Molecular Modeling Approaches for Determining Gene Function: Application to a Putative Poly-A Binding Protein from *Leishmania amazonensis* (LaPABP)

FP Silva-Jr, FZ Veyl*, J Clos*, S Giovanni De Simone/+

Laboratório de Bioquímica de Proteínas e Peptídeos, Departamento de Bioquímica e Biologia Molecular, Instituto Oswaldo Cruz-Fiocruz, Av. Brasil 4365, 21045-900 Rio de Janeiro, RJ, Brasil *Tropical Medicine Institute, Unit of *Leishmania* Cloning, Hamburg, Germany

The great expansion in the number of genome sequencing projects has revealed the importance of computational methods to speed up the characterization of unknown genes. These studies have been improved by the use of three dimensional information from the predicted proteins generated by molecular modeling techniques. In this work, we disclose the structure-function relationship of a gene product from Leishmania amazonensis *by applying molecular modeling and bioinformatics techniques.*

The analyzed sequence encodes a 159 aminoacids polypeptide (estimated 18 kDa) and was denoted LaPABP for its high homology with poly-A binding proteins from trypanosomatids. The domain structure, clustering analysis and a three dimensional model of LaPABP, basically obtained by homology modeling on the structure of the human poly-A binding protein, are described. Based on the analysis of the electrostatic potential mapped on the model's surface and conservation of intramolecular contacts responsible for folding stabilization we hypothesize that this protein may have less avidity to RNA than it's L. major *counterpart but still account for a significant functional activity in the parasite. The model obtained will help in the design of mutagenesis experiments aimed to elucidate the mechanism of gene expression in trypanosomatids and serve as a starting point for its exploration as a potential source of targets for a rational chemotherapy.*

Key words: molecular modeling - poly-A binding protein - *Leishmania amazonensis* - LaPABP - bioinformatics

The genome-sequencing projects are providing a detailed "parts list" of life. A key to comprehending this list is understanding the function of each gene and each protein at various levels (Skolnick & Fetrow 2000). Structure and function in proteins are closely related. Despite rapid growth of known protein sequences, direct experimental determination of their structure by nuclear magnetic ressonance (NMR) or X-ray crystallography is still quite time consuming and often limited by the protein size (NMR) or the availability of crystals (Dandekar & König 1997). Knowledge of protein structure is fundamental to understanding mechanism of action, and prediction of structure for new sequences is of great value to such studies (Westhead & Thornton 1998).

When considering whole parasite genomes, comprising thousands of genes, the actual challenge is to assemble, catalogue and analyze this information in a robust and useful manner (Fairlamb 2001). In this work, current public available bioinformatics and molecular modeling tools were used in a generic approach to determine the structure-function relationship of unknown genes. This methodology was applied to study a genomic sequence from *Leishmania amazonensis* (Veyl et al., unpublished data) which has been shown to have high homology to the poly-A binding protein class and for this reason hereafter named LaPABP.

A marked characteristic of trypanosomatid parasites (many pathogenic for humans) is the permutation between intra and extracellular forms in both invertebrate and mammalian hosts. Thus, stage specific gene expression in trypanosomatids must be efficiently regulated and this has been assumed to occur, mostly at a post-transcriptional level, either in the nucleus or in the mitochondrion by trans-splicing or editing (respectively) and polyadenylation (Vanhamme & Pays 1995). Diverse RNA binding proteins (RBPs) are likely to be involved in these processes and the primary structural characterization of these polypeptides from Trypanosomatidae have only recently began (Cross et al. 1993, Marchal et al. 1993, Metzenberg et al. 1993). Maybe, one of the mostly studied RBPs is the poly(A) binding protein 1 (PABP1) of eukaryotes (reviewed by Sachs & Wahle 1993). This protein is related to many biological roles involved with the presence of multiple adenine nucleotides runs in the 3' untranslated region (UTR) of mRNAs. Primary functions of PABP1 include stimulation of translation initiation, regulation of mRNA degradation and regulation of the poly(A) tail during the polyadenylation reaction. Characterization of genes encoding PABP-1 homologues in many organisms has shown that this protein is structurally conserved, consisting of four RNA binding domains (RBDs) also named RNA recognition motifs (RRMs; Burd & Dreyfuss 1994) on it's N-terminal two-thirds and a C-ter-

minal domain, also containing a conserved motif (unique PABP domain). Structural determination of some protein-sRNA complexes (reviewed by Antson 2000) showed that RBDs are usually 80-100 residues long, folded into a four-stranded antiparallel β sheet, comprising two conserved motifs, RNP1 (octamer) and RNP2 (hexamer), mapped onto the two central strands (reviewed by Kenan et al. 1991, Birney et al. 1993, Burd & Dreyfuss 1994). Although single domains do not bind poly(A) tails, the two N-terminal RNP domains interacts with RNA through a groove formed by the β-sheet surfaces of these domains, which are connected by a 9-residue linker. RNA binds to one side of the β-sheet, whereas the other side is protected from the solvent by two α-helices connecting the β-strands.

In this work, we describe the domain structure and clustering analysis of a *L. amazonensis* poly-A binding protein (LaPABP) sequence based on information gathered from multiple sequence alignments. We also propose a general procedure for building a theoretical 3D model and analyze the potential RNA binding property of LaPABP in terms of the electrostatic potential on its modeled surface and of fold stabilizing interactions between elements of secondary structure.

MATERIALS AND METHODS

Cloning of LaPABP - LaPABP had its gene cloned and sequenced by Veyl et al. (unpublished data) from a *L. amazonensis* (MHOM/BR/77/LTB0016) genomic library (Hubel & Clos 1996).

Sequence analysis - The sequence of LaPABP was used to perform a Position-Specific Iterated BLAST (PSI-BLAST; Altschul et al. 1997) search on non-redundant protein databank (GenBank CDS translations, PDB, SwissProt, PIR and PRF) using the BLOSUM62 substitution matrix (Gap costs: 11 existence and 1 for extension). Secondary structure, residue composition (PHD) and a multiple sequence alignment with detachment of physico-chemical properties conservation (MaxHom) were performed automatically using the PredictProtein Server (PP server; Rost 1996). For comparison, secondary structure elements were also predicted using the PSIPred program (Jones 1999), and a manual method based on the analysis of buried and exposed residues patterns.

Domain assignment and custering analysis - Reverse position-specific BLAST (RPS-BLAST; Altschul et al. 1997) was used to search matches on the NCBI conserved domains database (CDD) which is composed by entries from PFAM protein families (Bateman et al. 2000) and SMART (Schultz et al. 1998) databases. Mapping of residues belonging to the motifs involved in RNA recognition was done by careful analysis of multiple alignments generated in ClustalW, using default options (Gap Opening Penalty:10.00, Gap Extension Penalty: 0.05, Delay divergent sequences: 40%; Thompson et al. 1994) and colored by conservation pattern of homologous RBPs containing RNP motifs. Dendrograms were inferred from the aligned RBPs sequences described above. The trees generated by this procedure were displayed by the program NJPLOT.

Comparative molecular modeling - Template structures to be used in the homology modeling were selected by searching the Brookhaven Protein DataBank (PDB; www.rcsb.org/pdb/). The following structures were used to construct models of LaPABP: A chain of the paraneoplastic encephalomyelitis antigen (HUD; PDB code 1FXL), A chain of the second RNA-binding domain (RBD2) of hu antigen c (HUC; PDB code 1D9A), RBD1,2 of human hnRNP A1 (PDB code 2UP1), A chain of *Drosophila melanogaster* sex lethal protein (sxl-lethal; PDB code 1B7F) and E chain of the human Poly(A) binding protein (PDB code 1CVJ). Alignment of the selected templates to the LaPABP sequence was performed using the Hidden Markov Model (HMM) generated by the SAM-T99 program (Karplus et al. 1999) and optimized based on the conservation of secondary structures. Threading of LaPABP onto templates were done in Swiss-PDB Viewer v3.51 (Guex & Peitsch 1997) and then submitted for the automated homology modeling server Swiss-Model (http://www.expasy.ch/swissmod /SWISS-MODEL.html). The ProModII package (Peitsch 1996), implemented in Swiss-Model, automatically generates lacking loops either by searching loop databases (knowledge based approach) or by exploring conformational space. Next, completing of backbone (if necessary) and correction is done using a database of backbone fragments. Then, side chains are rebuilt and corrected based on a library of allowed side

TABLE

Homologous proteins to LaPABP found by PSI-BLAST search

Protein	Organism	Identity (%)	Access code
PABP	*Trypanosoma cruzi*	61 (157 aa)	gb|AAC46487.1
PABP-1	*Trypanosoma brucei*	59 (157 aa)	gb|AAD13337.1
PABP	*Arabidopsis thaliana*	40 (159 aa)	pir|T00497
PABP-4	*Homo sapiens*	36 (159 aa)	gb|AAB97309.1
PABP-5	*Arabidopsis thaliana*	36 (159 aa)	gb|AC012654
PABP-2	*Homo sapiens*	36 (157 aa)	pir|PS0381
PABP-1	*Mus musculus*	36 (158 aa)	sp|P29341|PAB1_MOUSE
PABP-1	*Homo sapiens*	36 (158 aa)	sp|P11940|PAB1_HUMAN
PABP-1	*Xenopus laevis*	36 (158 aa)	Emb|CAA40721.1
PABP-1	*Saccharomyces cerevisiae*	37 (156 aa)	sp|P04147|PABP_YEAST

PABP: poly(A) binding protein; gb: GenBank; pir: PIR; sp: SwissProt; emb: Tr-EMBL
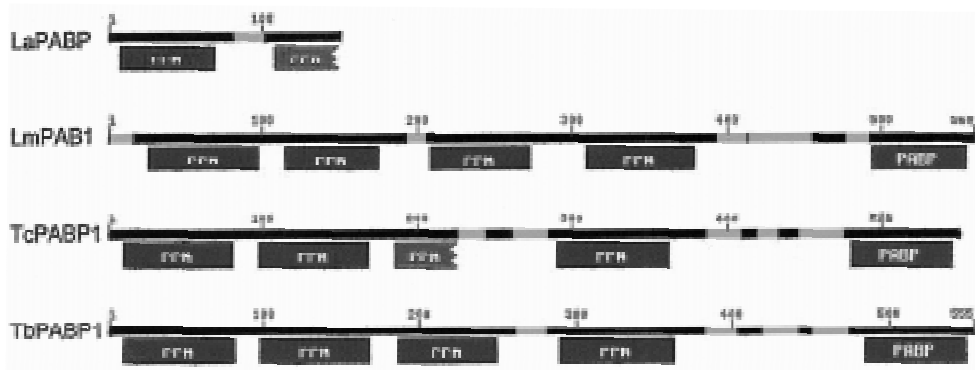
Fig. 1: domain mapping of LaPABP and other PABPs from Trypanosomatidae. Domains are represented by shaded boxes whereas low complexity regions are in light gray. RRM = RNA recognition motif and PABP = unique poly(A) binding protein domain.

chain rotamers. Finally, the overall model quality is verified by analyzing the 3D context of each residue and the packing of the structure is checked. The final model is obtained after refinement by 200 cycles of steepest descent and then 300 cycles of conjugate gradient energy minimization by the GROMOS96 force field implemented in Swiss-Model. Model quality was further assessed with the programs PROCHECK (Laskowski et al. 1993), PROVE (Pontius et al. 1996), WHATIF (Hooft et al. 1996) and the Swiss-PDB Viewer analytical tools.

RESULTS

RBPs which presented higher homology to LaPABP are listed in the Table. LaPABP showed to be approximately 60% identical to the PABPs from *Trypanosoma cruzi* and *T. brucei*. Two RNA binding domains could be assigned to LaPABP through RPS-blast searches on the CDD in NCBI (Fig. 1). This procedure also revealed that the second RBD of LaPABP is possibly incomplete. Residues that would be mapped to the fourth β-strand and the second a-helix are missing. Sequence alignment of LaPABP with representative PABPs from Trypanosomatidae and higher eukaryotes made it possible to map the RNP-1 and RNP-2 conserved motifs (data not shown). The clustering analysis of LaPABP and a set of representative PABPs sequences (Fig. 2) showed that LaPABP belongs to a branch where is located the subgroup containing the try-panomatid PABPs from *T. cruzi* and *T. brucei*.

The strategy used to ascertain the best alignment between LaPABP and amino acid sequences from template structures (30-35% homology to LaPABP) used on its modeling is summarized in Fig. 3. Briefly, a consensus between two secondary structure predictions made by the PSI-Pred and the PHD WWW servers and a manual prediction based on the pattern of hydrophobic/hydrophilic aminoacid conservation was used to improve the alignment between LaPABP and its templates proposed by an HMM model. The structure of human PABP1, which is deposited on the PDB as an octameric complex with polyadenylate RNA at 2.60 Å resolution, is the only PABP with a experimentally determined structure. Hence, it was chosen to be the primary template on the homology modeling procedure because it would furnish the right orientation of the two adjacent RRM linked by the variable loop region. This orientation seems to be of key importance to the RNA sequence binding specificity. Other RNA binding protein structures containing RRM domains which showed ≥ 30% identity to LaPABP were used in order to improve the confidence on the modeling of independent domains. The theoretical model of LaPABP was obtained after submitting the corrected alignment to the SwissModel automated homology modeling server. The overall conformation of LaPABP model is similar to other RRM containing RNA binding structures (Fig. 4A, B, C) showing the four antiparallel strands forming the β-sheet responsible for RNA binding and on the other face the two hydrophobic α-helices. The RMSD for the 134 C$^\alpha$ superposed on human PABP1 was 1.45 Å, showing that the RRM fold is high conserved (Fig. 4D). Comparison between the molecular surface responsible for polyadenylate binding in human PABP1 and the corre-
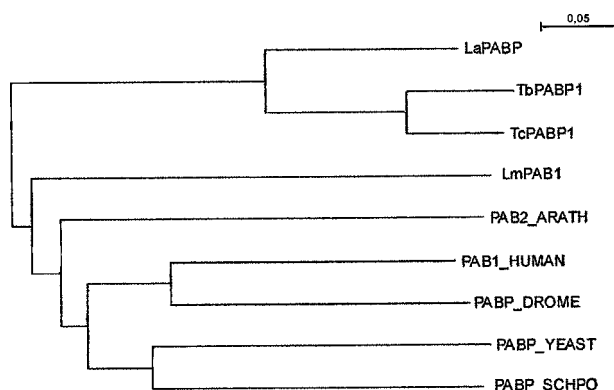

Fig. 2: clustering of representative trypanosomatid and higher eukaryotes PABP1s sequences. The guide tree obtained from the multiple alignment of these sequences in ClustalW v. 1.60 was displayed with the program NJPLOT. PAB1_HUMAN - *Homo sapiens*, PABP_SCHPO - *Schizosaccharomyces pombe*, PABP_YEAST - *Saccharomyces cerevisiae*, PABP_DROME - *Drosophila melanogaster*, PAB2_ARATH - *Arabdopsis thaliana*, TcPABP1 - *Trypanosoma cruzi*, TbPABP1 - *T. brucei*, LmPAB1 - *Leishmania major* and LaPABP – *L. amazonensis* (this work).
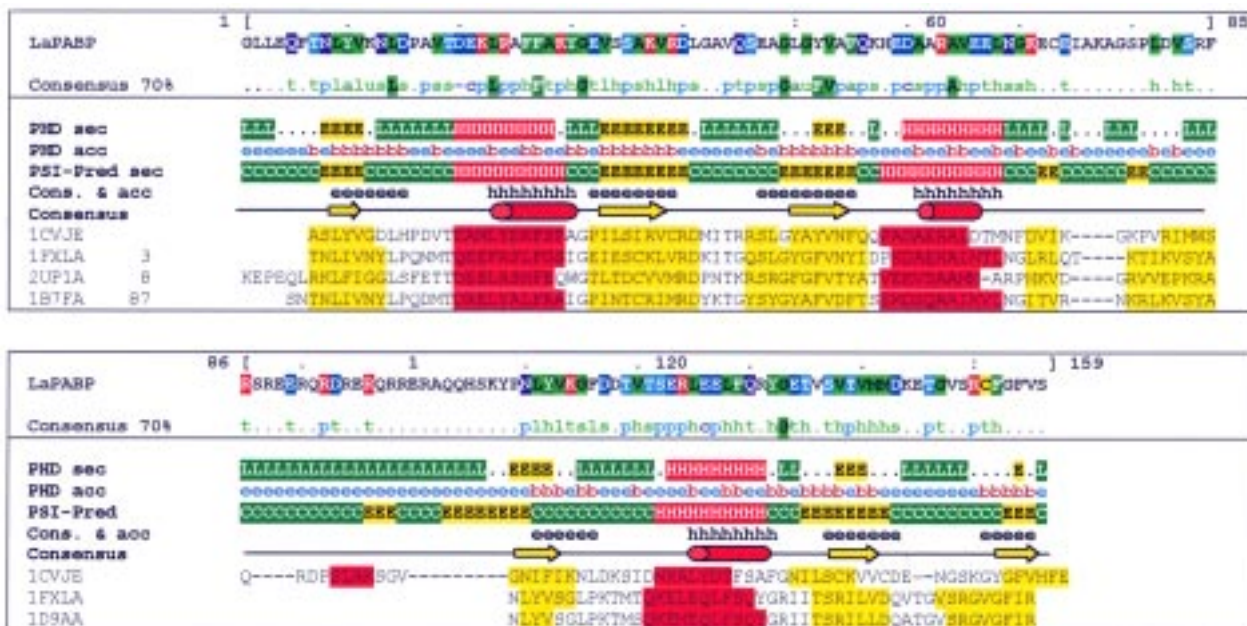
Fig. 3: summary of the alignment proposition strategy used in LaPABP homology modeling. On each panel, listed from top to bottom: LaPABP sequence colored by physicochemical properties conservation according to the consensus 70% of aligned homologous proteins obtained by the program MaxHom; light green: aliphatic hydrophobic (P,M,V,G,L,A,I); dark green: aromatic hydrophobic (Y,F,H,W); light blue: alcohols-polar (S,T); blue: negatively charged (D,E); purple: amides-polar (Q,N); red: positively charged (R,K); yellow: cysteine (C); in the consensus/70% line, greenish letters corresponds to hydrophobic residues classes and bluish letters corresponds to hydrophilic residues: t - turn-like; p -polar; l - aliphatic; a aromatic; u tiny; s small; (-) negatively charged; (+) positively charged; h hydrophobic; PHD sec and PHD acc: PHD predictions for secondary structure and accessibility; L - loop; H - helical; E - extended; e - exposed; b - buried; PSI-pred sec: Psi-Pred prediction of secondary structure; C - coil; H - helix; E - extended; Cons.&acc: manual secondary structure prediction based on the conservation of hydrophobic/buried residues. In the next line is represented the consensus between the methods used in the secondary structure prediction: yellow arrows - β-sheets and red cylinders - α-helix. In the bottom is shown the alignment between homologous RBP structures and LaPABP derived from the Hidden Markov Model maximized to preserve secondary structures.

sponding region in LaPABP shows that there is a lower positive charge density in the RMM2 region of LaPABP (Fig. 5). This data suggest that LaPABP would bind RNA with lower avidity. Residues involved in stabilizing packing interactions between RRMs in human PABP1 have been mapped (Deo et al. 1999). Superposition of the corresponding residues in LaPABP (Fig. 6) showed that only two pairs of molecular contacts were changed (Lys129-Phe74 for Thr142-Gly70 and Tyr116-Met85 for Glu129-Ser85).

DISCUSSION

As the first information about the function of an unknown gene often comes from homology to described sequence proteins deposited in databanks, the choice for the most adequate similarity searching software is essential. The PSI-BLAST program performs an iterative search in which sequences found in one round of execution are used to build a score model for the next round of searching. This tool is publicly available at the NCBI BLAST homepage (http://www.ncbi.nlm.nih.gov/ BLAST/) and is recommended for searching protein sequence databanks with a novel protein sequence because its algorithm enhance the probability of finding distant homologues.

The first step after finding a putative function for the novel gene is to assign its domain structure and study how its sequence aligns to the other sequences belonging to its class. Following this principle the sequence targeted in this work which is 159 residues long and is predicted to weight 18 kDa, showed to be homologous to poly-A binding proteins from trypanosomatids and other eukaryotes (Table). The first Trypanosomatidae PABP1 to have its gene cloned and sequenced was from *T. cruzi* (Batista et al. 1994). This protein (TcPABP1) has 66 kDa and is similar to PABP1 of other eukaryotic organisms, which show molecular weights ranging from 64-73 kDa. Recently, were characterized PABP1s from *T. brucei* (Hotchkiss et al. 1999) and *L. major* (Bates et al. 2000). The protein from *T. bruc*ei (TbPABP1) possess a predicted molecular weight of 62 kDa and is 88.7% similar to TcPABP1. As the protein from *T. cruzi*, TbPABP1 is nearly 40% similar to the PABPs from other organisms and show conservation of the four RBDs in the N-terminal two-thirds of the protein. On the other hand, the PABP1 from *L. major* (LmPAB1) shows no more overall conservation with other trypanosomatid PABP1s than with a range of other eukaryotic PABP1s. Interestingly, clustering analysis showed that LaPABP is nearer evolutionary to *Trypanosoma* sp. than to the *Leishmania* species analyzed (Fig. 2).

It is clear from its size that LaPABP present a different domain architecture from other trypanosomatid PABP. LaPABP possesses only two RBDs whereas the second one is incomplete (Fig. 1). Looking at this scenario one

should promptly argue if LaPABP sequence is truncated considering the evident absence of the two N-terminal RBDs. However, we propose that LaPABP sequence can correspond to a full functional protein. Our hypothesis is supported by a number of evidences: (i) it is estimated that LmPAB1 would account for only 50% of the cytoplasmic poly(A) binding activity observed in *L. major* cultures, suggesting the presence of another abundant

RNA binding protein which would interact with poly(A) tails with lower affinity (Bates et al. 2000); (ii) it has been shown that only two RRM are necessary for RNA binding in vitro and that the first two RRMs in PABP bind polyadenylate with higher affinity than the third and fourth domains do (Nietfeld et al. 1990); (iii) the PABP from the lower eukaryotes *Dictyostelium discoideum* and *Physarum polycephalum* were reported to have lower
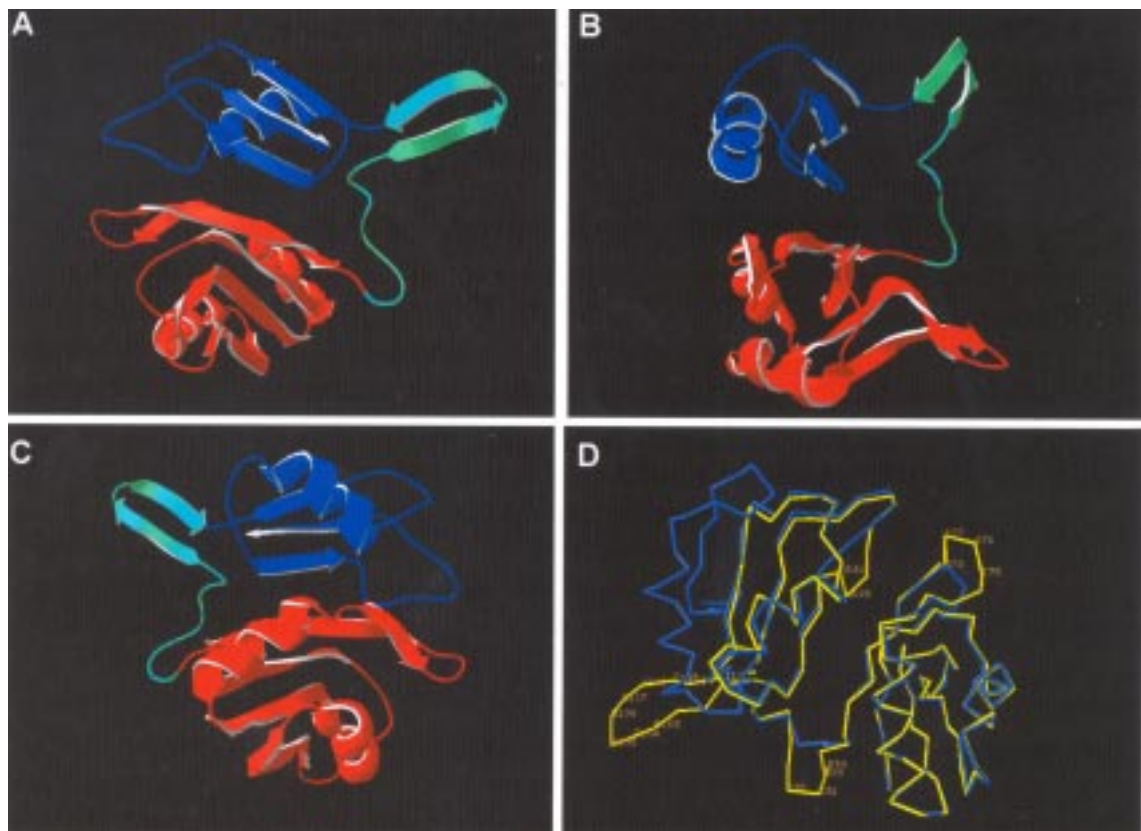

Fig. 4: theoretical model of LaPABP. Model colored by domain structure: RBD-1 in red, RBD-2 in blue and linker region in green. A: front view showing the RNA binding surface; B: left side view, C: back view. D: C$^\alpha$ traces superpositioning of LaPABP (yellow) on human PABP1 (blue; PDB: 1CVJE).
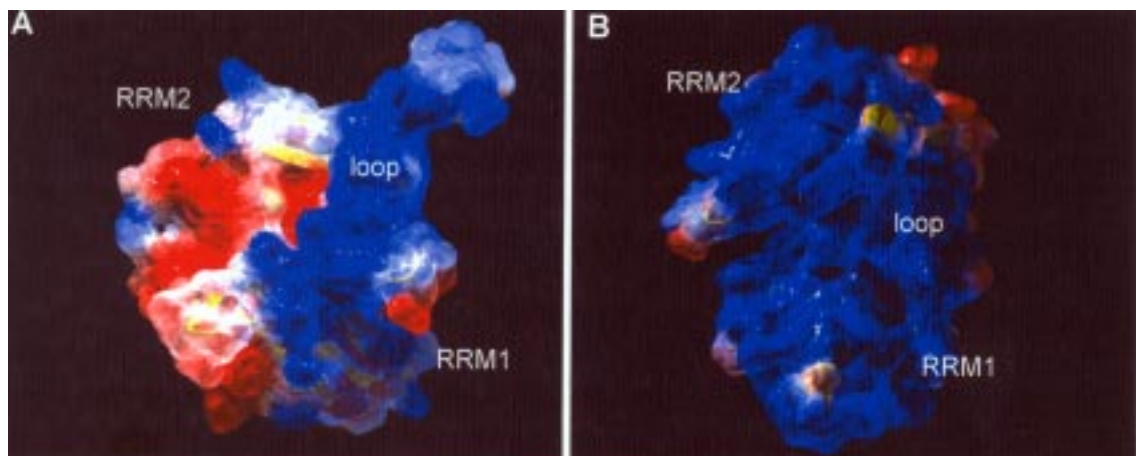

Fig. 5: comparison between RNA binding surfaces of LaPABP (A) and human PABP1 (B). Surfaces were colored by electrostatic potential: blue - positively charged, red - negatively charged and white - uncharged. In background, is shown the ribbons diagram in yellow.
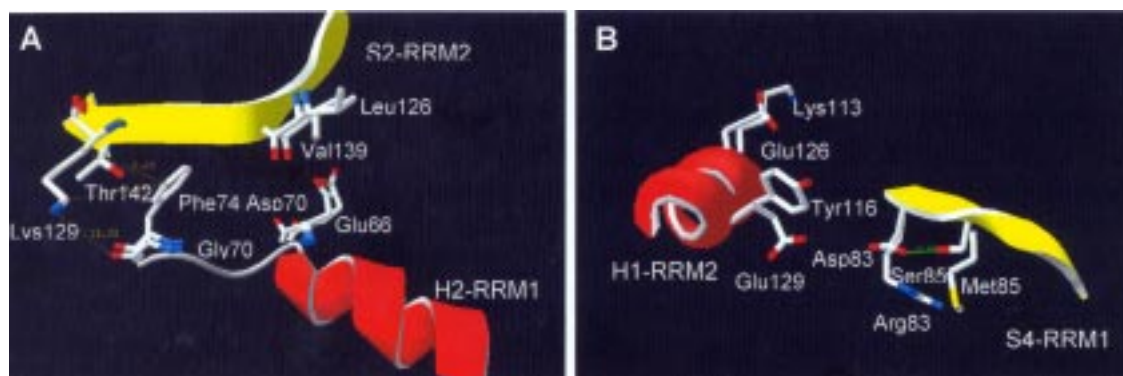
Fig. 6: analysis of residues involved in the stabilizing interactions between RRM1 and RRM2 of LaPABP and human PABP1. A: superposition of h2-RRM1 and s2-RRM2 residues of LaPABP (h2-RRM1: Gly70, Asp74; s2-RRM2: Val139, Thr142) and human PABP1 (h2-RRM1: Asp70, Phe74; s2-RRM2: Leu126, Lys129); B: superposition of h1-RRM2 and s4-RRM1 residues of LaPABP (h1-RRM2: Glu129, Glu126; s4-RRM1: Asp83, Ser85) and human PABP1 (h1-RRM2: Tyr116, Lys113; s4-RRM1: Arg83, Met85).

molecular weights than exhibited by the PABP from more complex organisms (hence with a different architecture from the classical PABP) but still showing a significant poly-A binding activity (Batista et al. 1994). However, the lack of a number of residues (less than 20) corresponding to the remainder of the second RRM domain could indicate that the clone obtained really correspond to a C-terminal truncated form of this parsimonious version of the observed PABP domain architecture. Nevertheless, two observations encouraged us to consider that the sequence analyzed could still correspond to a functional protein in vivo: (i) the second RBD found in the PABP1 from *T. cruzi* is also incomplete (Batista et al. 1994); (ii) although being incomplete, the second RBD from LaPABP presents both motifs involved in the recognition of RNA (RNP1 and RNP2; data not shown).

Thus, in order to further investigate this possibility a theoretical model of LaPABP was proposed (Fig. 4A, B, C). It is worth emphasizing that the main bottleneck of a homology modeling procedure is the quality of the alignment between the targeted sequence and its template (Dandekar & König 1997). The strategy used in this work to obtain a reasonable alignment between LaPABP, human PABP and other RRM containing RNA binding protein structures relied on the principle of secondary structure elements conservation (Fig. 3). The RMSD for the 134 C^α superposed on human PABP1 (Fig. 4D) was 1.45 Å, which is in accordance to the homology level shared between LaPABP and its templates sequences.

A common feature of all RBPs structures containing RRMs determined to date is that RNA binding occurs by interaction with the exposed β-sheet surfaces of two consecutive RRMs, while the other RRM face is protected from solvent by the two α-helices connecting the β-strands. Structural analysis, in terms of surface eletrostatic potential and fold stabilizing residues conservation, of the LaPABP 3D model generated in this work shows that the basic structural properties required for fulfilling poly(A) binding activity by the protein represented by the sequence analyzed are present. Absence of the forth β-strand in the second RRM was the main responsible for

the smaller positive charge density on the modeled protein surface (Fig. 5). This data can be direct extrapolated to a poor RNA binding avidity in vivo, which is compatible to the existence of a second lower affinity PABP in *Leishmania* as pointed out by others (Bates et al. 2000). The helices belonging to the solvent-protected face present in RRMs have been proposed to account for the specificity of protein-protein interactions in the cells (Deo et al. 1999). Thus the absence of the second α-helix on RRM2 of LaPABP may reflect a different pattern of protein-protein interations in the parasite. Finally, the conservation of the main interdomain contacts responsible for stabilizing two consecutive RBDs supports a functional folding (Fig. 6). In conclusion, the bioinformatics and molecular modeling approaches used in this work showed to be able to give relevant information on the biochemical properties and biological roles of a putative *Leishmania* protein. PABPs structures are scarce and the theoretical model of LaPABP generated here is the first among PABPs from *Leishmania* sp. Its importance can be relied on the intricate gene expression in trypanosomatids whose elucidation would be useful for understanding its roles on the parasite infection and its exploration as a potential source of targets for rational chemotherapy of parasitic diseases.

REFERENCES

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res 25*: 3389-3402.

Antson AA 2000. Single stranded RNA binding proteins. *Current Opinion Struc Biol 10*: 87-94.

Bateman EB, Durbin R, Eddy SR, Howe KL, Sonnhammer LLL 2000. The Pfam protein families database. *Nucleic Acids Res 28*: 263-266.

Bates EJ, Knuepfer E, Smith DF 2000. Poly(A)-binding protein 1 of *Leishmania*: functional analysis and localization in trypanosomatid parasites. *Nucleic Acids Res 28*: 1211-1220.

Batista JAN, Teixeira SMR, Donelson JE, Kirchhoff LV, Sá CM 1994. Characterization of a *Trypanosoma cruzi* poly(A)-binding protein and its genes. *Mol Biochem Parasitol 67*: 301-312.

Birney E, Kumar S, Krainer AR 1993. Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res 21*: 5803-5816.

Burd CG, Dreyfuss G 1994. Conserved structures and diversity of functions of RNA-binding proteins. *Science 265*: 615-621.

Cross M Wieland B, Palfi Z, Gunzl A, Rothlisberger U, Lahm H-W, Bindereif A 1993. The trans-spliceosomal U2 snRNP protein 40K of *Trypanosoma brucei:* cloning and analysis of functional domains reveals homology to a mammalian snRNP protein. *EMBO J 12:* 1239-1248.

Dandekar T, Konig R 1997. Computational methods for the prediction of protein folds. *Biochim Biophys Acta 1343*: 1-15.

Deo RC, Bonanno JB, Sonenberg N, Burley SK 1999. Recognition of Polyadenilate RNA by the Poly(A)-Binding Protein. *Cell 98*: 835-845.

Fairlamb AH 2001. Brave new world of postgenomics. *Trends Parasitol 17*: 255-256.

Guex N, Peitsch MC 1997. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis 18*: 2714-2723.

Hooft RWW, Vriend G, Sander C, Abola EE 1996. Errors in protein structures. *Nature 381*: 272-280.

Hotchkiss TL, Nerantzakis GE, Dills SC, Shang L, Read LK 1999. *Trypanosoma brucei* poly(A) binding protein I cDNA cloning, expression, and binding to 5' untranslated region sequence elements. *Mol Biochem Parasitol 98*: 117-29.

Hubel A, Clos J 1996.The genomic organization of the HSP83 gene locus is conserved in three *Leishmania* species. *Exp Parasitol 82*: 225-228.

Jones DT 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol 292*: 195-202.

Karplus K, Barrett C, Hughey R 1999. Hidden Markov models for detecting remote protein homologies. *Bioinformatics 14*: 846-856.

Kenan DJ, Query CC, Keene JD 1991. RNA recognition: towards identifying determinants of specificity. *Trends Biochem Sci 16*: 214-20.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst 26*: 283-291.

Marchal C, Ismaili N, Pays E 1993, A ribosomal S12-like gene of *Trypanosoma brucei. Mol Biochem Parasitol 57:* 331-334.

Metzenberg S, Joblet C, Verspieren P, Agabian N 1993. Ribosomal protein L25 from *Trypanosoma brucei*: phylogeny and molecular co-evolution of an rRNA-binding protein and its rRNA binding site. *Nucleic Acids Res 21*: 4936-4940.

Nietfeld W, Mentzel H, Pieler T 1990. The *Xenopus laevis* poly(A) binding protein is composed of multiple functionally independent RNA binding domains. *EMBO J 9*: 3699-3705.

Peitsch MC 1996. ProMod and Swiss-Model: Internet-based tools for automated comparative protein modeling. *Biochem Soc Trans 24*: 274-279.

Pontius J, Richelle J, Wodak SJ 1996. Quality assessment of protein 3D structures using standard atomic volumes. *J Mol Biol 264*: 121-136.

Rost B 1996. PHD: predicting one-dimensional protein structure by profile based neural networks. *Methods Enzymol 266*: 525-539.

Sachs A, Wahle E 1993. Poly(A) tail metabolism and function in eukaryotes. *J Biol Chem 268*: 22955-22958.

Schultz J, Milpetz F, Bork P, Ponting CP 1998. SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc Natl Acad Sci USA 95*: 5857-5864.

Skolnick J, Fetrow JS 2000. From genes to protein structure and function: novel applications of computational approaches in the genomic era. *Trends Biotechnol 18*: 34-39.

Thompson JD, Higgins DG, Gibson TJ 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res 22*: 4673-4680.

Vanhamme L, Pays E 1995. Control of gene expression in trypanosomes. *Microbiol Rev 59*: 223-240.

Westhead DR, Thornton JM 1998. Protein structure prediction. *Curr Opin Biotechnol 9*: 383-389.