in both sequence and length, and possessing a highly conserved carboxy-terminal core catalytic domain. This domain harbors the invariant tyrosine residue which attacks the phosphodiester bond of the DNA strand forming a transient 3′ phosphotyrosyl-enzyme intermediate and nicked DNA [11–13]. However, in kinetoplastids, which are flagellated protozoa like *Trypanosoma brucei* and *Leishmania donovani*, the TopIB enzyme exists as a dimer, the subunits of which are encoded by two different genes located at two different chromosomes [14,15]. In the present study, computational methods were employed to model the three-dimensional structure of *T. brucei* TopIB dimeric protein (henceforth referred to as TbTOP1B) and identify potential drug-like molecules as targets of this protein. Towards this direction, a comprehensive phylogenetic analysis was initially performed; conserved amino acid residues important for the structure and function of the TOP1B class of enzymes, as well as residues conserved between the kinetoplastid TOP1B proteins, were identified. A 3D pharmacophore model based on the structural and physicochemical properties of the TOP1B binding site was generated. This pharmacophore model was subsequently used for mining the chemical databases in search for novel small drug-like compounds (Fig. S1).

2. Material and methods

2.1. Phylogenetic analyses

2.1.1. Sequence database search

In order to identify homologous TOP1B protein sequences, the non-redundant publicly available databases: UniProtKB [16], GenBank [17] and *Cyanidioschyzon merolae* database [18] were searched with the entire amino acid sequences of all known TOP1Bs applying reciprocal BLASTp and tBLASTn [19].

* Corresponding author at: Bioinformatics & Medical Informatics Team, Biomedical Research Foundation, Academy of Athens, Soranou Efessiou 4, Athens 11527, Greece. Fax: +30 210 6597 545.
E-mail address: skossida@bioacademy.gr (S. Kossida).

0888-7543/$ – see front matter © 2013 Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.ygeno.2013.11.008

2.1.2. Phylogenetic tree construction

The entire TOP1B amino acid sequences were searched against PROSITE [20], in order to identify protein domains. The retrieved protein sequences were aligned using CLUSTALW [21]. To optimize the alignment, the amino acid sequences that correspond to the large and small subunit of the kinetoplastid TOP1B were joined and treated as a single sequence in the phylogenetic analysis. The resulting multiple sequence alignment was trimmed by applying Gblocks [22,23] with default options, and subsequently was used to reconstruct a phylogenetic tree by employing the neighbor-joining method implemented in MEGA5 [24]. The number of amino acid substitutions per site was estimated using the JTT model [25]. The robustness of the inferred tree was tested by performing bootstrap analysis (2000 replicates). The phylogenetic tree was visualized using Dendroscope [26].

2.1.3. Motif construction

The full-length amino acid sequences of TOP1Bs under study were aligned and edited by employing Utopia suite's CINEMA alignment editor [27]. The evolutionarily conserved sequence motifs that were derived from the alignment were submitted to Weblogo [28] in order to generate consensus sequences.

2.2. Homology modeling

The homology modeling of the TbTOP1B dimeric protein was carried out using a restraint-based approach, as implemented in the Modeller package [29]. The crystal structure of the heterodimeric *L. donovani* topoisomerase I-vanadate-DNA complex (PDB entry: 2B9S) [30] was used as template structure. The sequence alignment between the target sequence of TbTOP1B and the template sequence revealed almost 69% identity, which allowed for a reliable homology modeling to be performed. The homology modeling process in Modeller starts with the building of a profile using the "build_profile.py" script. Then the alignment was performed using the "align2d" command. As soon as the alignment between the target and the template was constructed, the 3D model of the target was calculated using the automodel class of Modeller. Finally, the model was evaluated using the DOPE approach via the "evaluate_model.py" file. The overall homology modeling process was divided into the following steps: First, the initial spatial constraints for the target sequence were derived from a large number of template protein structures; the target sequence was aligned to the backbone of a template structure copying the geometric coordinates of the template to the target sequence. Second, target regions, where geometric constrains could not be copied from the template easily, were modeled. These regions represented either deletions or insertions with respect to the template. The third step involved loop selection and side chain modeling, where a collection of independent models was obtained. Fourth, the final models were scored and ranked, after they had been stereochemically tested and evaluated with a built-in module for protein geometry error-checking. Since several models were calculated for the same target, we had to choose an evaluation method that would rank them. As mentioned above the DOPE approach was chosen. It is important to clarify that the DOPE score is not an absolute measure rather than a relative one. It has only meaning during the relatively comparison of models that have been generated based on the same initial alignment. DOPE method is considered more advanced to its rival GA341 method that is also available by Modeller.

2.2.1. Molecular electrostatic potential (MEP)

Electrostatic potential surfaces were calculated by solving the non-linear Poisson–Boltzmann equation using finite difference method as implemented in the PyMOL Software [31]. The potential was calculated on grid points per side (65, 65, 65) and the grid fill by solute parameter was set to 80%. The dielectric constants of the solvent and the solute were set to 80.0 and 2.0, respectively. An ionic exclusion radius of 2.0 Å, a solvent radius of 1.4 Å and a solvent ionic strength of 0.145 M were applied. Amber99 [32] charges and atomic radii were used for this calculation.

2.2.2. Model optimization

The models were initially subjected to energy minimization using the Gromacs-implemented force-fields to remove the geometrical strain [33,34]. The model was subsequently solvated with Simple Point Charge (SPC) water using the truncated octahedron box extending to 7 Å from the model. Molecular dynamics was performed after that at 300 K, 1 atm with 2 fsecond step size, and for a total of 10 ns, using the NVT ensemble in a canonical environment; NVT stands for Number of atoms, Volume and Temperature that remain constant throughout the simulation. The results of the molecular dynamics simulation were collected into a database by Gromacs which can be subjected to further analysis.

2.2.3. Model evaluation

The resultant models were initially evaluated within the Gromacs package, version 4.5.5 [33,34] by a residue packing quality function, which depends on the number of buried non-polar side chain groups and on hydrogen bonding. Moreover, the suite PROCHECK [35] was employed to further evaluate the quality of the produced TbTOP1B model. Finally, Verify3D [36] was used to evaluate whether the model of TbTOP1B is similar to any known protein structures. The resultant homology models were visualized using PyMOL [31].

2.3. Design of TbTOP1B specific ligands

The modeled 3D structure of the TbTOP1B dimer was used to generate TbTOP1B specific ligands by employing LigBuilder2 [37]. To this end, the structural constrains of the binding pocket of TbTOP1B were analyzed and ligands were constructed in a stepwise fashion using a library of fragments. A genetic algorithm, implemented in LigBuilder2, controlled the overall ligand construction process. During construction, minimization of conformation was performed.

The LigBuilder suite is capable of using both the "link" and the "grow" approaches and to implement a genetic algorithm based search for the determination of the best structures. The linking procedure, works by asking the algorithm to select the most suitable moieties to interact with the active site of the protein and to start linking them together in a chemically appropriate way. The growing procedure involves the determination of a group as a starting point and then the growing of a larger compound that would fit in the active site, while being capable of establishing interactions with it. It is important that the first approach requires that the user must supply the docking/interaction points of the moieties that will be used as anchors to the final structure. Once the compounds are drawn from the population, the new population is formed by combining the information from groups of selected individuals. This is done by randomly splitting the representative strings of the parents and recombining them to form new entities. This operation is called Crossover. Because the parents have been selected for their fitness, the children of these parents will, statistically, be more fit than the parents. This is done until a new population has been built from the old one. The final step is the evaluation of the new population, which may result in starting the whole process over again by changing a single or groups of parameters.

The program took into consideration the flexibility of the ligand, whereas the protein TbTOP1B was kept rigid. In this way, a database of novel ligand fragments was generated. The database was further filtered; thus, the included ligand molecules exhibited chemical properties which were correlated more strongly to the structural and physicochemical characteristics of the receptor binding pocket. A post-filtering step involved the use of statistical methods to deduce a pattern from the ligand molecules.

2.4. 3D pharmacophore elucidation

The "Complexed-based" pharmacophore module as implemented in the MOE suite [38] was employed to generate a 3D pharmacophore model. Initially, a set of Pharmacophoric Annotation Points (PAPs) was prepared for a series of amino acid residues in the TbTOP1B. These residues were selected based on their degree of conservation among kinetoplastids (i.e. the residues that were more conserved between kinetoplastid TOP1Bs and less conserved among other species). This was achieved by visual inspection of the multiple alignment of the TOP1B amino acid sequences. The highest ranking 3D pharmacophore hypothesis, as a grouped 3D arrangement of PAPs, was selected since it represents the best match to the collective structural and physicochemical properties necessary for optimal ligand–receptor interaction.

3. Results/discussion

3.1. Phylogenetic analysis of TOP1B

In the present study, a phylogenetic analysis of TOP1B homologs in all available genomes, along with domain organization analysis of the identified putative proteins, was performed. (See Fig. 1) Based on our findings, putative members of the TOP1B family were identified in the major eukaryotic and prokaryotic taxonomic divisions, as well as viruses (Fig. 2A and Supplementary Table 1). Furthermore, all protein sequences under study were found to possess five signature motifs where the five catalytically important residues are located, including tyrosine (Fig. 2B).

In eukaryotes, except kinetoplastids, the protein TOP1B exists as a composite protein (Fig. 2A, blue clade). In archaea, TOP1B homologs were detected as single fused proteins only in the phylum thaumarchaeota (Fig. 2A, blue clade); after extensive database searches, none of the two subunits was detected in the archaeal phyla euryarchaeota and crenarchaeota. The eukaryotic and thaumarchaeal TOP1Bs cluster together (Fig. 2A). It is suggested that either thaumarchaeota acquired a *TOP1B* gene from eukaryotes through horizontal gene transfer or a *TOP1B* copy might have existed also in euryarchaeota and crenarchaeota that got lost during evolution [39].

In our analysis, in accordance with previous reports [14,15,40], we found that in kinetoplastids the protein TOP1B is split into its two component proteins (Fig. 2A, red clade). There are two alternative possible hypotheses that could explain this phenomenon. According to the first hypothesis, a single ancestral *TOP1B* gene probably underwent fission and was split into two separate genes. Conversely, two independent genes may have existed which were fused over time, except in kinetoplastids. Given that TOP1B proteins are similar in sequence and mechanism to the tyrosine recombinases [41], Krogh and Shuman suggested that the two enzyme families harbor a common catalytic domain that fused during evolution with different amino-terminal domains [8]. Over the course of evolution, genes are fused in order to confer advantages such as decrease of the regulational load in the cell for certain processes, or to yield new genes with novel functions [42,43]. In a similar manner, the fusion of the genes encoding the DNA binding and the catalytic domain may have taken place in order to produce the novel gene *TOP1B*. However, this is not the case in the kinetoplastid TOP1B, where there must have been evolutionary pressure to TOP1B to maintain its dimeric architecture, as it allows to the two domains to function independently and, also be replaced in case of damage ensuring that the enzymatic activity is not interrupted [14].

Furthermore, in the phylogenetic tree in Fig. 2A, the kinetoplastid TOP1B proteins form a distinct monophyletic branch which is basal to the clade that corresponds to the other eukaryotic proteins (supported by a high bootstrap value). This adds further support to the hypothesis that the kinetoplastid TOP1Bs diverged early from the other eukaryotic TOP1Bs.

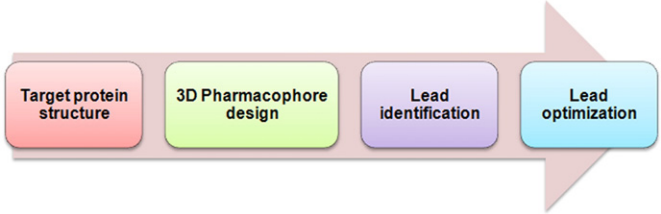

Fig. 1. Schematic illustration of the drug design process.

3.2. Homology modeling of the dimeric TbTOP1B

In the present study, the dimeric trypanosome TbTOP1B, was modeled complexed with a cleaved dsDNA substrate using the sequence alignment of Fig. 3. The two protomers together make the fused protein of the *T. brucei* DNA Topoisomerase IB. According to this model, the large DNA binding domain of TbTOP1B clamp wraps around the substrate DNA and the carboxy-terminal region of the large domain interacts with the small, catalytic domain (Fig. 4). This is a great opportunity that distinguishes the *T. brucei* DNA Topoisomerase from the corresponding human one, as through the fusion event it creates a unique structural conformation for the *T. brucei* that does not exist in the non-fused human Topoisomerase. This is very important for the specificity of *in silico* inhibitor screening as well as for reduced expected human toxicity. The template structure is colored yellow in Fig. 4, while the first component of the fused trypanosome protein is colored green and the second one is colored blue. The dsDNA substrate is colored gray. It is clear that due to the relatively high sequence identity (69%), the model of the trypanosome topoisomerase was very similar in its 3D structural arrangement to its template. The catalytic tyrosine residue is covalently attached to the nicked DNA backbone. Furthermore, as suggested by previous studies, although the human and kinetoplastid TOP1B proteins have different architecture (monomer versus dimer), the overall structure and catalytic mechanism are highly conserved [30]. The 3D structure of the trypanosome TOP1B and TbTOP1B was modeled so that structure-based drug design and pharmacophore elucidation studies could be carried out. The model was then *in silico* evaluated for its reliability. Firstly, it was structurally superimposed and subsequently compared to its template, where it exhibited an alpha-carbon RMSD of less than 0.5 Å (Fig. S1). Furthermore it was evaluated with MOE and PROCHECK for its geometry (see supplementary material) and then subjected to the Verify3D algorithm for a more thorough evaluation (Fig. S2). Verify3D assessed the compatibility of the 3D model of Topoisomerase model with its own amino acid sequence. Based on location and environment, a structural class is assigned for each residue. A collection of reference structures is used as a control in order to calculate a score for each residue. The Topoisomerase model ranged from +0.28 to +0.78. That confirmed that the model is of high quality, since Verify3D scores below +0.1 are indicative of serious problems in the model [23].

3.3. Pharmacophore elucidation and virtual screening

Camptothecins are specific anti-TOP1B agents that act as uncompetitive inhibitors that bind to the TOP1B enzymes and substrate DNA resulting in an immobile ternary complex which prevents religation of the DNA strands [44,45]. The kinetoplastid and mammalian cells exhibit comparable susceptibility to camptothecin [46,47]. This could be explained by the fact that the residues that are located in the camptothecin-binding pocket of TOP1B were found to be conserved between the kinetoplastid *Leishmania* and human [30].

The aim of our study was the identification of selective agents that specifically target the kinetoplastid TbTOP1B, avoiding in this way any harmful side effects on the human host. Towards this end, a set of 12 *de novo* designed drug-like compounds, which can potentially bind to TbTOP1B and inhibit its activity, was identified (Fig. S3). When we

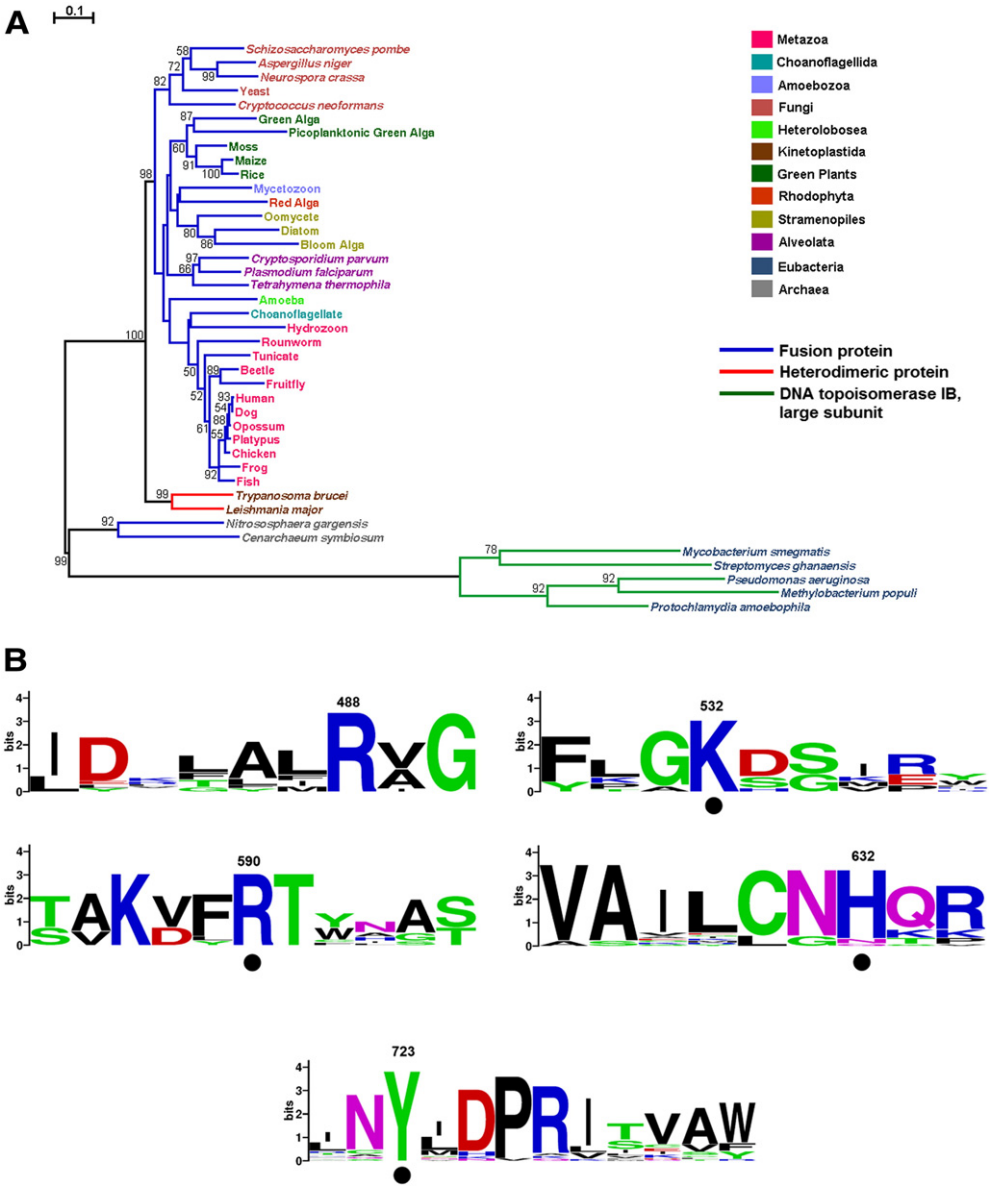

Fig. 2. A) Phylogenetic tree of TOP1B proteins. Bootstrap values (>50%) are shown at the nodes. The length of the tree branches reflects evolutionary distance. The scale bar at the upper left represents the length of amino acid substitutions per position. In some cases, the species common names were used instead. Different colors are used for different taxonomic divisions. The clades that correspond to species where the TOP1B is fused, are indicated by blue, split by red. B) Sequence logo of the five catalytic motifs identified in TOP1B proteins. The height of each letter represents the frequency of the corresponding amino acid residue at that position, and the letters are ordered in such a way that the most frequent is on the top. The residues essential for TOPiB catalytic activity are indicated by dots and are numbered according to the human TOP1B.

superimposed our modeled structure of TbTOP1B with its human counterpart (PDB ID: 1A31) [48], the residues that line the binding site of the 12 ligands were not conserved in the two species; instead they were specific to trypanosomes (data not shown).

Subsequently, a 3D pharmacophore model was designed that incorporated all spatial and electronic features that are necessary to ensure optimal non-covalent interactions with the TbTOP1B enzyme (Fig. 5). Evolutionary information in the form of conserved amino acids was also taken into account in pharmacophore design in order to increase the possibility of finding molecules with drug-like properties that specifically target the TbTOP1B protein.

The 3D Pharmacophore design method that was used in this study was a receptor-structure based one. A pharmacophore elucidation study was conducted within MOE by automatic detection of possible cavities as active site. The automatic cavity selection was conducted using the alpha-sphere cavity detector of MOE. The highest ranking cavity was located next to the interaction site of the two modeled components. This is very important, as the structure-based pharmacophore shared features from both protomers. The final pharmacophore model was the overlaying of 35 different pharmacophores that were reduced to their common features. In this way, a set of receptor-inhibitor ideal theoretical interaction fingerprints was deduced. Fig. 5 depicts the 3D pharmacophore that was established. It was found that the *T. brucei* DNA Topoisomerase IB pharmacophore comprised of five functional groups: two electron accepting groups (Fig. 5, blue colored), two aromatic regions (Fig. 5, green colored) and one electron donating group (Fig. 5, magenta colored). The ultimate aim for the establishment of the 3D pharmacophore for the *T. brucei* DNA Topoisomerase IB was to enable us to perform ultra fast *in silico* virtual screening of large drug-like databases towards the identification of putative compounds that may have anti-*T. brucei* DNA Topoisomerase IB activity. Collectively, according to our *in silico* prediction model, a potent candidate inhibitor of

```
protomer 1      KTVSLGTSKVNYIDPRIVCSWANENNVPISRLFSATLQKKFPWALKARDFTF--------
protomer 2      ----------------------------------------------------ETEWWTQG
2B9S            KAVSLGTSKINYIDPRIICSWAKAQDVPINKIFSATIQKKFPWAMNAENFDFDLNWWEQE

protomer 1      ------------------------------------------------------------
protomer 2      TLTITKKGEKRWDILLHNGVLFPPAYVPHGIPILYNQQKFEMTPEEEEVATMFAVLREHD
2B9S            NLRIAMKGERRWETLAHNGVLFPPEYEPHGIPIFYDGREFKMTPEEEEVATMFAVMKEHD

protomer 1      ------------------------------------------------------------
protomer 2      YYRNEVFRRNFFQSWREILDKRKHPIRCLELCDFSAIYEWHQREVEKRKSRTREEKKELK
2B9S            YYRMEVFRRNFFESWREILDKRQHPIRRLELCDFEPIYQWHLVQREKKLSRTKEEKKAIK

protomer 1      ------------------------------------------------------------
protomer 2      RIADEEAEPYKWCIWNGKKEQVANFRVEPPGLFRGRGEHPMRGKLKKRILPEDVVLNIGK
2B9S            EKQDAEAEPYRYCVWDGRREQVANFRVEPPGLFRGRGKHPLMGKLKVRVQPEDITINIGE

protomer 1      ------------------------------------------------------------
protomer 2      EAPIPQAPAGHKWKGVVHDQNVTWLAMWYEPTIGQCKYVMLAPSSTLKGQSDYAKFETAR
2B9S            TAEVPVPPAGHKWAAVQHDHTVTWLAMWRDSVAGNMKYVMLAPSSSVKGQSDMVKFEKAR

protomer 1      ------------------------------------------------------------
protomer 2      ELKNHIDDIRESYTKDFSSTDEMERQRAVATYFIDKLALRVGHEKGEEEADTVGCCSLRK
2B9S            KLKDKVDDIRASYMEDFKSNDLHVAQRAVAMYFIDRLALRVGNEKGEDEADTVGCCSLRV

protomer 1      ------------------------------------------------------------
protomer 2      EHIELRPDNVVRFDFLGKDSIRYVNEVTVLPEVYKLLGSFIKRTDS--EIFRKVTPTTLN
2B9S            EHIQLMPDNIVRFDFLGKDSIRYQNDVAVLPEVYALLQRFTRRKSPGMDIFDQLNPTQLN

protomer 1      ------------------------------------------------------------
protomer 2      NYLKSFLKDLSAKVFRTYNASITLDEWFREKPVDPKASLSDKLVYFNKANTEVAKLCNHQ
2B9S            DHLKSFMDGLSAKVFRTYNASITLDRWFKEKPW----STADKLAYFNKANTEVAILCNHQ

protomer 1      --
protomer 2      RS
2B9S            KS
```

Fig. 3. Sequence alignment. Each one of the two fused protomers is aligned against the template Crystal Structure of heterodimeric *L. donovani* topoisomerase I-vanadate-DNA complex (RCSB code: 2B9S).

T. brucei DNA Topoisomerase IB should satisfy all of the previously described pharmacophoric features. All pharmacophoric features were subsequently used as a query to sequentially screen the National Cancer Institute (NCI) database of chemical compounds through virtual High Throughput Screening (vHTS). After a series of *in silico* docking and ranking a total of 999 lead compounds were collected that conform to the pharmacophoric features and can be considered potential TbTOP1B inhibitors. The highest ranking compound was found to be compound DV1, which fitted accurately our model in its estimated bioactive conformation (Fig. 6). It is obvious that when the DV1 compound is bound to the *T. brucei* DNA Topoisomerase IB, it is blocking the DNA channel, thus inhibiting its passage and its processing by the *T. brucei* DNA Topoisomerase IB (Fig. 6A). More specifically in Fig. 6B, it is clear that the compound interacting residues Arg481 and Thr228 come from two different protomers of the *T. brucei* DNA Topoisomerase IB model. The two latter residues were found to interact with the DNA bases in the initial model. Therefore, establishing strong hydrogen bonding interactions with those amino acids will create a "roadblock" for the incoming DNA oligonucleotide chain. More specifically the DV1 compound competes with DNA since, its Adenine-like base set of conjugated rings, mimics the nucleotide binding form as it occupies the same 3D spatial coordinates of the DNA nucleotide (Fig. 6C). Gly227 stabilizes the conjugated aromatic rings by establishing hydrophobic, pi-stacking interactions.

Collectively, herein our aim is to model the three dimensional structure of the *T. brucei* DNA Topoisomerase IB towards the development of a 3D specific pharmacophore, due to its emerging potential as a pharmaceutical target suitable for drug design. In an effort to improve our understanding of residue conservation in the structure of the *T. brucei* DNA Topoisomerase IB, a full scale, across species phylogenetic analysis was conducted. In-depth analysis, cross merging of the phylogenetic and the structural data resulted in a concise 3D *in silico* model, which explains and rationalizes the structure of the *T. brucei* DNA Topoisomerase IB. The resulting 3D model suggested a new perspective that led to the establishment of a specific 3D pharmacophore for the *T. brucei* DNA Topoisomerase IB. We used the 3D model and the designed pharmacophore for the *in*

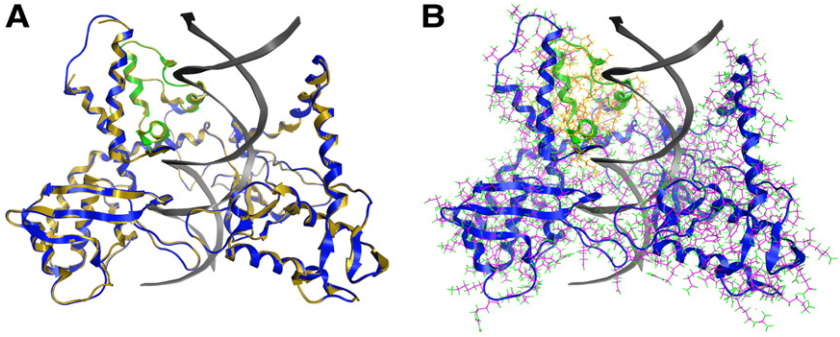

Fig. 4. Homology model of the TbTOP1B heterodimer bound to nicked DNA. A) The DNA binding domain of TbTOP1B is shown in blue, the catalytic domain is indicated by green and the DNA substrate is depicted in gray color. B) Uses same color-conventions as A, but is also a full atom display of the model. The binding domain is magenta atom color, while the catalytic domain is orange atom color.

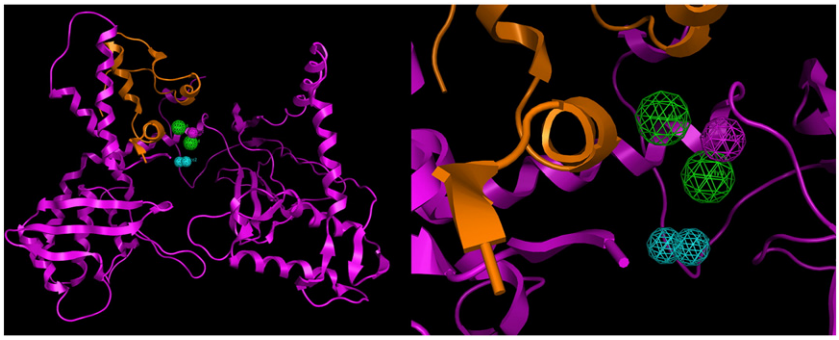

Fig. 5. The 3D Pharmacophore of the *Trypanosoma brucei* DNA Topoisomerase IB model structure. Here, the binding domain of TbTOP1B is shown in magenta and the catalytic domain in orange color. Purple and blue colors correspond to electron donating and accepting groups, respectively and green to aromatic groups.

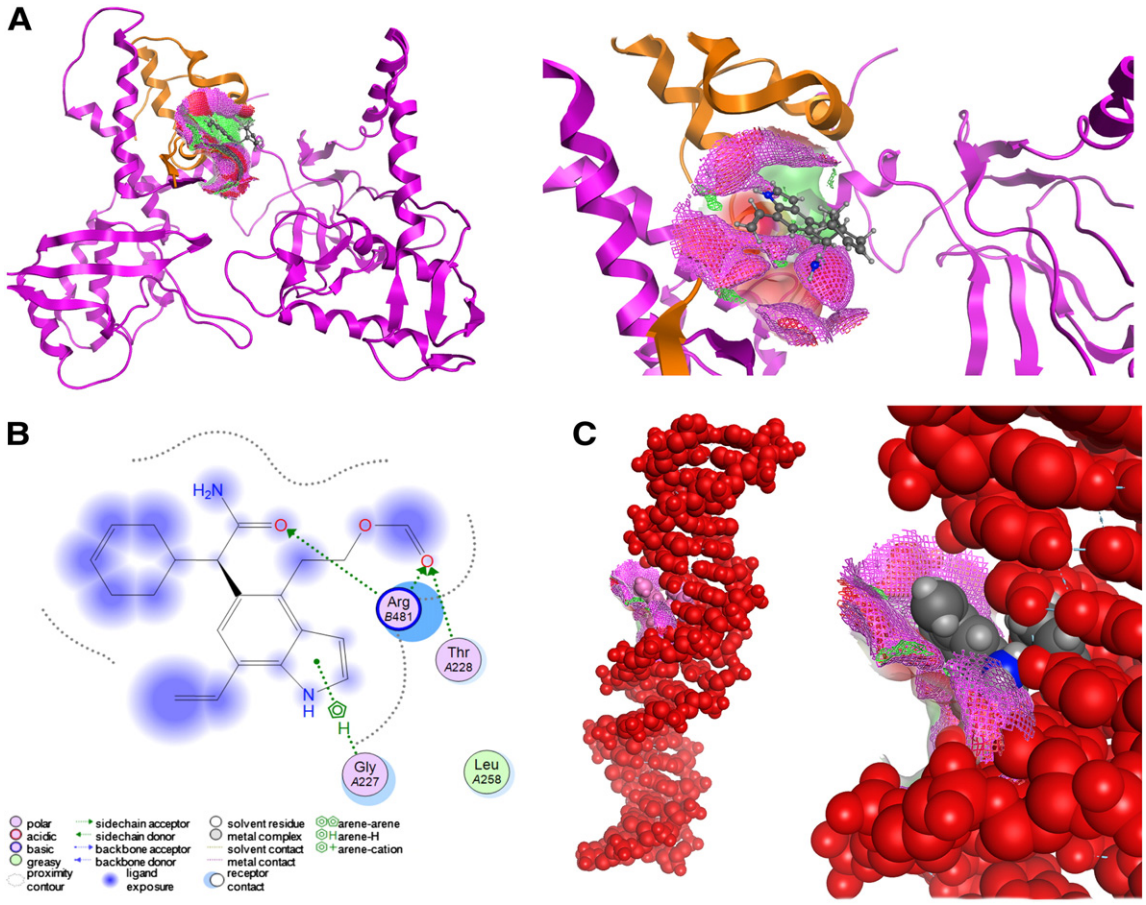

Fig. 6. *In silico* virtual screening. A) The 3D pharmacophore was used to screen the full NCI compound database. B) This compound (DV1) was the highest ranking one as it established interactions with residues from both protomers of the *Trypanosoma brucei* DNA Topoisomerase IB model. C) The selected compound clearly blocks the passage of the ssDNA substrate in the topoisomerase model. The dsDNA is shown in red spacefill format.

silico virtual screening of large compound databases that led to the identification of a lead compound as a promising inhibitor, which will modulate the activity of the enzyme. The present work provides insights for future drug design of novel compounds with improved biochemical and clinical characteristics as anti-*T. brucei* DNA Topoisomerase IB agents.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2013.11.008.

Acknowledgments

The authors would like to sincerely thank Dr Lila Koumandou for critically reviewing the manuscript.

References

[1] J.J. Champoux, DNA topoisomerases: structure, function, and mechanism, Annu. Rev. Biochem. 70 (2001) 369–413.
[2] J. Roca, The mechanisms of DNA topoisomerases, Trends Biochem. Sci. 20 (1995) 156–160.
[3] J.C. Wang, Cellular roles of DNA topoisomerases: a molecular perspective, Nat. Rev. Mol. Cell Biol. 3 (2002) 430–440.
[4] A. Patel, L. Yakovleva, S. Shuman, A. Mondragon, Crystal structure of a bacterial topoisomerase IB in complex with DNA reveals a secondary DNA binding site, Structure 18 (2010) 725–733.
[5] W.R. Bauer, E.C. Ressner, J. Kates, J.V. Patzke, A DNA nicking-closing enzyme encapsidated in vaccinia virus: partial purification and properties, Proc. Natl. Acad. Sci. U. S. A. 74 (1977) 1841–1845.
[6] J.J. Champoux, R. Dulbecco, An activity from mammalian cells that untwists superhelical DNA—a possible swivel for DNA replication (polyoma-ethidium

bromide-mouse-embryo cells-dye binding assay), Proc. Natl. Acad. Sci. U. S. A. 69 (1972) 143–146.
[7] P. D'Arpa, P.S. Machlin, H. Ratrie III, N.F. Rothfield, D.W. Cleveland, et al., cDNA cloning of human DNA topoisomerase I: catalytic activity of a 67.7-kDa carboxyl-terminal fragment, Proc. Natl. Acad. Sci. U. S. A. 85 (1988) 2543–2547.
[8] B.O. Krogh, S. Shuman, A poxvirus-like type IB topoisomerase family in bacteria, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 1853–1858.
[9] S. Shuman, Vaccinia virus DNA topoisomerase: a model eukaryotic type IB enzyme, Biochim. Biophys. Acta 1400 (1998) 321–337.
[10] D.A. Koster, V. Croquette, C. Dekker, S. Shuman, N.H. Dekker, Friction and torque govern the relaxation of DNA supercoils by eukaryotic topoisomerase IB, Nature 434 (2005) 671–674.
[11] B.O. Krogh, S. Shuman, Catalytic mechanism of DNA topoisomerase IB, Mol. Cell 5 (2000) 1035–1041.
[12] L. Stewart, G.C. Ireton, J.J. Champoux, The domain organization of human topoisomerase I, J. Biol. Chem. 271 (1996) 7602–7608.
[13] L. Stewart, G.C. Ireton, L.H. Parker, K.R. Madden, J.J. Champoux, Biochemical and biophysical analyses of recombinant forms of human topoisomerase I, J. Biol. Chem. 271 (1996) 7593–7601.
[14] A.L. Bodley, A.K. Chakraborty, S. Xie, C. Burri, T.A. Shapiro, An unusual type IB topoisomerase from African trypanosomes, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 7539–7544.
[15] H. Villa, A.R. Otero Marcos, R.M. Reguera, R. Balana-Fouce, C. Garcia-Estrada, et al., A novel active DNA topoisomerase I in *Leishmania donovani*, J. Biol. Chem. 278 (2003) 3521–3526.
[16] M. Magrane, U. Consortium, UniProt Knowledgebase: a hub of integrated protein data. Database (Oxford), 2011. (bar009).
[17] D.A. Benson, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, E.W. Sayers, GenBank, Nucleic Acids Res. 39 (2011) D32–D37.
[18] M. Matsuzaki, O. Misumi, I.T. Shin, S. Maruyama, M. Takahara, et al., Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D, Nature 428 (2004) 653–657.
[19] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.
[20] C.J. Sigrist, L. Cerutti, E. de Castro, P.S. Langendijk-Genevaux, V. Bulliard, et al., PROSITE, a protein domain database for functional characterization and annotation, Nucleic Acids Res. 38 (2010) D161–D166.
[21] J.D. Thompson, T.J. Gibson, D.G. Higgins, Multiple sequence alignment using ClustalW and ClustalX, Curr. Protoc. Bioinforma. 2 (2002)(Chapter 2).
[22] J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis, Mol. Biol. Evol. 17 (2000) 540–552.
[23] G. Talavera, J. Castresana, Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments, Syst. Biol. 56 (2007) 564–577.
[24] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, et al., MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, Mol. Biol. Evol. 28 (2011) 2731–2739.
[25] D.T. Jones, W.R. Taylor, J.M. Thornton, The rapid generation of mutation data matrices from protein sequences, Comput. Appl. Biosci. 8 (1992) 275–282.
[26] D.H. Huson, D.C. Richter, C. Rausch, T. Dezulian, M. Franz, et al., Dendroscope: an interactive viewer for large phylogenetic trees, BMC Bioinforma. 8 (2007) 460.
[27] S. Pettifer, D. Thorne, P. McDermott, J. Marsh, A. Villeger, et al., Visualising biological data: a semantic approach to tool and database integration, BMC Bioinforma. 10 (Suppl. 6) (2009) S19.
[28] G.E. Crooks, G. Hon, J.M. Chandonia, S.E. Brenner, WebLogo: a sequence logo generator, Genome Res. 14 (2004) 1188–1190.
[29] A. Sali, L. Potterton, F. Yuan, H. van Vlijmen, M. Karplus, Evaluation of comparative protein modeling by MODELLER, Proteins 23 (1995) 318–326.
[30] D.R. Davies, A. Mushtaq, H. Interthal, J.J. Champoux, W.G. Hol, The structure of the transition state of the heterodimeric topoisomerase I of *Leishmania donovani* as a vanadate complex with nicked DNA, J. Mol. Biol. 357 (2006) 1202–1210.
[31] W.L. DeLano, The PyMOL User's Manual, DeLano Scientific, San Carlos, CA, USA., 2002.
[32] J.C. Wang, P. Cieplak, P.A. Kollman, How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules, Comput. Chem. Mol. Model. 21 (2000) 1049–1074.
[33] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation, J. Chem. Theory Comput. 4 (2008) 435–447.
[34] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, et al., GROMACS: fast, flexible, and free, J. Comput. Chem. 26 (2005) 1701–1718.
[35] R.A. Laskowski, J.A. Rullmannn, M.W. MacArthur, R. Kaptein, J.M. Thornton, AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, J. Biomol. NMR 8 (1996) 477–486.
[36] D. Eisenberg, R. Luthy, J.U. Bowie, VERIFY3D: assessment of protein models with three-dimensional profiles, Methods Enzymol. 277 (1997) 396–404.
[37] Y. Yuan, J. Pei, L. Lai, LigBuilder 2: A Practical de Novo Drug Design Approach, J. Chem. Inf. Model. 51 (5) (2011) 1083–1091, http://dx.doi.org/10.1021/ci100350u.
[38] CC Group, Molecular Operating Environment (MOE). 1010 Sherbrooke St. West, Suite 910, Montreal, Canada, H3A 2R, 2012.
[39] C. Brochier-Armanet, S. Gribaldo, P. Forterre, A DNA topoisomerase IB in Thaumarchaeota testifies for the presence of this enzyme in the last common ancestor of Archaea and Eucarya, Biol. Direct 3 (2008) 54.
[40] D. Dimitriadis, V.L. Koumandou, P. Trimpalis, S. Kossida, Protein functional links in *Trypanosoma brucei*, identified by gene fusion analysis, BMC Evol. Biol. 11 (2011) 193.
[41] C. Cheng, P. Kussie, N. Pavletich, S. Shuman, Conservation of structure and mechanism between eukaryotic topoisomerase I and site-specific recombinases, Cell 92 (1998) 841–850.
[42] A.J. Enright, C.A. Ouzounis, Functional associations of proteins in entire genomes by means of exhaustive detection of gene fusions, Genome Biol. 2 (9) (2001) research0034.1–research0034.7, (PMCID: PMC65099).
[43] M. Long, A new function evolved from gene fusion, Genome Res. 10 (2000) 1655–1657.
[44] T.K. Li, L.F. Liu, Tumor cell death induced by topoisomerase-targeting drugs, Annu. Rev. Pharmacol. Toxicol. 41 (2001) 53–77.
[45] L.F. Liu, S.D. Desai, T.K. Li, Y. Mao, M. Sun, et al., Mechanism of action of camptothecin, Ann. N. Y. Acad. Sci. 922 (2000) 1–10.
[46] A.L. Bodley, T.A. Shapiro, Molecular and cytotoxic effects of camptothecin, a topoisomerase I inhibitor, on trypanosomes and *Leishmania*, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 3726–3730.
[47] A.L. Bodley, M.C. Wani, M.E. Wall, T.A. Shapiro, Antitrypanosomal activity of camptothecin analogs. Structure-activity correlations, Biochem. Pharmacol. 50 (1995) 937–942.
[48] M.R. Redinbo, L. Stewart, P. Kuhn, J.J. Champoux, W.G. Hol, Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA, Science 279 (1998) 1504–1513.