# Genetic and structural study of DNA-dependent RNA polymerase II of Trypanosoma Brucei, towards the designing of novel antiparasitic agents.

**Louis Papageorgiou** [1, 2, 3] , **Vasileios Megalooikonomou** [3] , **Dimitrios Vlachakis** [Corresp. 2, 3]

[1] Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Athens, Greece

[2] Computational Biology & Medicine Group, Biomedical Research Foundation, Academy of Athens, Athens, Attica, Greece

[3] Computer Engineering and Informatics Department, University of Patras, Patra, Greece

Corresponding Author: Dimitrios Vlachakis
Email address: dvlachakis@bioacademy.gr

*Trypanosoma brucei brucei*  (TBB) belongs to the unicellular parasitic protozoa organisms, specifically to the  *Trypanosoma* genus of the  *Trypanosomatidae* class. A variety of different vertebrate species can be infected by TBB including humans and animals. Under particular conditions, the TBB can be hosted by wild and domestic animals; thereby an important reservoir of infection always remains available to transmit through the tsetse flies. Although the TBB parasite is one of the leading causes of death in the most underdeveloped countries, to date, there is neither vaccination available nor any drug against TBB infection. The subunit RPB1 of the TBB DNA-directed RNA polymerase II (DdRpII) constitutes an ideal target for the design of novel inhibitors, since it is instrumental role is vital for the parasite's survival, proliferation, and transmission. A major goal of the described study is to provide insights for novel anti-TBB agents via a state of the art drug discovery approach of the TBB DdRpII RPB1. In an attempt to understand the function and action mechanisms of this parasite enzyme related to its molecular structure, an in-depth evolutionary study has been conducted in parallel to the  *in silico* molecular designing of the 3D enzyme model, based on state of the art comparative modelling and molecular dynamics techniques. Based on theevolutionary studies results nine new invariant, first-time reported, highly conserved regions have been identified within the DdRpII family enzymes. Consequently, those patches have been examined both at the sequence and structural level and have been evaluated in regards to their pharmacological targeting appropriateness. Finally, the pharmacophore elucidation study enabled us to virtually  *in silico*  screenhundreds of compounds and evaluate their interaction capabilities with the enzyme. It was found that a series of Chlorine-rich set of compounds were the optimal inhibitors for the TBB DdRpII RPB1 enzyme. All-in-all, herein we present a series of new sites on the TBB DdRpII RPB1 of high pharmacological interest, alongside the

construction of the 3D model of the enzyme and the suggestion of a new *in silico* pharmacophore model for fast screening of potential inhibiting agents.

1 **Genetic and structural study of DNA-dependent RNA polymerase II of Trypanosoma Brucei,**
2 **towards the designing of novel antiparasitic agents.**

4 **Louis Papageorgiou[1,2,3], Vasileios Megalooikonomou[2] and Dimitrios Vlachakis[,1,2,#]**
5

7 [1]Computational Biology & Medicine Group, Biomedical Research Foundation, Academy of Athens, Soranou
8 Efessiou 4, Athens 11527, Greece
9 [2]Computer Engineering and Informatics Department, University of Patras, Patras University Campus, 26504,
10 Greece
11 [3]Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, University
12 Campus, Athens, 15784, Greece

14

29

30

31

32 #Correspondence to:
33 Dimitrios Vlachakis,

34    Bioinformatics & Medical Informatics Team, Biomedical Research Foundation, Academy of
35    Athens, SoranouEfessiou 4, Athens 11527, Greece
36    Tel: + 30 210 6597 647, Fax: +30 210 6597 545
37    E-mail: dvlachakis@bioacademy.gr

38    **Abstract**

39    *Trypanosoma brucei brucei* (TBB) belongs to the unicellular parasitic protozoa organisms,

40    specifically to the *Trypanosoma* genus of the *Trypanosomatidae* class.  A variety of different

41    vertebrate species can be infected by TBB including humans and animals. Under particular

42    conditions, the TBB can be hosted by wild and domestic animals; thereby an important

43    reservoir of infection always remains available to transmit through the tsetse flies. Although the

44    TBB parasite is one of the leading causes of death in the most underdeveloped countries, to

45    date, there is neither vaccination available nor any drug against TBB infection. TBB DNA-

46    dependent RNA polymerase II (DdRpII subunit RPB1) is an ideal target for the design of novel

47    inhibitors against TBB. This enzyme plays a critical role in parasite's survival, proliferation, and

48    transmission. A major goal of the described study is to provide insights for novel anti-TBB

49    agents via a state of the art drug discovery approach of the TBB DdRpII RPB1. In an attempt to

50    understand the function and action mechanisms of this parasite enzyme related to its

51    molecular structure, an in-depth evolutionary study has been conducted in parallel to the *in*

52    *silico* molecular designing of the 3D enzyme model, based on state of the art comparative

53    modelling and molecular dynamics techniques. Based on the evolutionary studies results nine

54    new invariant, first-time reported, highly conserved regions have been identified within the

55    DdRpII family enzymes. Consequently, those patches have been examined both at the sequence

56    and structural level and have been evaluated in regards to their pharmacological targeting

57    appropriateness. Finally, a 3D pharmacophore model was constructed specifically for the TBB

58    DdRpII RPB1 enzyme. All-in-all, herein we present a series of new sites on the TBB DdRpII RPB1

59    of high pharmacological interest, alongside the construction of the 3D model of the enzyme and

60    the suggestion of a new *in silico* pharmacophore model for fast screening of potential inhibiting

61    agents.

62

63

**Introduction**

African trypanosome parasites cause human sleeping sickness and nagana in Africa, Asia, and South America. More than 95% of reported cases are caused by two subspecies of *Trypanosoma brucei brucei* (TBB), the *Trypanosoma brucei gambiense* (TBG) and the *Trypanosoma brucei rhodesiense* (TBR) which is found in western and central Africa (Berriman et al. 2005; World Health Organization 2015). The parasitic infection is transmitted by tsetse flies, which breed in warm and humid areas.  Tsetse flies are found living in 36 countries in sub-Saharan Africa, thus putting 60 million people at risk. Currently, about 10,000 new cases each year are reported by the World Health Organization (WHO).  Moreover, it is believed that many cases are undiagnosed and unreported.  Sleeping sickness can be curable with medication, but it may be fatal if it is left untreated. It is estimated that Human deaths caused by Sleeping sickness are of about 48,000 annually. Bites by the tsetse fly erupt into a red sore on the skin and in the following weeks, the person may have to deal with several symptoms including fever, swollen lymph glands, aching muscles, headaches, and irritability.  In advanced stages, the TBB parasite attacks the central nervous system of the host, and in general consul some disorders in personality, circadian rhythm, serenity, speech, and difficulties in walking. Despite the significant treatment advances for patients with sleeping sickness, the parasite's progression is often inevitable and needs more treatment options. Until today, drugs can only be used in the early stages of the disease and without providing 100% reassurance for full convalesce of the patient (Ridley 2002; Ross et al. 2007; Trouiller et al. 2002). The TBB parasite starts its activity after each invasion through its proteins, specifically with its replication enzymes including helicases and polymerases. Such enzymes are ideal targets for inhibitor design since those proteins are crucial for the TBB parasite survival. Being already in possession of the widely known sequence of the DNA-dependent RNA polymerase II (DdRpII) RPB1 (Chung et al. 1993) which plays a significant role in the replication of the parasite, our primary goal is to suppress its function towards replication itself when it infects a human. Although TBB has been reported many times in the past, the three-dimensional structure of its essential enzymes like DdRpII remains unknown so far (Malvy & Chappuis 2011).

Protein structure has been found to be three to ten times more conserved than sequence (Illergard et al. 2009). Thus, when possible, it is preferable to study an enzyme's 3D structure rather than its sequence. Knowledge of the tertiary structure can assist in the understanding of relationships between structure and function (Berg et al. 2002). Herein, the three-dimensional structure of DdRpII subunit RPB1 has been modelled, in an effort to predict the 3D molecular structure that is linked to the function of this enzyme (Bayele 2009; Koch et al. 2016). The molecular model has been constructed using conventional molecular modelling techniques and a known 3D crystal structure of *Schizosaccharomyces Pombe* RNA Polymerase II as a template (Papageorgiou et al. 2014b). The established molecular model of the DdRpII RPB1 enzyme of TBB exhibits all known structural motifs that are unique to the DdRpII RPB1 enzymes.

Upon successful completion of the 3D structure prediction of the TBB DdRpII RPB1 protein, molecular dynamics simulations have been performed to structurally improve and benchmark the quality of the 3D model. Moreover, the reliability and viability of the TBB DdRpII RPB1 model were checked using several *in silico* scoring tools such as MOE and Procheck. After

108  the model validation process, a *de novo* structure-based drug design approach has been
109  performed, which led to the establishment of a 3D novel pharmacophore model that is highly
110  specific for the DdRpII RPB1 enzyme of TBB. The generated pharmacophore model may be used
111  in future experiments involving the high throughput virtual screening of large compound
112  databases towards the identification of novel anti-TBB agents (Loukatou et al. 2014). The
113  present work opens the field for the design of novel compounds with improved biochemical
114  and clinical characteristics in the future.
115  **Methods**
116
117  **Database sequence search**
118  The full-length protein sequences related to the DdRpII family were extracted from the NCBI
119  database. In total, 36 DdRpII protein sequences were downloaded from several species with
120  fully sequenced genomes (Supplementary data 1).
121
122  **Genetic and evolutionary analyses**
123  Multiple sequence alignment of the DdRpII protein family sequences were performed using two
124  different programs, MUSCLE (Edgar 2004) and CLUSTALW (Chenna et al. 2003; Thompson et al.
125  1994). In the next step, multiple sequence alignment was checked with ProtTest3 (Darriba et al.
126  2011) to estimate the appropriate model of sequence evolution. Phylogenetic analyses were
127  performed by two different ways, and two representative phylogenetic trees were constructed
128  for the DdRpII dataset (Vlachakis et al. 2014b). The first phylogenetic tree was constructed
129  using the MEGA software (Stecher et al. 2014) utilizing Bayesian and Maximum Likelihood
130  statistical methods as described in with 100 bootstrap replicates (Papageorgiou et al. 2016)
131  (Figure 1 and Supplementary data 2). The second phylogenetic tree was constructed using the
132  Jalview software (Waterhouse et al. 2009) utilizing the neighbour joining statistical method in
133  with 100 bootstrap replicates (Supplementary Figure 1 and Supplementary data 3).
134
135  **Conserved motifs exploration**
136  The phylogenetic trees that derived from the phylogenetic analyses (Jalview and MEGA) were
137  separated in sub-trees, in order to extract the most highly related protein sequences of the TBB
138  DdRpII RPB1 family for the conserved motifs exploration (Figure 2). The full-length amino acid
139  sequences of the closely related proteins with the TPP DdRpII RPB1 protein were aligned using
140  the CLUSTALW (Papageorgiou et al. 2016; Thompson et al. 1994) statistical method. The
141  evolutionary conserved sequences motifs that were derived from the multiple sequence
142  alignment were identified through the consensus sequence and logo graph where generated
143  using Jalview software (Papageorgiou et al. 2016; Waterhouse et al. 2009). (Figure 2)
144
145  **Molecular modelling**
146  All calculations and visual constructions were performed using the Molecular Operating
147  Environment (MOE) version 2013.08 software package developed by Chemical Computing
148  Group (Montreal, Canada) on a cloud-based multi core High Performance Computing (HPC)
149  cluster (Loukatou et al. 2014).
150
151  **Identification of templates structures and sequence alignment**

152   The amino acid sequence of the TBB DdRpII RPB1 was retrieved from the conceptual translation
153   of the trypanosomal RNA polymerase largest subunit genes at the NCBI database
154   (http://www.ncbi.nlm.nih.gov/) (UniProtKB/Swiss-Prot: P17545.1) (Das et al. 2006; Evers et al.
155   1989). The blastp algorithm (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to identify
156   homologous structures by searching in the Protein Data Bank (PDB). The multiple sequence
157   alignment was performed using MOE (Vilar et al. 2008).
158
159   **Homology Modelling**
160   The homology modelling of the Tbb DdRPII RPB1 was carried out using MOE. The selection of
161   template crystal structures for homology modelling was based on the primary sequence
162   identity and similarity (Figure 3), and the crystal resolution (Nayeem et al. 2006). The crystal
163   structure of *Schizosaccharomyces Pombe* RNA polymerase II subunit RPB1 (PDB: 3H0G) was
164   used as template structure. The MOE homology model method is separated into four main
165   steps. First, comes a primary fragment geometry specification. Second the insertion and
166   deletions task. The third step is the loop selection and the side-chain packing, and the last step
167   is the final model selection and refinement (Figure 4 and Supplementary data 4) (Papageorgiou
168   et al. 2014a; Vlachakis et al. 2013b). Subsequently, energy minimization was done in MOE
169   initially using the Amber99 (Wang et al. 2000) force-field as implemented into the same
170   package. The energy minimization process was applied up to a gradient of 0.0001, in an effort
171   to remove the geometrical strain (Vlachakis et al. 2013a).
172
173   **Molecular electrostatic potential**
174   Molecular electrostatic potential surfaces were calculated by solving the non-linear Poisson –
175   Boltzmann equation using finite difference method as implemented into the MOE and PyMol
176   Software (Seeliger & de Groot 2010; Vilar et al. 2008). The potential was calculated on solid
177   points per side. Protein contact potential is an automated representation where the false
178   red/blue charge-smoothed surface is shown on the protein (Figure 4). Amber99 charges and
179   atomic radii were used for this calculation.
180
181   **Molecular dynamics**
182   The Molecular Dynamics simulations of the Tbb DdRPII RPB1 3D model were executed in a
183   periodic cell, which was explicitly solvated with simple point charge (SPC) water. The truncated
184   octahedron box was chosen for solvating of the model, with a set distance of 7Å clear of the
185   protein. The molecular dynamic simulations were conducted at 300 K, 1 atm with a set 2
186   fsecond step size for a total of one hundred nanoseconds (Supplementary Figure 2). For the
187   purposes of this study we opted for a NVT ensemble in a canonical environment (Vlachakis et
188   al. 2014a). NVT stands for Number of atoms, Volume, and Temperature that remain constant
189   throughout the calculation (Vlachakis 2009). The intricate zinc ions were included in the
190   molecular dynamics simulations as integral parts of the modelled biological system
191   (Chakravorty & Merz 2014; Temiz et al. 2010). However, due to the nature of the ions, we had
192   to limit the allowed degrees of freedom for those molecules. Thus, the potential of the zinc ions
193   was constrained in the three dimensional conformational space in the vicinity of the Tbb DdRPII
194   RPB1 3D model. The ions were prepositioned in the 3D model of Tbb DdRPII RPB1, after
195   structural superposition to the template x-ray structure. The model was structurally optimized

196  and adjusted locally by subsequent energy minimizations, in an effort to eliminate any
197  molecular clashes and minimize the constrain energy. A radius of 6Å around each ion was given
198  full degrees of freedom during the abovementioned structural optimizations. Provided that the
199  Tbb DdRPII RPB1 is a nucleotide processing enzyme, whose structure coordinates a repertoire
200  of ions (e.g. Zinc, Mg++), the AMBER99 forcefield was selected. The AMBER99 forcefield is fully
201  parameterized for our biological system as it implements ff10 parameters for amino acids and
202  nucleic acids as well as EHT for small molecules, such as ions/cations at the same time (Vilar et
203  al. 2008). AM1-BCC charges were applied since the molecular system included the ion
204  molecules. The results of the molecular dynamics simulation were collected into a database by
205  MOE for further analysis.
206
207  **Model evaluation**
208  The produced models were initially evaluated within the MOE package by a residue packing
209  quality function, which depends on the number of buried non-polar side-chain groups and on
210  hydrogen bonding. Moreover, the suite PROCHECK (Laskowski et al. 1996) was employed to
211  further evaluate the quality of the produced models. Finally, MOE and its build in protein check
212  module was used to evaluate whether the models of DdRpII RPB1 domains are similar to known
213  protein structures of this family (Supplementary data 5, 6 and 7).
214
215  **Pharmacophore Elucidation**
216  A pharmacophoric feature characterizes a particular property and is not tied to a specific
217  chemical structure; indeed different chemical groups may share the same property and so be
218  represented by the same feature (Vlachakis et al. 2013c). It is thus a mistake to name as
219  pharmacophoric features chemical functionalities such as guanidines or sulfonamides or typical
220  structural skeletons such as flavones or steroids.
221  The term *pharmacophore modeling* refers to the generation of a pharmacophore hypothesis for
222  the binding interactions in a particular active site (Vlachakis et al. 2015). Several different
223  pharmacophore models for the same active site can be overlaid and reduced to their shared
224  features so that common interactions are retained. Such a *consensus pharmacophore* can be
225  considered as the largest common denominator shared by a set of active molecules.
226  In MOE, the computerized representation of a hypothesized pharmacophore is called a
227  *pharmacophore query*. A MOE pharmacophore query is a set of *query features* that are typically
228  created from ligand *annotation points*. Annotation points are markers in space that show the
229  location and type of biologically important atoms and groups, such as hydrogen donors and
230  acceptors, aromatic centers, projected positions of possible interaction partners or R-groups,
231  charged groups, and bioisosteres. The annotation points on a ligand are the potential locations
232  of the features that will constitute the pharmacophore query. Annotation points relevant to the
233  pharmacophore are converted into query features with the addition of an extra parameter: a
234  non-zero radius that encodes the permissible variation in the pharmacophore query's
235  geometry.
236  Once generated, a pharmacophore query can be used to screen virtual compound libraries for
237  novel ligands. Pharmacophore queries can also be used to filter conformer databases, e.g.
238  output from molecular docking runs, for biologically active conformations.
239

240 **Results**

241

242 **Phylogenetic Analysis**

243 In the present study, two phylogenetic analyses of DdRpII family proteins in all available
244 genomes, with putative full-length protein sequences were performed using two different
245 statistical methods from the Jalview and MEGA software. Based on findings, putative members
246 of the DdRpII family were identified in the *Animalia*, *Fungi*, *Plantae, Protista* and
247 C*hromalveolata* kingdom major eukaryotic taxonomic division, as well as viruses (Figure 1 and
248 Supplementary Figure 1). In our analyses, in agreement with previous reports (Smith et al.
249 1989), we found that DdRpII family is split into two main subunits the RPB1 and the RPB2. The
250 two subunits of the DdRpII family are clearly separated in the phylogenetic trees as two major
251 sub-trees were obtained for each one of them (Figure 1 and Supplementary Figure 1). The
252 monophyletic sub-tree of the RPB1 subunit contains the TBB DdRpII RPB1, as well as another 17
253 leaves, which are related to RPB1 subunit. Furthermore, in the phylogenetic trees, the TBB
254 DdRpII RBP1 forms a distinct monophyletic branch with the *Euplotes octocarinatus* DdRpII RPB1
255 and the *Plasmodium falciparum* DdRpII RPB1, which is basal to a clade that corresponds to
256 other parasites. The Newick format of the phylogenetic trees is provided (Supplementary Data
257 1 and 2).

258

259 **Conserved motifs exploration**

260 Multiple sequence alignment of the DdRpII subunit RPB1 protein sequences from a variety of
261 several species were included in the first sub-tree, highlights important conserved functional
262 domains as described previously by Janet L. Smith and Judith R. Levin (Smith et al. 1989). Good
263 conservation is evident throughout the whole length of the sequence, especially among species
264 that belong to the same taxonomic division (Figure 2).

265 In this study, an effort has been done to suggest motifs that were probably included in the
266 DdRpII of the subunit RPB1. Regions conserved across all species (eukaryotic and viruses) are
267 indicative of important functional domains of the DdRpII RPB1 enzyme. Finally, the consensus
268 sequence of the multiple sequence alignment highlights nine conserved motifs which are
269 conserved between all species. All of the conserved motifs identified here have not been
270 reported previously, and indisputably deserve further study (Figure 2, 3 and Supplementary
271 Figure 3). It is remarkable that all 18 polymerases, from the phylogenetic sub-tree of the
272 subunit RPB1, have high identity score and remain undamaged during the evolution (Figure 1
273 and 2). As we know, the highly conserved motifs in protein families are directly related to their
274 active sites and functionality (Koonin & Galperin 2003; Papageorgiou et al. 2016). Therefore,
275 the suggested conserved motifs were further studied in the TBB DdRiII 3D model.

276

277 **Structural models of the Trypanosoma Brucei DdRpII RPB1**

278 Homologous solved 3D structures from the Protein Data Bank (PDB) have been identified from
279 the Protein Data Bank (PDB) using the NCBI/BLASTp algorithm. Based on BLASTp report many
280 3D structures were determined suitable as templates for the homology modelling including the
281 crystal structure of the *Schizosaccharomyces Pombe* RNA Polymerase II subunit RPB1 (PDB:
282 3H0G) (Spahr et al. 2009), the crystal structure of the *Saccharomyces cerevisiae* RNA
283 Polymerase II subunit RPB1 (PDB: 4A3C and 1I3Q) (Cheung et al. 2011; Cramer et al. 2001), and

284  the electron microscopy structure of the *Human* RNA Polymerase II subunit RPB1 (PDB: 3J0K)
285  (Bernecky et al. 2011). The final choice of a template structure was not only based on the
286  percent sequence identity/similarity and the structure resolution, but also on the results of the
287  phylogenetic trees. The *Schizosaccharomyces Pombe* RNA Polymerase II subunit RPB1 is the
288  most suitable template for the design of the *Trypanosoma Brucei* DdRp II RPB1 (Figure 3 and
289  Supplementary Figure 3). Although the *Human* RNA Polymerase II RPB1 could also be used to
290  build the *Trypanosoma Brucei* DdRpII RPB1 3D model, it was avoided in an effort to minimize
291  potential toxicity issues during the drug design process. Nonetheless, the sequence of the
292  Human RNA Polymerase II and the corresponding sequence of the *Trypanosoma Brucei* were
293  aligned in an effort to identify sequence-based differences and/or similarities for the modelling
294  and drug design process (Supplementary Figure 3). A multiple sequence alignment was
295  constructed including the *Trypanosoma Brucei Brucei* DdRpII RPB1 (NCBI: P17545.1) (Das et al.
296  2006), the *Trypanosoma Brucei Gambiense* DdRpII RPB1(NCBI: XP_011773113.1) (Jackson et al.
297  2010), the crystal structure of *Schizosaccharomyces pombe* DdRpII RPB (PDB: 3H0G A chain)
298  (Spahr et al. 2009), the crystal structure of *Saccharomyces Cerevisiae* DdRpII RPB1 (PDB: 1I3Q A
299  chain) (Cramer et al. 2001)  and the crystal structure of Human DdRpII DdRpII RPB1 (PDB: 3J0K
300  A chain) (Bernecky et al. 2011) towards to identify all the suggested conserved motifs within
301  the  highlighted  domains  of  the  RPB1  and  the  major  sequences  differences  and
302  similarities(Supplementary Figure 3).
303  Sequence alignment of the *Trypanosoma Brucei* DdRpII RPB1 and *Schizosaccharomyces Pombe*
304  RNA Polymerase II RPB1 identified all the nine conserved motifs as expected (Figures 2 and 3).
305  The  sequence  alignment  between  the  *Trypanosoma  Brucei*  DdRpII  RPB1  and  the
306  *Schizosaccharomyces Pombe* RNA Polymerase II RPB1 (PDB 3H0G, resolution 3.65 Å) template
307  revealed 40% Identity and 56% similarity. These scores allow for a conventional homology
308  modelling approach to be considered (Figure 3). The model of TPP DdRpII was first structurally
309  superimposed and subsequently structurally compared to its template using the MOE software
310  (Figure 4). The TPP DdRpII model exhibited an alpha-carbon RMSD lower than 1.3 angstroms
311  (Figure 5 and Supplementary Data 7). Furthermore, the model was evaluated in regards to its
312  geometry and its compatibility with the template structure using the build in protein check
313  module of MOE (Supplementary Data 7). These results, confirmed the structural viability of the
314  3D *in silico* model.
315
316  **Discussion**
317
318  **Description of the *Trypanosoma Brucei  Brucei* DdRPII RPB1 model**
319  RNA Polymerase II is a multi-subunit enzyme that transcribes protein-coding genes in
320  eukaryotes (Sentenac 1985). Transcription in eukaryotes is dependent by three classes of
321  nuclear RNA polymerases I-III. The genes encoding the largest subunits of eukaryotic RNA
322  polymerases I, II and III have been isolated and are single copy genes, except *Trypanosoma* RNA
323  polymerase II which contain two alleles (Smith et al. 1989). Structural and sequence differences
324  between the two alleles are minor, but the C-terminal domain of those enzymes has a highly
325  unusual structure. TBB DdRpII RPB1 model is the first protein subunit of the ten subunits multi-
326  complex of RNA Polymerase II (Hahn 2004; Suh et al. 2013). The RPB1 subunit is very critical in
327  RNA polymerase formation and function. The RPB1 active site and the RPB2 hybrid-binding

328   region combine in a single fold that forms the active centre of the RpII (Figure 4). There are two
329   metal ions at the RNA polymerase II active site. It has been previously reported that a Mg metal
330   ion interacts with the three invariant aspartates of RPB1 (Cramer et al. 2001). The latter
331   aspartate residues, which were found in all RPB1 sequences were aligned and fitted in a motifs
332   exploration study. Consequently, those residues have now been marked as motif 4b in the TBB
333   DdRpII RPB1 structural model.

334         The swinging motion of the clamp dictates the degree of opening of the cleft in DdRpII
335   and permits the insertion of promoter DNA for the initiation of transcription (Suh et al. 2013).
336   Based on previous studies, it is established that, upon closure of a transcribing complex, the
337   RPB1 clamp serves as a multi-functional tool, sensing the DNA/RNA hybrid conformation and
338   splitting DNA and RNA strands at the upstream end of the transcription complex (Cramer et al.
339   2001). The clamp is formed by N- and C-terminal regions of RPB1 and a part of the C-terminal
340   region of RPB2 (Chen et al. 2007; Hahn 2004; Li et al. 2014). The clamp is primarily stabilized by
341   three Zn ions within the RPB1 subunit (also marked in the TPP DdRpII RPB1) which forms zinc –
342   finger conformations; two within the "clamp core" and one in the "clamp head". Accordingly,
343   two Zinc-finger formations were identified and highlighted in the TBB DdRpII RPB1 model
344   (Figure 6). The first formation can be recognized between a Zn ion and four cysteine residues in
345   the suggested motif 1a, also known as CX(2)CXnCX2C/H (Das et al. 2006) (Figure 3).  Mutations
346   in the first Zn-finger formation confer a lethal phenotype of RNA polymerase II (Donaldson &
347   Friesen 2000). The second Zinc –finger can be recognized in the next four cysteine residues
348   (Figure 3 and 6). In the proposed motif 1b, the first two cysteine residues were identified, which
349   constitute part of the second Zing finger formation. Finally, according to our molecular
350   dynamics simulations, the main role of the Rpb1 and Rpb2 subunits is to provide stability within
351   the overall structure formation of the RNA polymerase II molecule in the 3D space.
352
**353   3D Pharmacophore Elucidation**
354
355   3D Pharmacophore design techniques take into account both the three-dimensional structures
356   and binding modes of receptors and inhibitors towards identifying regions that are favorable or
357   not for a particular receptor-inhibitor interaction (Vlachakis & Kossida 2013). The description of
358   the receptor-inhibitor interaction pattern is determined through a correlation between the
359   specific properties of the inhibitors and their action on enzymatic activity (Balatsos et al. 2009;
360   Vlachakis et al. 2012). The pharmacophore for TBB DdRpII RPB1 (Figure 7) was based on
361   structural information from the enzyme's catalytic site including all steric and electronic
362   features that are necessary to ensure optimal non-covalent interactions. The pharmacophoric
363   features were investigated including positively or negatively ionized regions, hydrogen bond
364   donors and acceptors, aromatic regions and hydrophobic areas. Firstly, there should be one
365   electron-donating group in the proximity of the Ser1172 (colored green). The electron-donating
366   region indicates a particular property of the inhibitor and is not necessarily confined to a
367   specific chemical structure. Moreover, this interaction site may not strictly represent a
368   hydrogen bond, but water or ion mediated bridges since the distance from the catalytic amino
369   acids varies between 3-9 Å. An aromatic PAP (colored orange) was positioned in the proximity
370   of Phe1179, which established pi-stacking interactions. Two electron accepting PAPs (colored
371   red) were positioned in the proximity of the two Arginine residues (Arg1171 and Arg1203).

372    Finally, a set of two adjacent PAPs were positioned in the center of the active site, where the
373    Zn++ is coordinated in the crystal structure. Those yellow-colored PAPs are indicative of S-S
374    bonds and bridges or even S-C interactions, following the Michael acceptor moiety pattern. The
375    surrounding Cysteines are Cys1173, Cys1155, Cys1152, and Cys1270. However, the most
376    important factor of the latter PAPs was the optimal positioning of these groups in the 3D
377    conformational space of the TBB DdRpII RPB1 active site, rather than the amount of
378    conjugation or interaction with the protein.
379

380    **Conclusion**
381    The *Trypanosoma Brucei Brucei* DdRpII RPB1 enzyme was evolutionary analyzed, and nine new
382    conserved motifs were identified. Using the X-ray crystal structure of the *Schizosaccharomyces*
383    *Pombe* DdRpII RPB1, the 3D model of the *Trypanosoma Brucei Brucei* DdRpII RPB1 was designed
384    using homology modelling techniques. The model was *in silico* evaluated and displayed high
385    conservation of the functional domains previously reported in other DdRpII subunit RPB1
386    species. The *Trypanosoma Brucei Brucei* DdRpII RPB1 model structure provides a basis for
387    interpretation of available data and the design of new experiments towards the *Trypanosoma*
388    *Brucei Brucei* inhibition. We, therefore, propose the use of the *Trypanosoma Brucei Brucei*
389    DdRpII RPB1 model as a pharmacological targeting platform for advanced, *in silico* drug design
390    experiments using the novel findings of this study, both in the sequence and structural level.
391    The 3D models and sequence datasets that derived from this study will be made available to
392    the public, in an effort to pave the way for fellow scientists of multidiscipline backgrounds to
393    word in a synergic way towards the designing of novel anti-malarial agents with improved
394    biochemical and clinical characteristics in the future.
395

396    **Abbreviations**
397    DdRpII        DNA-dependent RNA polymerase II
398    TBB            Trypanosoma brucei brucei
399    TBG            Trypanosoma brucei gambiense
400    TBR            Trypanosoma brucei rhodesiense
401    MOE           Molecular Operating Environment

403 **Figures and Data legend**

404

405 **Figure 1: Phylogenetic reconstruction of *Trypanosoma Brucei Brucei* DdRpII RPB1 protein**
406 **sequences.** The tree was generated using the DdRpII family dataset (36 foul length protein
407 sequences samples). The tree was constructed by Matlab Bioinformatics Toolbox utilizing
408 Neighbour – Joining statistical method for 100 bootstrap replicates and visualized using MEGA
409 cycle option. In the tree representation there are clearly separated in two monophyletic
410 branches the RNA polymerases II subunits RPB1 (colored green) and RPB2 (colored blue).
411 *Trypanosoma Brucei* DdRpII RPB1 protein sequence was correctly classified and separated in
412 the monophyletic sub-tree of the RPB1 group (highlight with red dots).

413

414

415 **Figure 2: Representative conserved motifs for the DdRpII subunit RPB1.** The nine suggested
416 conserved motifs were extracted based on the multiple sequence alignment of the 18 protein
417 sequences were classified and clearly separated in the DdRpII subunit RPB1 monophyletic sub-
418 tree. The conserved motifs were identified through the consensus sequence and logo graph
419 where generated using Jalview software.

420

421 **Figure 3: Sequence alignment between the *Trypanosoma Brucei Brucei* DdRpII RPB1 and the**
422 **corresponding sequence of the crystal structure of the *Schizosaccharomyces Pombe* DdRpII**
423 **RPB1**. **(A)** Alignment of DdRpII RPB1 from *Trypanosoma Brucei* DdRpII RPB1 (Labeled as "TB")
424 with *Schizosaccharomyces Pombe* DdRpII RPB1 (Labeled as "SB") was initially carried out with
425 BLASTp and then manually adjusted. The nine suggested conserved motifs (Motifs 1a, 1b, 2, 3a,
426 3b, 3c, 4a, 4b, 4c) based on figure 2, domains and domain-like regions of *Trypanosoma Brucei*
427 DdRpII RPB1 represented in different colours. The amino acid residue numbers at the domain
428 boundaries are indicated. Important structural elements and prominent regions involved in
429 subunit interactions are also noted. Residues involved in the Zn and Mg coordination are
430 highlighted in blue. **(B)** Domains and domain-like regions of the DdRpII subunit Rpb1. The
431 amino acid residue numbers at the domain boundaries are indicated.

432

433 **Figure 4: Model of the *Trypanosoma Brucei Brucei* DdRPII RPB1**. **(A and B)** Ribbon
434 representation of the produced *Trypanosoma Brucei Brucei DdRPII RPB1* model (colored
435 Orange) superposed with the corresponding *Schizosaccharomyces pombe* DdRpII RPB1 (in
436 purple). **(C and D)** The nine suggested conserved motifs and the domains and domain-like
437 regions of the *Trypanosoma Brucei Brucei DdRPII RPB1. The motifs and RPB1 domains have*
438 *been color-coded according* to the Figures 2 and 3, and are shown in CPK format (Usual space
439 filling). **(E and F)** Electrostatic surface potential for the *Trypanosoma Brucei Brucei DdRPII RPB1*.
440 Represented with blue is the area of negative charge. Red is the area of positive charge and
441 white is the un-charged region.

442

443 **Figure 5: Structural superposition of the TBB DdRPII RPB1 model and the**
444 ***Schizosaccharomyces pombe* crystal structural (template).** The produced *Trypanosoma Brucei*
445 *Brucei DdRPII RPB1* molecular surface (colored Orange) next to the corresponding

446  *Schizosaccharomyces pombe* DdRpII RPB1 molecular surface (in purple). The two molecular
447  surfaces are highly conserved in their's active sites with few differences in the outer layer.
448
449  **Figure 6: Zinc-finger formations in the *Trypanosoma Brucei Brucei* DdRpII RPB1 model.** Ribbon
450  representation of the produced *Trypanosoma Brucei Brucei* DdRPII RPB1 model. In the
451  produced model were highlighted 3 main zing-finger domain formations (colored grey) were
452  contained in the clam core, clam head and active site region.  Domains and domain-like regions
453  of the *Trypanosoma Brucei Brucei* DdRPII RPB1 *have been color-coded according* to conventions
454  of Figures 3.
455
456  **Figure 7: The 3D pharmacophore model for the *Trypanosoma Brucei Brucei* DdRPII RPB1**
457  **model.** In total 5 distinct pharmacophoric features were identified. An aromatic region (colored
458  orange), an electron donating region (colored green), two electron accepting regions (colored
459  red) and a sulphur specific S-S interacting region (colored yellow).
460
461  **Supplementary Figure 1: Phylogenetic reconstruction of *Trypanosoma Brucei Brucei* DdRPII**
462  **RPB1 model DdRpII RPB1 protein sequences.** The tree was generated using the DdRpII family
463  dataset (36 foul length protein sequences samples) and the Jalview software. Tree was
464  constructed using the average distance statistical method with PAM 250. In the tree
465  representation there are clearly shown the two RNA polymerases II subunits RPB1 and RPB2 as
466  two main monophyletic sub-trees. *Trypanosoma Brucei* DdRpII RPB1 protein sequence was
467  correctly classified in the monophyletic sub-tree of the RPB1 group.
468
469  **Supplementary Figure 2: Molecular dynamics graph.**  The energy (Kcal/mol) vs time (ns) plot of
470  the 100ns simulation trajectory of the TBB DdRpII RPBI model.
471
472  **Supplementary Figure 3: Multiple sequence alignment.** The alignment was performed using
473  the *Trypanosoma Brucei Brucei DdRPII RPB1*, the *Trypanosoma Brucei Gambiense* DdRpII RPB1,
474  the crystal structure of *Schizosaccharomyces pombe* DdRpII RPB, the crystal structure of
475  *Saccharomyces Cerevisiae* DdRpII RPB1 and the crystal structure of Human DdRpII DdRpII RPB1.
476  **(A)** All nine suggested conserved motifs and major domains of DdRpII RPB1 have been marked
477  (Motifs 1a, 1b, 2, 3a, 3b, 3c, 4a, 4b, 4c). Additionally, in the multiple sequence alignment were
478  presented the major differences.  **(B)**  Domains and domainlike regions of the DdRpII subunit
479  Rpb1. The amino acid residue numbers at the domain boundaries are indicated.
480
481
482  **Supplementary Data 1: DdRPII related proteins dataset.**
483
484  **Supplementary Data 2: MEGA software phylogenetic tree in newick format.** The tree was
485  constructed the Neighbour – Joining statistical method for 100 bootstrap replicates and the 36
486  extracted samples of the DpRpII.
487

488   **Supplementary Data 3: Jalview software phylogenetic tree in newick format.** The tree was
489   constructed using the average distances statistical method and the 36 extracted samples of the
490   DpRpII.
491
492   **Supplementary Data 4: *Trypanosoma Brucei Brucei* DdRPII RPB1 model in .pdb format.**
493
494   **Supplementary Data 5: Protein structure report of the template.**
495
496   **Supplementary Data 6: Protein structure report of the model.**
497
498   **Supplementary Data 7: Protein structure report of the superposed model and template.**
499
500

501 **References**

502

503 Balatsos NA, Vlachakis D, Maragozidis P, Manta S, Anastasakis D, Kyritsis A, Vlassi M, Komiotis D, and
504         Stathopoulos C. 2009. Competitive inhibition of human poly(A)-specific ribonuclease (PARN) by
505         synthetic fluoro-pyranosyl nucleosides. *Biochemistry* 48:6044-6051. 10.1021/bi900236k
506 Bayele HK. 2009. Trypanosoma brucei: a putative RNA polymerase II promoter. *Exp Parasitol* 123:313-
507         318. 10.1016/j.exppara.2009.08.007
508 S0014-4894(09)00233-1 [pii]
509 Berg JM, Tymoczko JL, and Stryer L. 2002. *Biochemistry*. New York: W.H. Freeman.
510 Bernecky C, Grob P, Ebmeier CC, Nogales E, and Taatjes DJ. 2011. Molecular architecture of the human
511         Mediator-RNA      polymerase      II-TFIIF      assembly.      *PLoS      Biol*      9:e1000603.
512         10.1371/journal.pbio.1000603
513 Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E,
514         Hamlin NE, Haas B, Bohme U, Hannick L, Aslett MA, Shallom J, Marcello L, Hou L, Wickstead B,
515         Alsmark UC, Arrowsmith C, Atkin RJ, Barron AJ, Bringaud F, Brooks K, Carrington M, Cherevach I,
516         Chillingworth TJ, Churcher C, Clark LN, Corton CH, Cronin A, Davies RM, Doggett J, Djikeng A,
517         Feldblyum T, Field MC, Fraser A, Goodhead I, Hance Z, Harper D, Harris BR, Hauser H, Hostetler
518         J, Ivens A, Jagels K, Johnson D, Johnson J, Jones K, Kerhornou AX, Koo H, Larke N, Landfear S,
519         Larkin C, Leech V, Line A, Lord A, Macleod A, Mooney PJ, Moule S, Martin DM, Morgan GW,
520         Mungall K, Norbertczak H, Ormond D, Pai G, Peacock CS, Peterson J, Quail MA, Rabbinowitsch E,
521         Rajandream MA, Reitter C, Salzberg SL, Sanders M, Schobel S, Sharp S, Simmonds M, Simpson
522         AJ, Tallon L, Turner CM, Tait A, Tivey AR, Van Aken S, Walker D, Wanless D, Wang S, White B,
523         White O, Whitehead S, Woodward J, Wortman J, Adams MD, Embley TM, Gull K, Ullu E, Barry JD,
524         Fairlamb AH, Opperdoes F, Barrell BG, Donelson JE, Hall N, Fraser CM, Melville SE, and El-Sayed
525         NM. 2005. The genome of the African trypanosome Trypanosoma brucei. *Science* 309:416-422.
526         10.1126/science.1112642
527 Chakravorty DK, and Merz KM, Jr. 2014. Studying allosteric regulation in metal sensor proteins using
528         computational      methods.      *Adv      Protein      Chem      Struct      Biol*      96:181-218.
529         10.1016/bs.apcsb.2014.06.009
530 S1876-1623(14)00010-8 [pii]
531 Chen HT, Warfield L, and Hahn S. 2007. The positions of TFIIF and TFIIE in the RNA polymerase II
532         transcription preinitiation complex. *Nat Struct Mol Biol* 14:696-703. nsmb1272 [pii]
533 10.1038/nsmb1272
534 Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, and Thompson JD. 2003. Multiple
535         sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31:3497-3500.
536 Cheung AC, Sainsbury S, and Cramer P. 2011. Structural basis of initial RNA polymerase II transcription.
537         *EMBO J* 30:4755-4763. 10.1038/emboj.2011.396
538 emboj2011396 [pii]
539 Chung HM, Lee MG, Dietrich P, Huang J, and Van der Ploeg LH. 1993. Disruption of largest subunit RNA
540         polymerase II genes in Trypanosoma brucei. *Mol Cell Biol* 13:3734-3743.
541 Cramer P, Bushnell DA, and Kornberg RD. 2001. Structural basis of transcription: RNA polymerase II at
542         2.8 angstrom resolution. *Science* 292:1863-1876. 10.1126/science.1059493
543 1059493 [pii]
544 Darriba D, Taboada GL, Doallo R, and Posada D. 2011. ProtTest 3: fast selection of best-fit models of
545         protein evolution. *Bioinformatics* 27:1164-1165. 10.1093/bioinformatics/btr088
546 btr088 [pii]
547 Das A, Li H, Liu T, and Bellofatto V. 2006. Biochemical characterization of Trypanosoma brucei RNA
548         polymerase II. *Mol Biochem Parasitol* 150:201-210. S0166-6851(06)00232-5 [pii]

549    10.1016/j.molbiopara.2006.08.002

550    Donaldson IM, and Friesen JD. 2000. Zinc stoichiometry of yeast RNA polymerase II and characterization
551        of mutations in the zinc-binding domain of the largest subunit. *J Biol Chem* 275:13780-13788.
552        275/18/13780 [pii]

553    Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic*
554        *Acids Res* 32:1792-1797. 10.1093/nar/gkh340

555    32/5/1792 [pii]

556    Evers R, Hammer A, Kock J, Jess W, Borst P, Memet S, and Cornelissen AW. 1989. Trypanosoma brucei
557        contains two RNA polymerase II largest subunit genes with an altered C-terminal domain. *Cell*
558        56:585-597. 0092-8674(89)90581-3 [pii]

559    Hahn S. 2004. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct*
560        *Mol Biol* 11:394-403. 10.1038/nsmb763

561    nsmb763 [pii]

562    Illergard K, Ardell DH, and Elofsson A. 2009. Structure is three to ten times more conserved than
563        sequence--a study of structural response in protein cores. *Proteins* 77:499-508.
564        10.1002/prot.22458

565    Jackson AP, Sanders M, Berry A, McQuillan J, Aslett MA, Quail MA, Chukualim B, Capewell P, MacLeod A,
566        Melville SE, Gibson W, Barry JD, Berriman M, and Hertz-Fowler C. 2010. The genome sequence
567        of Trypanosoma brucei gambiense, causative agent of chronic human african trypanosomiasis.
568        *PLoS Negl Trop Dis* 4:e658. 10.1371/journal.pntd.0000658

569    Koch H, Raabe M, Urlaub H, Bindereif A, and Preusser C. 2016. The polyadenylation complex of
570        Trypanosoma brucei: Characterization of the functional poly(A) polymerase. *RNA Biol* 13:221-
571        231. 10.1080/15476286.2015.1130208

572    Koonin EV, and Galperin MY. 2003. *Sequence - evolution - function : computational approaches in*
573        *comparative genomics*. Boston: Kluwer Academic.

574    Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, and Thornton JM. 1996. AQUA and PROCHECK-
575        NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR*
576        8:477-486.

577    Li W, Giles C, and Li S. 2014. Insights into how Spt5 functions in transcription elongation and repressing
578        transcription coupled DNA repair. *Nucleic Acids Res* 42:7069-7083. 10.1093/nar/gku333

579    gku333 [pii]

580    Loukatou S, Papageorgiou L, Fakourelis P, Filntisi A, Polychronidou E, Bassis I, Megalooikonomou V,
581        Makalowski W, Vlachakis D, and Kossida S. 2014. Molecular dynamics simulations through GPU
582        video games technologies. *J Mol Biochem* 3:64-71.

583    Malvy D, and Chappuis F. 2011. Sleeping sickness. *Clin Microbiol Infect* 17:986-995. 10.1111/j.1469-
584        0691.2011.03536.x

585    S1198-743X(14)61376-8 [pii]

586    Nayeem A, Sitkoff D, and Krystek S, Jr. 2006. A comparative study of available software for high-accuracy
587        homology modeling: from sequence alignments to structural models. *Protein Sci* 15:808-824.
588        10.1110/ps.051892906

589    Papageorgiou L, Loukatou S, Koumandou VL, Makalowski W, Megalooikonomou V, Vlachakis D, and
590        Kossida S. 2014a. Structural models for the design of novel antiviral agents against Greek Goat
591        Encephalitis. *PeerJ* 2:e664. 10.7717/peerj.664

592    Papageorgiou L, Loukatou S, Sofia K, Maroulis D, and Vlachakis D. 2016. An updated evolutionary study
593        of Flaviviridae NS3 helicase and NS5 RNA-dependent RNA polymerase reveals novel invariable
594        motifs as potential pharmacological targets. *Mol Biosyst*. 10.1039/c5mb00706b

595    Papageorgiou L, Vlachakis D, Koumandou VL, Papangelopoulos N, and Kossida S. 2014b. Computer-
596        Aided Drug Design and Biological Evaluation of Novel Anti-Greek Goat Encephalitis Agents.

597     *International Journal of Systems Biology and Biomedical Technologies* 2.
598         10.4018/ijsbbt.2013100101
599     Ridley RG. 2002. Medical need, scientific opportunity and the drive for antimalarial drugs. *Nature*
600         415:686-693. 10.1038/415686a
601     415686a [pii]
602     Ross L, Lim ML, Liao Q, Wine B, Rodriguez AE, Weinberg W, and Shaefer M. 2007. Prevalence of
603         antiretroviral drug resistance and resistance-associated mutations in antiretroviral therapy-
604         naive HIV-infected individuals from 40 United States cities. *HIV Clin Trials* 8:1-8.
605         L61238775J512643 [pii]
606     10.1310/hct0801-1
607     Seeliger D, and de Groot BL. 2010. Ligand docking and binding site analysis with PyMOL and
608         Autodock/Vina. *J Comput Aided Mol Des* 24:417-422. 10.1007/s10822-010-9352-6
609     Sentenac A. 1985. Eukaryotic RNA polymerases. *CRC Crit Rev Biochem* 18:31-90.
610     Smith JL, Levin JR, Ingles CJ, and Agabian N. 1989. In trypanosomes the homolog of the largest subunit of
611         RNA polymerase II is encoded by two genes and has a highly unusual C-terminal domain
612         structure. *Cell* 56:815-827.
613     Spahr H, Calero G, Bushnell DA, and Kornberg RD. 2009. Schizosacharomyces pombe RNA polymerase II
614         at 3.6-A resolution. *Proc Natl Acad Sci U S A* 106:9185-9190. 10.1073/pnas.0903361106
615     Stecher G, Liu L, Sanderford M, Peterson D, Tamura K, and Kumar S. 2014. MEGA-MD: molecular
616         evolutionary genetics analysis software with mutational diagnosis of amino acid variation.
617         *Bioinformatics* 30:1305-1307. 10.1093/bioinformatics/btu018
618     btu018 [pii]
619     Suh H, Hazelbaker DZ, Soares LM, and Buratowski S. 2013. The C-terminal domain of Rpb1 functions on
620         other RNA polymerase II subunits. *Mol Cell* 51:850-858. 10.1016/j.molcel.2013.08.015
621     S1097-2765(13)00588-1 [pii]
622     Temiz AN, Benos PV, and Camacho CJ. 2010. Electrostatic hot spot on DNA-binding domains mediates
623         phosphate desolvation and the pre-organization of specificity determinant side chains. *Nucleic
624         Acids Res* 38:2134-2144. 10.1093/nar/gkp1132
625     gkp1132 [pii]
626     Thompson JD, Higgins DG, and Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive
627         multiple sequence alignment through sequence weighting, position-specific gap penalties and
628         weight matrix choice. *Nucleic Acids Res* 22:4673-4680.
629     Trouiller P, Olliaro P, Torreele E, Orbinski J, Laing R, and Ford N. 2002. Drug development for neglected
630         diseases: a deficient market and a public-health policy failure. *Lancet* 359:2188-2194. S0140-
631         6736(02)09096-7 [pii]
632     10.1016/S0140-6736(02)09096-7
633     Vilar S, Cozza G, and Moro S. 2008. Medicinal chemistry and the molecular operating environment
634         (MOE): application of QSAR and molecular docking to drug discovery. *Curr Top Med Chem*
635         8:1555-1572.
636     Vlachakis D. 2009. Theoretical study of the Usutu virus helicase 3D structure, by means of computer-
637         aided homology modelling. *Theor Biol Med Model* 6:9. 10.1186/1742-4682-6-9
638     1742-4682-6-9 [pii]
639     Vlachakis D, Bencurova E, Papangelopoulos N, and Kossida S. 2014a. Current state-of-the-art molecular
640         dynamics methods and applications. *Adv Protein Chem Struct Biol* 94:269-313. 10.1016/B978-0-
641         12-800168-4.00007-X
642     B978-0-12-800168-4.00007-X [pii]
643     Vlachakis D, Fakourelis P, Megalooikonomou V, Makris C, and Kossida S. 2015. DrugOn: a fully integrated
644         pharmacophore modeling and structure optimization toolkit. *PeerJ* 3:e725. 10.7717/peerj.725

645   725 [pii]

646   Vlachakis D, Kontopoulos DG, and Kossida S. 2013a. Space constrained homology modelling: the
647          paradigm of the RNA-dependent RNA polymerase of dengue (type II) virus. *Comput Math*
648          *Methods Med* 2013:108910. 10.1155/2013/108910

649   Vlachakis D, and Kossida S. 2013. Molecular modeling and pharmacophore elucidation study of the
650          Classical Swine Fever virus helicase as a promising pharmacological target. *PeerJ* 1:e85.
651          10.7717/peerj.85

652   Vlachakis D, Koumandou VL, and Kossida S. 2013b. A holistic evolutionary and structural study of
653          flaviviridae provides insights into the function and inhibition of HCV helicase. *PeerJ* 1:e74.
654          10.7717/peerj.74

655   Vlachakis D, Pavlopoulou A, Roubelakis MG, Feidakis C, Anagnou NP, and Kossida S. 2014b. 3D molecular
656          modeling and evolutionary study of the Trypanosoma brucei DNA Topoisomerase IB, as a new
657          emerging pharmacological target. *Genomics* 103:107-113. 10.1016/j.ygeno.2013.11.008

658   S0888-7543(13)00224-3 [pii]

659   Vlachakis D, Pavlopoulou A, Tsiliki G, Komiotis D, Stathopoulos C, Balatsos NA, and Kossida S. 2012. An
660          integrated in silico approach to design specific inhibitors targeting human poly(a)-specific
661          ribonuclease. *PLoS One* 7:e51113. 10.1371/journal.pone.0051113

662   PONE-D-12-20013 [pii]

663   Vlachakis D, Tsagrasoulis D, Megalooikonomou V, and Kossida S. 2013c. Introducing Drugster: a
664          comprehensive and fully integrated drug design, lead and structure optimization toolkit.
665          *Bioinformatics* 29:126-128. 10.1093/bioinformatics/bts637

666   bts637 [pii]

667   Wang J, Cieplak P, and Kollman P. 2000. How well does a restrained electrostatic potential (resp) model
668          perform in calculating conformational energies of organic and biological molecules. *J Comp*
669          *Chem* 21:1049-1071.

670   Waterhouse AM, Procter JB, Martin DM, Clamp M, and Barton GJ. 2009. Jalview Version 2--a multiple
671          sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189-1191.
672          10.1093/bioinformatics/btp033

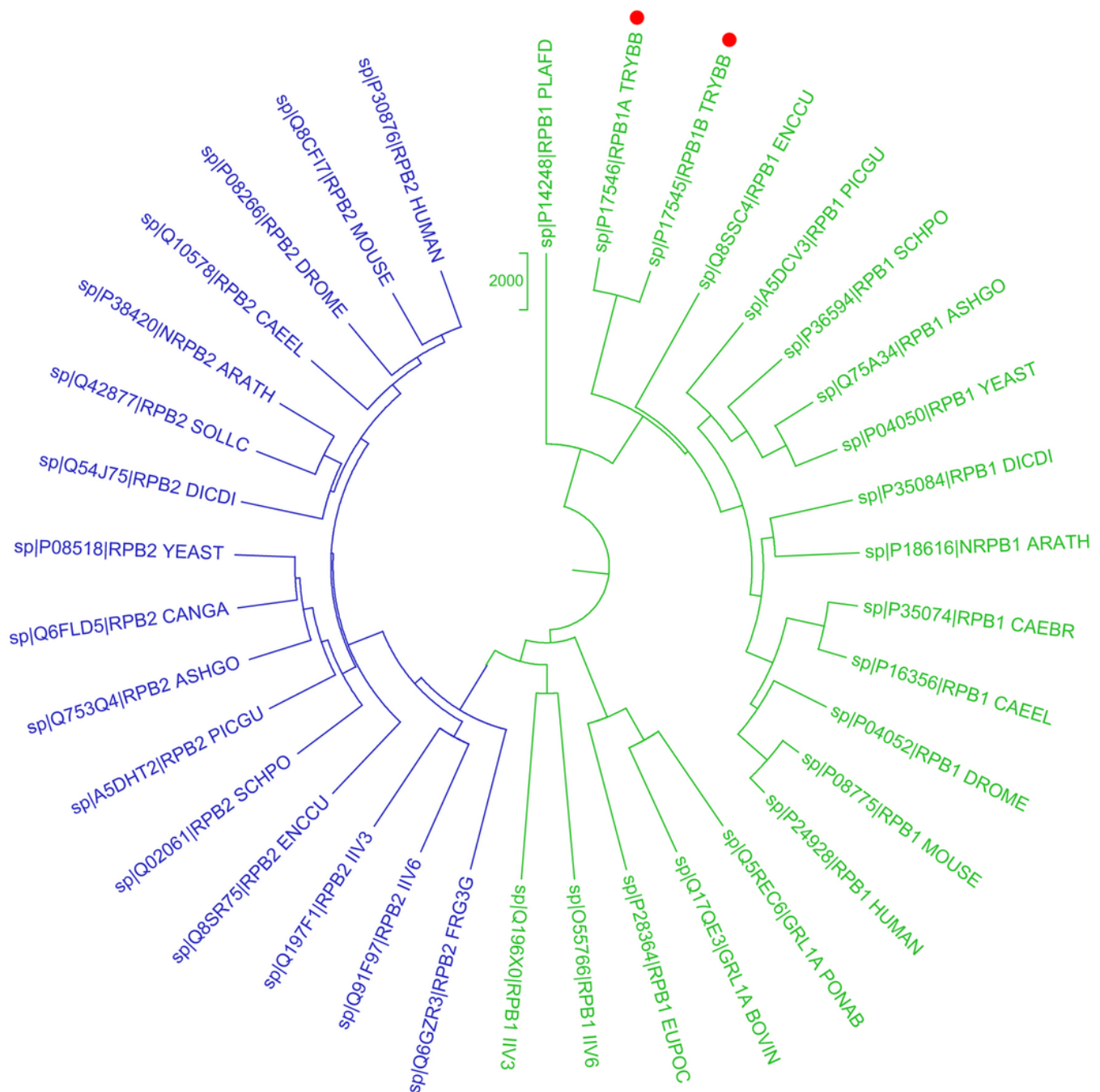
673   btp033 [pii]

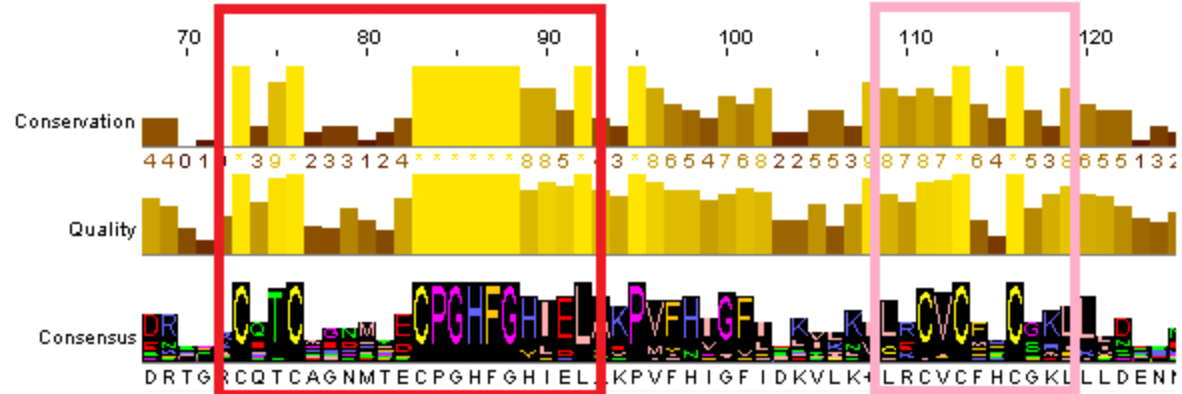
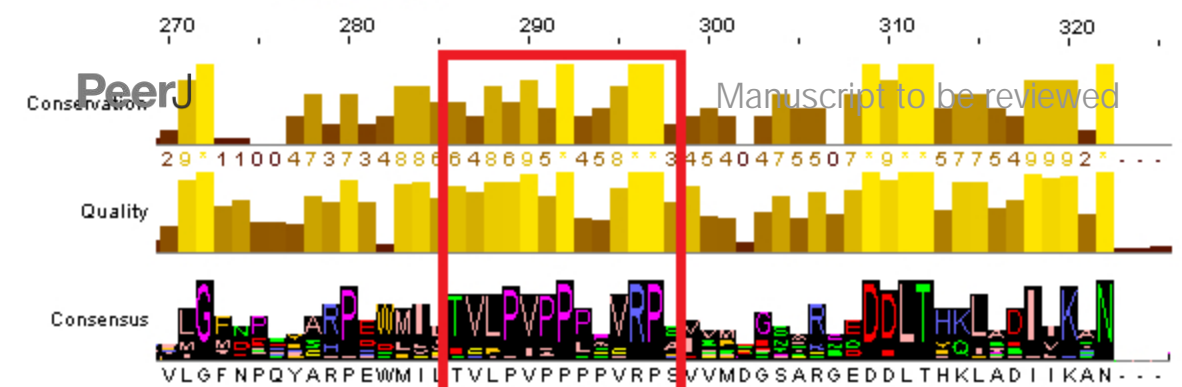
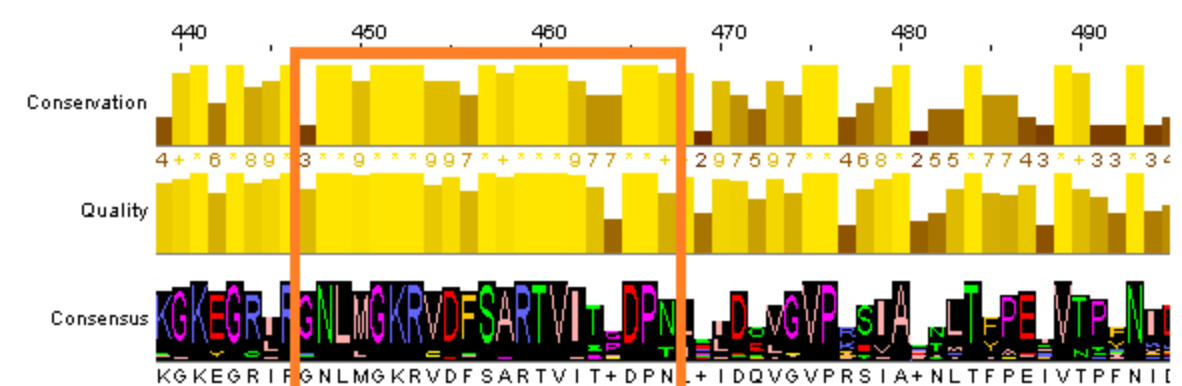
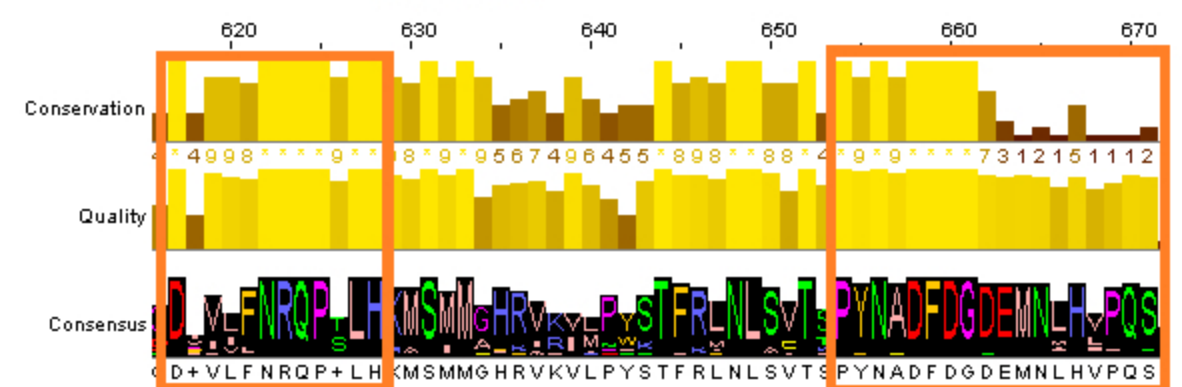
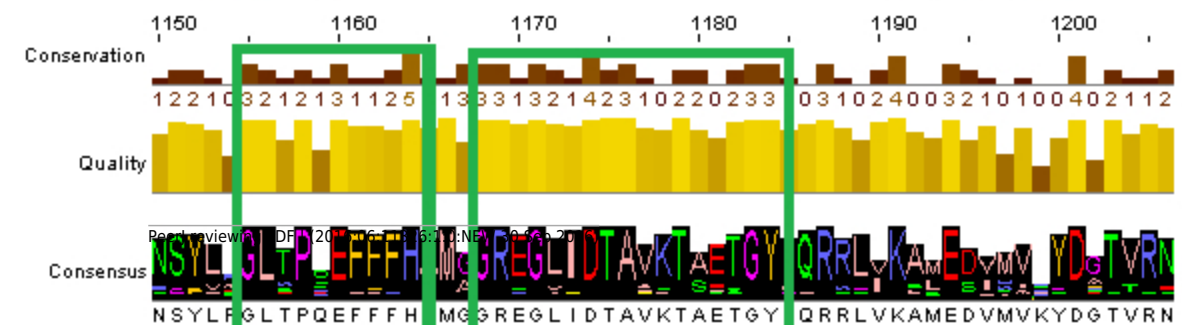
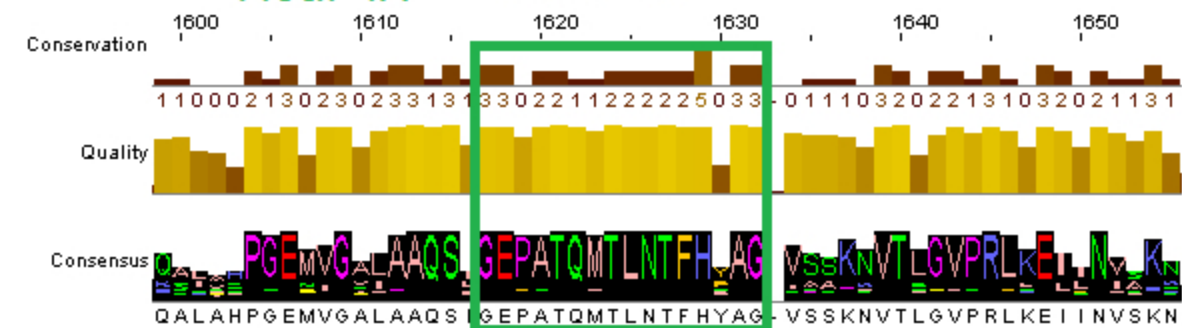
674   World Health Organization. 2015. Human African trypanosomiasis

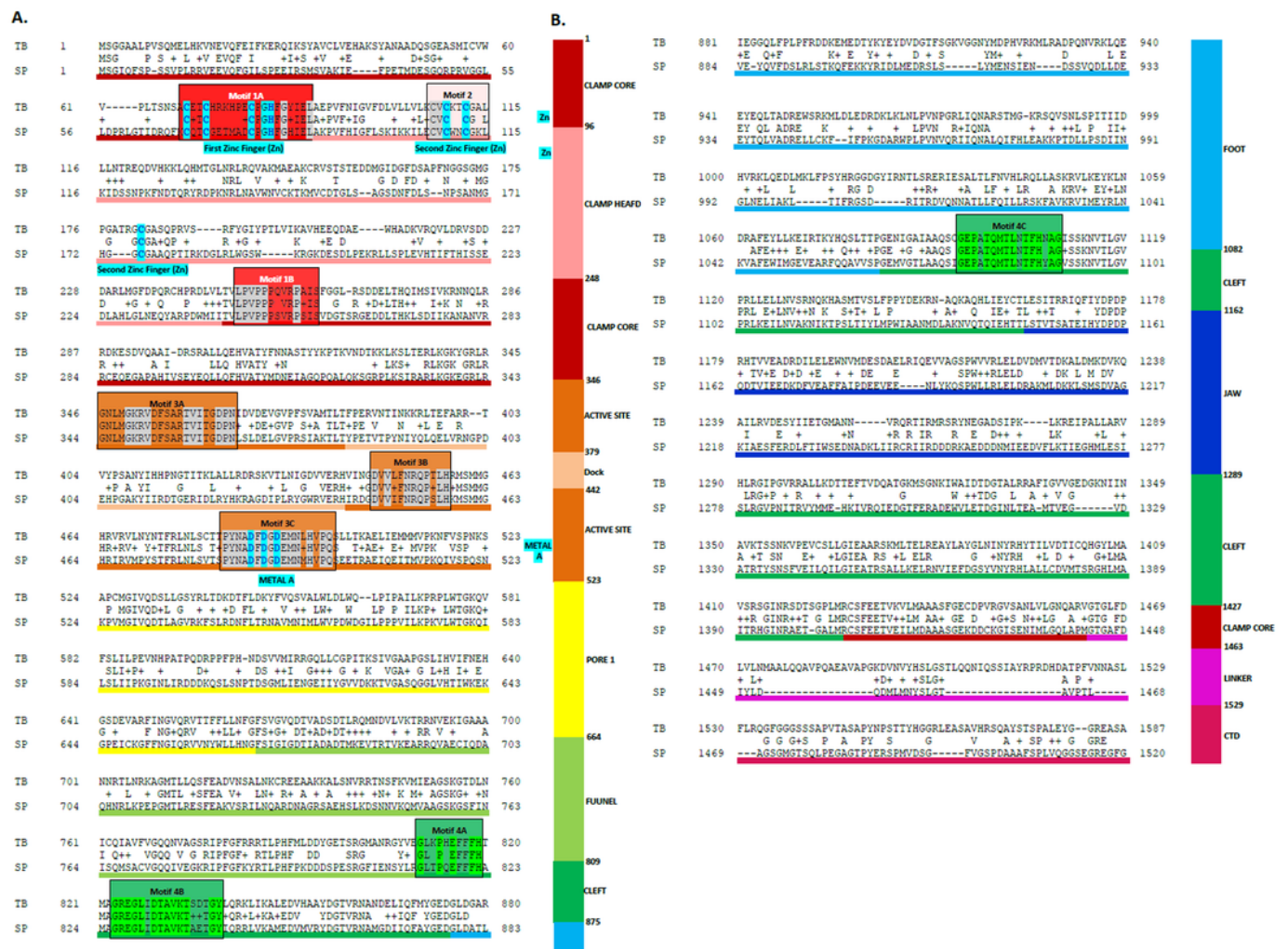
675

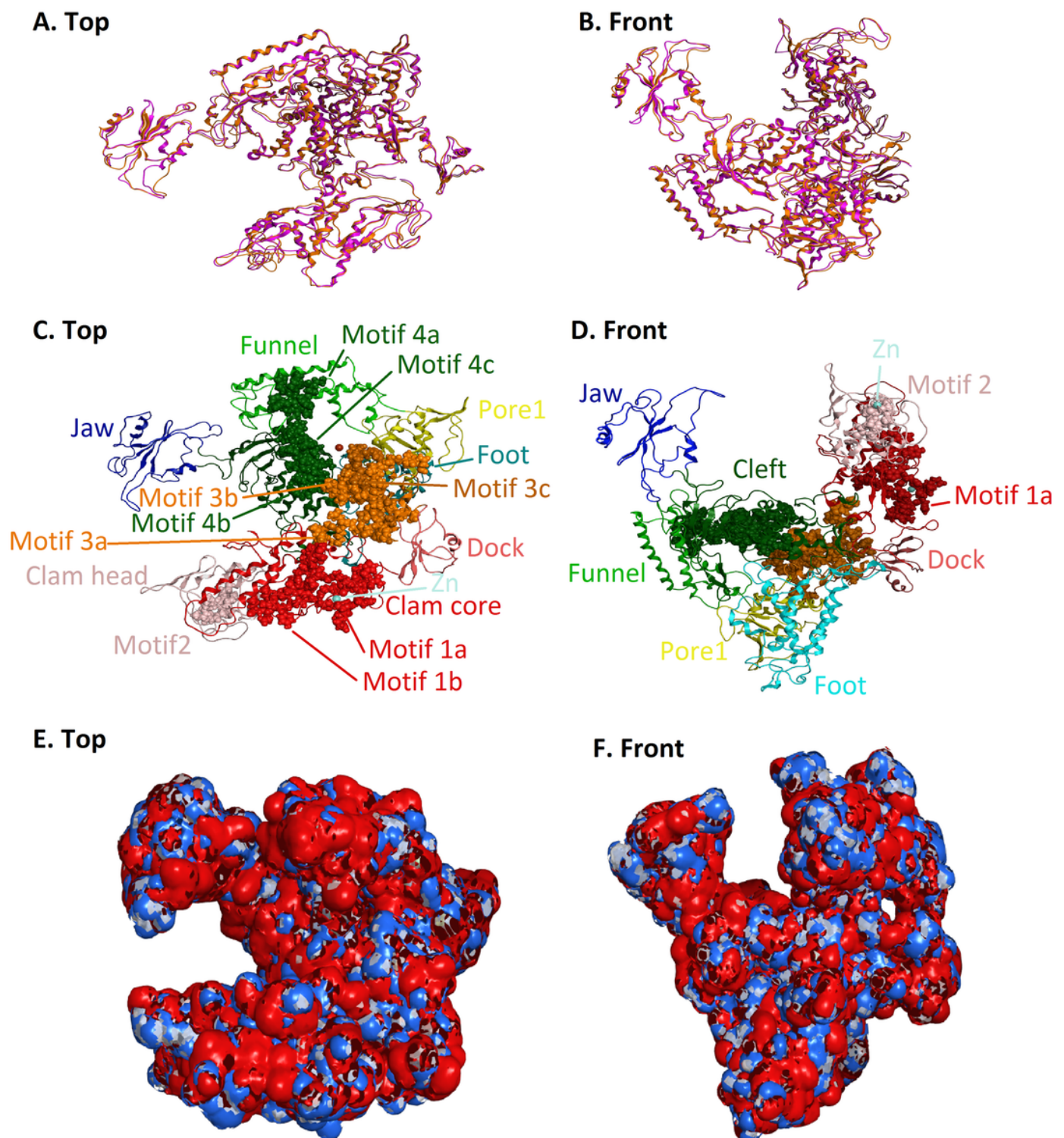
676

# Figure 1

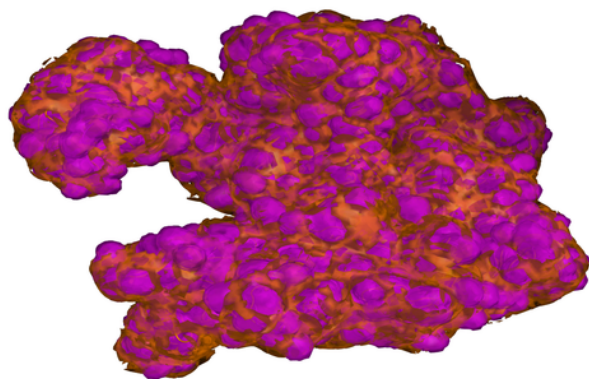
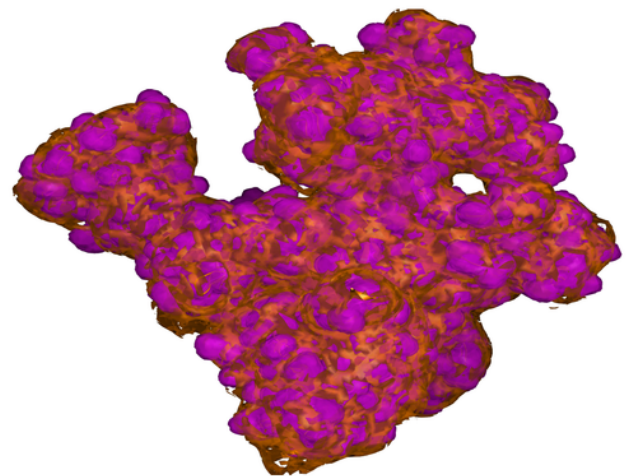
**Phylogenetic reconstruction of** *Trypanosoma Brucei Brucei* **DdRpII RPB1 protein sequences.**

**Representative conserved motifs for the DdRpIIsubunit RPB1.**

The nine suggested conserved motifs were extracted based on the multiple sequence alignment of the 18 protein sequences were classified and clearly separated in the DdRpII subunit RPB1 monophyletic sub-tree. The conserved motifs were identified through the consensus sequence and logo graph where generated using Jalview software.

# Figure 3

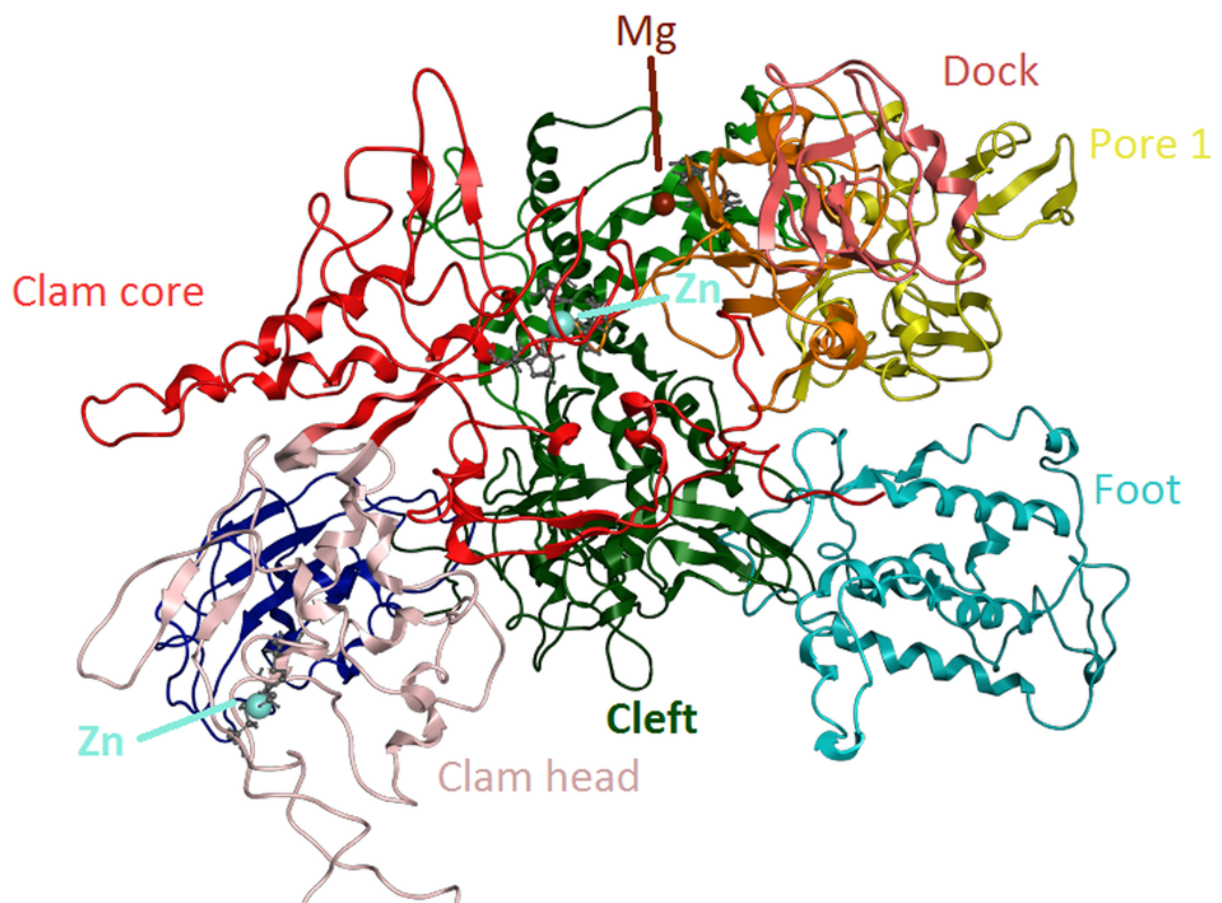
**Sequencealignment between the** *Trypanosoma Brucei Brucei* **DdRpIIRPB1 and the corresponding sequence of the crystal structureof the** *Schizosaccharomyces Pombe* **DdRpII RPB1**

# Figure 4

**Model of the _TrypanosomaBrucei Brucei_ DdRPII RPB1** .

# Figure 5

**Structural superposition ofthe  TBB DdRPII RPB1 model and the**
***Schizosaccharomyces pombe*  crystal structural (template)**

A. Top

B. Front

# Figure 6

**Zinc-finger formationsin the    *Trypanosoma Brucei Brucei*    DdRpII RPB1 model.**

Ribbon representation of the produced *Trypanosoma Brucei Brucei* DdRPII RPB1   model. In the produced model were highlighted 3 main zing-finger domain formations (colored grey) were contained in the clam core, clam head and active site region. Domains and domain-like regions   of the *Trypanosoma Brucei Brucei* DdRPII RPB1 *have been color-coded according* to conventions of Figures 3.

# Figure 7

**The 3D pharmacophore model for the *Trypanosoma Brucei Brucei* DdRPII RPB1 model.**

In total 5 distinct pharmacophoric features were identified. An aromatic region (colored orange), an electron donating region (colored green), two electron accepting regions (colored red) and a sulphur specific S-S interacting region (colored yellow).