Molecular & Biochemical Parasitology 217 (2017) 23–31

Contents lists available at ScienceDirect

Molecular & Biochemical Parasitology

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

Research paper

Structural and evolutionary analysis of *Leishmania* Alba proteins

Kauê Santana da Costa[a,1], João Marcos Pereira Galúcio[a,1], Elvis Santos Leonardo[a], Guelber Cardoso[b], Élcio Leal[b], Guilherme Conde[a], Jerônimo Lameira[b,*]

[a] Institute of Biodiversity, Federal University of West of Pará, Santarém, Pará, Brazil
[b] Institute of Biological Sciences, Federal University of Pará, 66075-110 Belém, Pará, Brazil

ARTICLE INFO

Keywords:
Leishmaniasis
Alba protein
Leishmania
Structural biology

ABSTRACT

The Alba superfamily proteins share a common RNA-binding domain. These proteins participate in a variety of regulatory pathways by controlling developmental gene expression. They also interact with ribosomal subunits, translation factors, and other RNA-binding proteins. The *Leishmania infantum* genome encodes two Alba-domain proteins, *Li*Alba1 and *Li*Alba3. In this work, we used homology modeling, protein–protein docking, and molecular dynamics (MD) simulations to explore the details of the Alba1-Alba3-RNA complex from *Leishmania infantum* at the molecular level. In addition, we compared the structure of *Li*Alba3 with the human ribonuclease P component, Rpp20. We also mapped the ligand-binding residues on the Alba3 surface to analyze its druggability and performed mutational analyses in Alba3 using alanine scanning to identify residues involved in its function and structural stability. These results suggest that the RGG-box motif of *Li*Alba1 is important for protein function and stability. Finally, we discuss the function of Alba proteins in the context of pathogen adaptation to host cells. The data provided herein will facilitate further translational research regarding Alba structure and function.

1. Introduction

Proteins in the Alba (Acetylation Lowers Binding Affinity) superfamily are involved in RNA/DNA-binding and share a common domain among the three domains of life (Pfam PF01918) [1,2]. The Alba proteins expressed by protozoa species have multiple functions and are highly important in regulating machinery at the RNA level. However, its mode of action was only recently described. RNA-binding proteins that participate in translation or post-transcriptional RNA regulation have been suggested to be potential drug targets against pathogenic protozoa species [3,4].

Target-based screening strategies to identify new drugs against parasites are based on the following four premises: (1) the candidate protein is considered essential for parasite survival in the host; (2) the protein exhibits little to no structural similarity with host proteins; (3) the protein contains small pockets or cavities to accommodate drug-like molecules; (4) the protein is conserved in different taxon of pathogenic species [5,6].

Alba proteins in *Leishmania* participate in various regulatory pathways controlling developmental gene expression and have been reported to interact with other RNA-binding proteins such as ribosomal subunits to regulate translation. The Alba3, previously named Alba20, expressed in *Leishmania* species are homologs of Alba3 and Alba4 from *Trypanosoma brucei* (*Tb*). Alba3 is constitutively expressed in both the amastigote and promastigote stages of the *Leishmania* life cycle.

Previous studies showed that *Li*Alba3 is involved in regulating the expression of δ-amastin during the amastigote stage of *Leishmania*. Dupé et al. conjecture that Alba proteins of *Leishmania* probably binds RNA through a conserved RNA motif flanking the U-rich region (URE) in the 3′-UTR of the δ-amastin transcript (5′-GUGCGUGYGCGUGC-3′) [7]. In addition, genetic depletion of Alba3 altered the regulation and expression of virulence factors involved in other cellular processes, such as transport, metabolism, biosynthetic processes, and parasite motility. Moreover, Alba1 (previously named Alba13) from *Leishmania infantum* (*Li*) forms a heterodimer with Alba3; although, it is not directly involved in the interaction or regulation of the δ-amastin transcript URE [7].

In *T. brucei*, Alba3 and Alba4 are constitutively expressed during the trypomastigote stage [8]. They interact with cytoplasmatic translational machinery such as poly(A)-binding proteins and the ribosomal protein, P0 [9], and are also participate in the regulatory activity of amastin mRNA [10]. It has been noted that Alba3 from *Leishmania* species; and Alba3/Alba4 from *T. brucei* contain a RGG-box domain located at the C-terminal region, which is involved in its RNA binding [7,8].

In vertebrates, RNAse P/MRP endoribonucleases are involved in

* Corresponding author.
E-mail address: lameira@ufpa.br (J. Lameira).
[1] The authors contributed equally to this work.

http://dx.doi.org/10.1016/j.molbiopara.2017.08.006
Received 27 March 2017; Received in revised form 23 July 2017; Accepted 18 August 2017
Available online 25 August 2017

tRNA and rRNA processing, respectively. These endoribonucleases contain two subunits, Rpp20 and Rpp25, with Alba domains that are homologous with protozoan Alba proteins. In vertebrates, biochemical and structural studies have shown that the Rpp25 subunit is structurally and functionally related to POP6 from *S. cerevisiae*, and the Rpp20 subunit is analogous to POP7 subunit [11,12]. In addition, studies have shown that Rpp25 and Rpp20 form a heterodimer that is crucial for its interaction with the P3 loop of RNA [13,14].

Regulation of gene expression by Alba proteins is an emerging field in the study of pathogenic parasites. In this article, we analyzed different structural determinants of *Leishmania* Alba proteins. Our recent articles used homology modeling to determine the 3D structures of the regulatory response protein (PhoP) of *C. pseudotuberculosis* [15], sialidases from *T. brucei* and *T. evansi* [16], and the virion infectivity factor from HIV-1 [17]. Herein, we used protein and RNA modeling, molecular dynamics (MD) simulations, protein–protein docking simulations, and phylogenetic analyses to obtain insights into the structure of the Alba1-Alba3-RNA complex from *Leishmania infantum*. Furthermore, we discuss these results from a functional and evolutionary perspective, deliberate on the implications of this work on parasite pathogenesis, and provide perspectives for drug-oriented therapies targeting Alba proteins.

2. Material and methods

2.1. Phylogenetic analysis

The coding DNA sequences (CDSs) of Alba proteins from different pathogenic protozoa species were obtained from the EuPathDB database [18]. Multiple sequence alignments were performed using MUSCLE [19]. MEGA7 [20] was used to select the best nucleotide model with the lowest Bayesian information criterion (BIC) scores. Next, phylogenetic analyses were performed using the maximum likelihood method and the Kimura 2-parameter [21] to calculate an evolutionary model following gamma distribution (K2 + G) with 1000 bootstrap replicates.

2.2. RNA secondary and tertiary structure modeling

There are multiple copies of the δ-amastin gene in the *Leishmania* genome. The CDS of δ-amastin was retrieved using the accession number LinJ.34.1010 in the TriTrypDB database. The chromosomal position, 34:411,171-411,806, was identified using the forward sequence (chromosome 34) with BLASTn (Ensembl Protists genome browser) [22] and the flanking sequences were determined. An 800 nt sequence located adjacent to an ORF of the δ-amastin gene within the 3′-UTR was selected. This sequence was modeled using RNAStructure to obtain the secondary structure [23]. The 3-dimensional structure of the transcript was determined using RNAComposer [24].

2.3. Protein structure predictions and complex assembly

*Li*Alba1 and *Li*Alba3 sequences were obtained from the TriTrypDB database. The *Li*Alba1 structure (LinJ.13.0270) was predicted by homology modeling, using a crystallographic structure of AT2G34160, a hypothetical protein from *Arabidopsis thaliana*, as a template, (PDB ID: 1VM0, chain A; x-ray resolution: 1.8 Å). The *Li*Alba3 structure (LinJ.34.2410) was predicted using the iterative threading assembly refinement (I-TASSER) server [25] using the archean *Aeropyrum pernix* Alba, K1, crystal structure as the template (PDB ID: 2H9U, chain A; x-ray resolution: 2.0 Å). In the I-TASSER server, regions without a reference structure in the alignment were predicted using *ab initio* modeling. To compare the Alba structures from protozoa parasites with their eukaryotic counterparts, we also modeled the hRpp25 subunit (Uniprot ID: Q9BUL9) using the I-TASSER server. Models were further refined using ModRefiner to improve the stereochemical quality [26]. Structures were energetically minimized using the steepest-descent method followed by the conjugated gradient method (1000 cycles each). Electrostatic potential maps of *Li*Alba1 and *Li*Alba3 were created in the PDB2PQR server [27], which utilizes the Poisson-Boltzmann equation to determine the distribution of electrostatic charges on proteins surfaces when solvated in a biomolecular medium. The proteins structures and the electrostatics maps were analyzed with UCSF Chimera [28].

The *Li*Alba3 and *Li*Alba1 models were aligned with the yeast structures of POP7 and POP6, respectively. The yeast proteins were used as references to superimpose the Alba domains of these two proteins. Manual adjustments were performed to prevent atomic clashes between the structures. Proteins were docked in RosettaDock and structures selected based on the Rosetta energy score [29]. The modeled structure of the δ-amastin mRNA was also docked into the *Li*Alba3/*Li*Alba1 heterodimer with NPdock server [30] using the RGG-box motif of *Li*Alba3 and the 5′-GUGCGUGYGCGUGC-3′ motif of the URE sequence as the binding interface.

2.4. Cavities, ligand binding hotspot and structural alignment search

We performed a structural alignment search of the MMBD database with NCBI/VAST [31] using the *Li*Alba3 structure (*Tb*Alba3 and *Tb*Alba4 homolog), which was modeled by threading, and *Li*Alba1 (*Tb*Alba1/*Tb*Alba2 homolog) to identify similar structures and potential structural correlations with mammalian proteins. The VAST server performs a structural alignment using only geometric criterion in all structures deposited in the NCBI/MMDB database. This server provided functional insights about protein structures with similar folding, but without sequence homology. To determine possible cavities present in the surface of the *LiAlba*1-Alba3 heterodimer, a prerequisite to assess its druggability, we performed a geometric analysis using MetaPocket (version 2.0) [32]. Ligand binding residues within the cavities of *Li*Alba3 were determined using molecular dynamics with FTMap [33].

2.5. Molecular dynamics simulations

To analyze the stability and conformational differences between the *Li*Alba1-*Li*Alba3-RNA complex and this complex with the inhibitor, we performed molecular dynamics simulations using Amber16 package [34]. The total simulation time for the complex containing 67,720 atoms was 100 ns, while the complex bound with inhibitor took 50 ns. The Amberff14SB and Amberff99bsc0χOL3 force fields were used to parameterize the proteins and RNA, respectively. The General Amber Force Field (GAFF) was used to parameterize the ligand. The system was solvated in a truncated octahedral water box using the explicit solvation model, TIP3P [35]. A distance of 10.0 Å was used between the cell wall and the solvated atoms of the system, and a distance of 0.8 Å was used between water molecules and the solute. Na^+ counter-ions were also added to neutralize the system. Initially, all of the hydrogen atoms within the system were minimized within 3000 cycles using the steepest-descent algorithm [36] and 3000 cycles using the conjugate gradient method [37]. Water and ions were minimized with 2000 cycles of the steepest-descent algorithm and 3000 cycles of the conjugate gradient method. Then, the whole system was minimized for 2000 cycles using the steepest-descent method followed by 3000 cycles of the conjugate gradient method; seven repetitions were performed with progressive relaxing of restraints. After minimization, the system was gradually heated to 300 K during 4.25 ns of equilibration. Then, molecular dynamics were performed using the isobaric-isothermal ensemble for 100 ns. The temperature was maintained using the Langevin thermostat and SHAKE was used to maintain the hydrogen bonds at their pre-defined equilibrium distances during minimization, which allowed us to use integration cycles of 2.0 fs. A cutoff of 10.0 Å was used to determine the minimum image convention. The RMSD values based on the heavy atoms of the *Li*Alba1 and *Li*Alba3 backbones (C, N,