Structure and mechanism of dimer-monomer transition of a plant poly(A)-binding protein upon RNA interaction: insights into its poly(A) tail assembly

Mariane Noronha Domingues[1†], Mauricio Luis Sforça[1†], Adriana Santos Soprano[1], Jack Lee[2] ~~Adriana Santos Soprano[1]~~, Tatiana de Arruda Campos Brasil de Souza[3], Alexandre Cassago[4], Rodrigo Villares Portugal[4], Ana Carolina de Mattos Zeri[1], Mario Tyago Murakami[1], Ari Sadanandom[2], Paulo Sergio Lopes de Oliveira[1] and Celso Eduardo Benedetti[1*]

[1]Laboratório Nacional de Biociências, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, SP, CP6192, Brazil

[2]School of Biological and Biomedical Sciences, Durham University, Durham, DH1, UK

[3]Instituto Carlos Chagas, FIOCRUZ, Curitiba, PR, Brazil

[4]Laboratório Nacional de Nanotecnologia, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, SP, CP6192, Brazil

[†]Equal contributors

[*]Correspondence to: Celso E. Benedetti, Laboratório Nacional de Biociências, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, SP, CP6192, Brazil. Phone: +55 19 35121111; Fax: +55 19 35121006; Email: celso.benedetti@lnbio.cnpem.br.

Abstract

Poly(A)-binding proteins (PABPs) play crucial roles in mRNA biogenesis, stability, transport and translational control in most eukaryotic cells. Although animal PABPs are well-studied proteins, the biological role, three-dimensional structure and RNA-binding mode of plant PABPs remain largely uncharacterized. Here, we report the structural features and RNA-binding mode of a *Citrus sinensis* PABP (CsPABPN1). CsPABPN1 has a domain architecture of nuclear PABPs (PABPN) with a single RNA recognition motif (RRM) flanked by an acidic N-terminal and GRPF-rich C-terminal. The RRM domain of CsPABPN1 displays virtually the same three-dimensional structure and poly(A)-binding mode of animal PABPNs. However, while the CsPABPN1 RRM domain specifically binds poly(A), the full-length protein also binds poly(U). CsPABPN1 localizes to the nucleus of plant cells and undergoes a dimer-monomer transition upon poly(A) interaction. We show that poly(A) binding by CsPABPN1 begins with the recognition of the RNA-binding sites RNP1 and RNP2, followed by interactions with residues of the β2 strands, which stabilize the dimer, thus leading to dimer dissociation. Like human PABPN1, CsPABPN1 also seems to form filaments in the presence of poly(A). Based on these data, we propose a structural model in which contiguous CsPABPN1 RRM monomers wrap around the RNA molecule creating a super-helical structure that could not only shield the poly(A) tail, but also serve as a scaffold for the assembly of additional mRNA processing factors.

Keywords: CsPABPN1, poly(A)-binding protein, RRM domain, *Citrus sinensis*, dimer-monomer transition,

Introduction

Most eukaryotic mRNAs are post-transcriptionally modified by processes like capping, splicing, 3'-end cleavage, polyadenylation and deadenylation [1,2]. The process of polyadenylation involves the addition of a poly(A) tail to the 3'-end of mRNAs generating a scaffold for the binding of poly(A)-binding proteins (PABPs), the main regulatory proteins that interact with the poly(A) tail [1,2].

PABPs play multiple roles in post-transcriptional gene regulation, controlling not only the 3'-end processing, but also the stability, nuclear export, translational activity and decay of mRNAs [1-4]. Two evolutionary conserved PABPs exist in most eukaryotic cells: the cytoplasmic PABPCs and nuclear PABPNs [1,2].

PABPC1 is the most studied variant among the PABPC family members; it contains four copies of non-identical RNA recognition motifs (RRMs) in ~~its~~ the N-terminal and a Pro-rich domain required for oligomerization and interaction with other proteins in ~~its~~ the C-terminal [5-7]. In contrast, eukaryotic cells seem to have only one nuclear PABP (PABPN1), characterized by a coiled-coil N-terminal, a single internal RRM domain and a C-terminal region with a nuclear localization signal (NLS). The PABPN1 NLS is dependent on a Pro-Tyr dipeptide (PY-NLS) and a Gly-Arg-rich (GR) region that is also important for PABPN1 self-interaction [1,4,8-11].

Despite sharing structurally similar RRM domains, PABPC1 and PABPN1 have different domain architectures and play distinct roles in the life cycle or fate of mRNAs [1-4,12-15]. For instance, while PABPC1 has been implicated in the control of mRNA translation and miRNA-dependent deadenylation and gene silencing, PABPN1 regulates the interaction between Poly(A) polymerase and the cleavage and polyadenylation specific factor during poly(A) synthesis, controlling the length of the poly(A) tail and the processes of alternative cleavage [2,15-22]. More recently though, human PABPN1 and related yeast PAB2 have also been implicated in mRNA decay [23,24].

In contrast to animal cells, the biological roles played by plant PABPs are less clear. Flowering plants have multiple PABPs with four RRM domains, which are related to the mammalian PABPCs [25,26]. Some of them have been reported to interact with translation initiation factors, complement yeast

translational and mRNA decay processes, and to mediate translation of plant virus RNA [27-31]. However, plant PABPNs have not yet been functionally or structurally characterized.

In a previous protein-protein interaction study, we have identified two sweet orange (*Citrus sinensis*) PABPs, designated CsPABP1 and CsPABP2, as interacting partners of PthA4, the main transcriptional activator-like (TAL) effector of the citrus canker pathogen *Xanthomonas* citri [32]. CsPABP1 and CsPABP2, which are presumed to be the orthologs of mammalian PABPN1 and PABPC1, respectively, based on sequence similarities, also interacted with other PthA4-binding proteins implicated in chromatin remodeling, transcription regulation, mRNA processing and translational control [32]. In particular, both interacted with the citrus high mobility group protein CsHMG that selectively binds poly(U) RNA [32].

Here, to gain new insight into the three-dimensional structure and RNA-binding properties of CsPABP1, we solved the solution structure of a CsPABP1 deletion derivative carrying the entire RRM domain. We present evidence indicating that CsPABP1 is a dimer in solution and undergoes a dimer-to-monomer transition upon poly(A) binding. The identification of the key residues implicated in poly(A) interaction and dimer stabilization led us to propose a mechanism of dimer dissociation upon RNA binding. Our results also show that CsPABP1 localizes to the nucleus of plant cells. Collectively, our data reveal that CsPABP1, to our knowledge the first plant nuclear PABP to be characterized, indeed represents an authentic ortholog of the animal PABPNs and was thus renamed to CsPABPN1. Moreover, based on the three-dimensional structure of the CsPABPN1 RRM domain and electron microscopy images of the protein in the presence of poly(A), we propose a structural model where single RRM domains of CsPABPN1 arranged in tandem assemble into a super-helical structure that covers the poly(A) tail.

Results

CsPABPN1 is homologous to type II PABPs and localizes to the nucleus of plant cells

CsPABPN1 is a 234 amino acid protein bearing an acidic DE-rich N-terminal, a single central RRM domain and a basic GRPF-rich C-terminal region (Fig. 1a). This domain architecture is typical of type II or nuclear PABPs. CsPABPN1 is closely related to several uncharacterized plant PABPs belonging to

dicot species, and yet shares 54% and 43% identity with the *Xenopus* embryonic XlePABP2 and human PABPN1, respectively (Fig. 1a).

An interesting feature of CsPABPN1 is that it carries a DEH-rich insertion motif in ~~its~~ the N-terminal region (27-DEHEHEHDQDHEHEHDADNENEED-50) that seems unique to citrus species, since this motif was not found in any of the PAPB sequences so far available in the Genbank (Fig. 1a). In the *Ricinus communis*, *Lotus japonicus* and *Arabidopsis thaliana* PABP orthologs, however, the DEH-rich motif is partially replaced by the SSRPEEEEEEVEEEYDE, DADADAEQQDHDFASNH and TEEYEEHGGEEGAAAGDEELE regions, respectively (Fig. 1a). In addition, in the *Citrus sinensis* genome [33], we found two genes encoding proteins related to the CsPABPN1 isolated from cultivar "Pera". The orange1.1g027515m gene encodes a protein that is nearly identical to CsPABPN1; nevertheless, it has a shorter C-terminal and a "DH" polymorphism within the DEH-rich motif. On the other hand, the second PAPB, encoded by the orange1.1g041633m gene, lacks the DEH-rich motif and shows a longer C-terminal (Fig. 1a).

Because CsPABPN1 has a domain architecture that is common to nuclear PABPs, and it carries a putative PY-NLS in ~~its~~ the C-terminal end that is closely related to those of mammalian and yeast PABPNs [10,11] (Fig. 1a), we tested its subcellular localization in plant cells. We found that the full-length CsPABPN1 localizes to the nucleus of *Nicotiana benthamiana* cells, thus confirming that it belongs to the PABPN subfamily (Fig. 1b). In addition, CsPABPN1 without the PY-NLS motif was detected only in a dotted-like pattern in the cytosol of *N. benthamiana* cells (Fig. 1b), indicating that this motif, like in yeast [11], is required for the nuclear import of CsPABPN1.

The RRM domain of CsPABPN1 specifically binds poly(A), whereas full-length CsPABPN1 also binds poly(U)

RRM domains of PABPs are known to specifically bind poly(A) RNA [12,13,34]. To tested whether the CsPABPN1 RRM domain would show similar properties, a truncated version of CsPABPN1 (CsΔPABPN1), carrying the entire RRM domain (Fig. 1a), was used in RNA-binding assays with labeled

single-strand RNA probes. This truncation was designed base on secondary structure predictions, which indicated that the first 56 and last 46 residues flanking the CsPABPN1 RRM domain are unstructured. As expected, CsΔPABPN1 specifically bound the poly(A_{15}) probe in gel shift assays (Fig. 2a). Importantly, CsΔPABPN1 formed three major CsΔPABPN1:poly(A_{15}) complexes of increased molecular weight, which were more evident as the amount of CsΔPABPN1 in the binding mix increased (Fig. 2a). Given that single RRM domains bind up to five consecutive adenines [12,13,34], the three CsΔPABPN1:poly(A_{15}) complexes observed in Fig. 2a suggest a 1:1, 2:1 and 3:1 CsΔPABPN1:poly(A_{15}) stoichiometry.

In contrast, we found that, in addition to poly(A_{15}), full-length CsPABPN1 also bound poly(U_{15}) (Fig. 2b). Moreover, only a single major CsPABPN1:RNA complex of higher molecular weight was observed with the poly(A_{15}) and poly(U_{15}) probes, suggesting that the N- and/or C-terminal of CsPABPN1 affect the binding specificity of the protein to RNA molecules (Fig. 2b).

The RRM domain of CsPABPN1 adopts a canonical RRM fold

To investigate the structural features of CsPABPN1 and provide insights into how this protein binds poly(A), we solved the solution structure of CsΔPABPN1, since the full-length protein is fairly insoluble and could not be purified or concentrated in the amounts required for structural resolution. Circular dichroism (CD) measurements revealed that CsΔPABPN1 is structured in solution showing two major negative peaks at 208 and 222 nm, suggesting the presence of α-helices (Supplementary Fig. S1a). These data, and the fact that CsΔPABPN1 bound poly(A_{15}) (Fig. 2a), indicated that purified CsΔPABPN1 was suitable for structure resolution. Accordingly, ^{15}N-labeled CsΔPABPN1 showed excellent HSQC peak dispersion (Supplementary Fig. S1b).

A total of 2416 distance restraints, including 1023 medium and long NOEs, were used for structure calculations (Table 1). The twenty structures with the lowest energy were selected to represent the ensemble of protein structures (Fig. 3a), and the statistics show no significant violations of the molecular geometry parameters (Table 1). In addition, the Ramachandran plot analysis shows that 99.5% of *phi-psi*

angles are in the most favorable and additional allowed regions, with no angles in the disallowed region, whereas the root mean square deviations (RMSD) for the backbone and side chains are within expected values, indicating a good stereochemical quality of the ensemble (Table 1).

The first twenty N-terminal residues of CsΔPABPN1 were disordered, as revealed by their heteronuclear NOE values, chemical shifts and ^{15}N relaxation measurements, and thus were omitted from the structures represented in Fig. 3. In the structured N-terminal region, two extended α-helices precede the RRM domain of CsPABPN1 (Fig. 1a and 3b). These helices are not present or fully represented in the 3D structures of the RRM domains of human PABPN1 and *Xenopus* XlePABP2 [13,14], the only 3D structures of single-RRM domain PABPs reported to date (Fig. 3b).

On the other hand, as typically observed for RRM domains of PABPs [12-14], the CsPABPN1 RRM domain is composed of four-stranded antiparallel β-sheet spatially arranged and backed by two α-helices (Fig. 3c). The two conserved RNA-binding sites, known as RNP1 (K148 to F155) and RNP2 (I110 to V115) [12-14,34], correspond to the central β3 and β1 strands, respectively (Fig. 1a and 3c). A comparison of type I and type II PABP structures revealed that the RRM domain of CsPABPN1 shows the same topology of the RRM domains of PABPCs and PABPNs (Fig. 3d). However, as previously noticed for the human and *Xenopus* PABPNs [14], the so-called loop3 or variable loop [35], which joins strand β2 with β3, adopts a different orientation compared to those of PABPC RRMs (Fig. 3d).

The molecular basis of poly(A) binding in CsPABPN1

To investigate the residues implicated in CsΔPABPN1 binding to poly(A), we performed NMR titrations of ^{15}N-CsΔPABPN1 in the presence of poly(A_{15}). We found specific chemical shift changes in the ^{1}H-^{15}N-HSQC signals between poly(A)-bound and free protein (Fig. 4a). At a 1:30 poly(A):CsΔPABPN1 ratio, we observed significant changes in the chemical shifts of V112, N114, L140, G149, R183 and T184, followed by changes in I110, Y111, G113, C119, K148, Y152, V179, N185 and I186 at a 1:20 ratio (Fig. 4a and b). Not surprisingly, I110, Y111, V112, G113 and N114 comprise the RNP2 site, whereas K148,

G149 and Y152 belong to the RNP1 site (Fig. 1a). Increased amounts of poly(A) (1:10 ratio) led to additional changes in the chemical shifts of E103, V115, D116, Y117, T120, T133, R136, V137, T138, I139, T141, F150, E154, A160, L165, H173, K178, A181, and K182, thus confirming that poly(A) binding in CsΔPABPN1 is restricted to the concave side of the β-sheet (Fig. 4b), which has a large positively charged surface (Fig. 4c).

To provide further insights into how CsΔPABPN1 binds poly(A), we generated a structural model of CsΔPABPN1 in complex with poly(A_9), based on the 3D structures of the human PABPC RRM domains bound to poly(A) [12,34] and our NMR titration studies (Fig. 4d). According to this model, adenine recognition by CsΔPABPN1 is greatly mediated by Van der Waals contacts, hydrogen bonds and stacking interactions with the conserved residues of the RNP1 (F150, Y152, E154) and RNP2 (Y111, N114) sites, but also with K182 and residues of strand β2 (R136, T138) and helix α2 (Q99, K102) (Fig. 4d). Interestingly, Y111, N114, R136, F150, Y152 and E154 are among the residues that first underwent chemical shifts changes upon poly(A) interaction in our NMR titration experiments (Fig. 4a and b).

To verify these results, some of the residues implicated in adenine binding, including Y111, N114, R136 and F150, were replaced by alanines. As anticipated, we found that all the mutant proteins lost their ability to bind poly(A_{15}) in gel shift assays (Fig. 4e), thus corroborating our NMR studies and the structural model of the CsΔPABPN1-poly(A_9) complex.

Altogether, the results show that the poly(A) binding mode of the citrus CsΔPABPN1 is highly conserved with respect to that of *Xenopus* XlePABP2 and human PABPC RRM domains [12,13,34].

CsPABPN1 forms dimers in solution and yeast cells

CsΔPABPN1 was found as a monomer in the NMR experiments, as ~~estimated~~ revealed by ~~its~~ the ~~correlation time~~tau-c value of 8.2 ± 0.5 ~~7.8 ± 0.8~~ ns (Supplementary Fig. S2). However, when the protein was analyzed by size-exclusion chromatography, it eluted as a single peak with an estimated molecular weight of approximately three times of its predicted size (Fig. 5a). Since XlePABP2 is a dimer in solution

[13] and CsΔPABPN1 has an extended and partially structured N-terminal region (Fig. 3a), we presumed that the CsΔPABPN1 elongated and partially unfolded N-terminal would likely affect its behavior in gel-filtration chromatography. To investigate this assumption, we analyzed the hydrodynamic behavior of CsΔPABPN1 by SAXS. Surprisingly, we found that under lower ionic strength conditions, as those of the NMR experiments, CsΔPABPN1 is a monomer in solution (Fig. 5b). In contrast, under higher salt conditions, the protein seems to dimerize (Fig. 5b). The change in size and shape of CsΔPABPN1 between the lower and higher ionic strength conditions can be observed by the corresponding scattering and pair distance distribution function *p(r)* curves (Fig. 5b). In addition, we found that while the CsΔPABPN1 NMR structure fitted remarkably well into the SAXS envelope model obtained under low-salt conditions, the SAXS envelope obtained under high-salt conditions showed a more elongated or extended shape consistent with a dimer (Fig. 5c).

To further investigate the dimeric nature of CsPABPN1, yeast two-hybrid assays were performed with CsΔPABPN1 and the full-length protein. We found that although the RRM domain of CsΔPABPN1 was not sufficient for a self-interaction, the full-length protein strongly self-interacted in yeast cells (Fig. 5d). This suggests that the negatively charged N-terminal or the C-terminal region, which carries the NLS and is required for bovine PABPN1 self-association [8], is important for CsPABPN1 self-interaction. The fact that full-length CsPABPN1 self-interacted in yeast is also in accord with the observation that it formed a single and higher molecular weight complex with the poly(A_{15}) compared with CsΔPABPN1 (Fig. 2). Taken together, the results show that CsPABPN1 can be found as monomer or dimer depending on the environmental conditions, thus suggesting a putative regulatory mechanism for poly(A) recognition and binding.

CsΔPABPN1 shows a dimer-monomer transition upon RNA binding

Given the overall structural similarities shared between CsΔPABPN1 and XlePABP2, we tested whether CsΔPABPN1 would also display a dimer-monomer transition upon RNA interaction [13]. Chemical cross-

linking was used to investigate the oligomeric state of CsΔPABPN1 in the presence and absence of poly(A_{15}). A cross-linking of the CsΔPABPN1 protein produced an approximately 32 kDa band observed in the SDS-PAGE gels, which is consistent with a CsΔPABPN1 dimer (Fig. 6a). However, in the presence of increased amounts of poly(A_{15}), a ~16 kDa band accumulated predominantly in the SDS-PAGE gels (Fig. 6a), showing that CsΔPABPN1 undergoes a dimer-monomer transition upon poly(A) binding. The full-length CsPABPN1 also undergoes a dimer-monomer transition upon poly(A) but not poly(U) interaction (Supplementary Fig. S3).

To understand how this transition might occur, we generated *in silico* models for dimeric CsΔPABPN1 free or bound to poly(A_9), based on the 3D structures of XlePABP2 (PDB code 2JWN) and human PABPC (PDB code 1CVJ), respectively [12,13]. The structural model of the dimer in the absence of RNA shows that, similar to XlePABP2, the CsΔPABPN1 dimer is stabilized by the antiparallel pairing of the β2 and β2' strands and by two salt bridges formed between R136 and D142, located in the ends of the opposing β2 strands (Fig. 6b). Consistent with the cross-linking assay showing that dimeric CsΔPABPN1 does not fully hinder poly(A) binding (Fig. 6a), the structural model of the CsΔPABPN1 dimer in complex with poly(A) suggests that the positively charged cleft created by the joined concave faces of the RRM domains would allow the binding of the poly(A) molecule (Fig. 6c). However, the dimer model also predicts steric clashes of two adjacent adenine bases with L140 and D142 of the β2' strand (Fig. 6d). Because the dimer is stabilized by the pairing of the β2 strands, and L140 is one of the first residues to show chemical shift changes upon poly(A) interaction (Fig. 4a), our results strongly indicate that poly(A) recognition by CsΔPABPN1 begins with the RNP sites, followed by interactions with the β2' residues (Fig. 4b and d), leading to dimer dissociation.

A proposed super-helical structure for CsPABPN1 bound to a poly(A) tail

As shown for XlePABP2 [13] and CsΔPABPN1 (Fig. 5d and 6b), dimerization of this group of PABPs is in part mediated by the antiparallel pairing of the β2 and β2' strands. However, in this dimeric

arrangement, the two RRM domains adopt an inverted orientation relative to the RRM domains of PABPCs, which are contiguously connected by a nine-residue linker [12]. In addition, in the crystal structure of the human PABPC (RRM1/2) bound to poly(A), the β-sheet surfaces of the RRM1 and 2 domains form a platform that binds poly(A) in an extended conformation of about four nucleotides per RRM domain [12]. Thus, considering that CsΔPABPN1 binds poly(A) preferentially as a monomer, and that the antiparallel dimer of CsΔPABPN1 is predicted to have steric clashes with adenine bases (Fig. 6d), it is expected that the RRM domains of multiple CsPABPN1 molecules would adopt a distinct arrangement when assembling into the poly(A) tail. Based on this assumption, and on the binding mode displayed by contiguous RRM domains of PABPCs, we generated a supra-macromolecular structural model for contiguous CsΔPABPN1 molecules bound to an extended poly(A_{50}) strand (Fig. 7a and Supplementary Data S1). According to this model, adjacent CsΔPABPN1 molecules adopt almost the same poly(A)-binding topology as the contiguous PABPC RRM1/2 domains [12], but with a small twist (Fig. 7b) that seems to favor the wrapping of CsΔPABPN1 molecules around the RNA. In fact, the model shows remarkably well that the tandem arranged CsΔPABPN1 monomers wrap around the RNA strand creating a superhelical shield (Fig. 7a and Supplementary Data S1). In addition, the model shows that seven CsΔPABPN1 molecules are sufficient to make one full turn around the poly(A) tail. This CsΔPABPN1 multimeric assembly is not only stabilized by interactions with the adenine bases but also by hydrogen bonds and van der Waals contacts between the CsΔPABPN1 monomers (Supplementary Data S1). Notably, the interaction interface between monomers A and B are similar to that of monomers B and C (Fig. 7c), indicating a symmetric arrangement of the CsΔPABPN1 monomers along the poly(A) tail. The model also suggests that this macromolecular assembly results from the polymerization of CsPABPN1 units upon poly(A) interaction and dimer disruption, allowing the formation of long filaments over the length of the poly(A) tail (Fig. 7a).

Because calf thymus PABP2 forms both linear filaments and oligomeric particles in the presence of poly(A) [36], we decide to investigate if CsΔPABPN1 could also form filaments or particles with

poly(A). Similar to calf thymus PABP2, CsΔPABPN1 formed both filaments and particles associated with filaments in the presence of poly(A_{50}) (Fig. 7d). These structures were not observed with either the protein or poly(A_{50}) alone, indicating that they are products of the CsΔPABPN1-poly(A_{50}) interaction. The average diameter of the observed filaments and particles are approximately 4 and 11 nm, respectively. The overall shape and dimensions of these structures are thus reminiscent of the 7 nm filaments and 21 nm particles of full-length PABP2, which is twice as big as CsΔPABPN1 [36]. Furthermore, the diameter of the observed CsΔPABPN1-poly(A_{50}) filaments is consistent with that of the structural model (Fig. 7a).

Discussion

While PABPCs have been extensively studied and shown to play roles in translation initiation and mRNA decay in several plant species, and to mediate virus RNA translation in host-virus interactions [27-31], plant PABPNs have not yet been functionally or structurally characterized.

Here, we describe the first three-dimensional structure of the RNA-binding domain of a plant PABPN family member. The solution structure of the citrus CsPABPN1 RRM domain shows that the protein adopts a canonical RRM fold, with a poly(A)-binding mode that is also conserved among animal PABPNs. Our NMR titration studies revealed that poly(A) recognition by CsΔPABPN1 starts with residues of the RNP2 (β1 strand) and RNP1 (β3 strand) sites, followed by interactions with residues of the β2 strand, implicated in protein dimerization. Indeed, we show that CsΔPABPN1 can be found as monomer or dimer in solution, and that similar to the *Xenopus* XlePABP2 [13], undergoes a dimer-monomer transition upon poly(A) interaction. Our structural model of dimeric CsΔPABPN1 strongly suggests that dimerization of CsΔPABPN1 is, to some extent, mediated by the antiparallel pairing of the β2 strands, as demonstrated for the XlePABP2 protein [13]. However, yeast two-hybrid assays also indicate an important role of the N- and/or C-terminal domains in CsPABPN1 self-interaction, since the RRM domain alone was not sufficient for an interaction in yeast cells (Fig. 5d). This result is consistent with the observation that dimerization and nuclear localization of mammalian PABPN1 depend on the

arginine-rich C-terminal region and its methylated status, and that the nuclear import of yeast PAB2 requires a PY-NLS motif located in ~~its~~ the C-terminal end [8,10,11,37]. Given that the CsPABPN1 C-terminal region is also rich in arginines and carries a putative PY-NLS motif (Fig. 1a), it is likely that this region is important not only for CsPABPN1 dimerization but also for its nuclear import. In fact, deletion of the CsPABPN1 PY-NLS motif caused the protein to accumulate in the cytosol of *N. benthamiana* cells (Fig. 1b), suggesting that the mechanism for the nuclear targeting of CsPABPN1 is similar to that of yeast PAB2 [11].

The fact that CsPABPN1 undergoes a dimer-to-monomer transition upon poly(A) interaction prompted us to investigate the molecular basis of such transition. Based on our NMR titration and cross-linking experiments combined with the molecular modeling studies of monomeric and dimeric CsΔPABPN1, free and bound to poly(A), we propose a mechanistic model for the dimer-monomer transition. The well-formed concave and positively charged cavity of the joined RRM domains in the dimeric arrangement (Fig. 6) serves as a platform for the poly(A) binding. However, as the RNA docks into the dimer cleft, it experiences some clashes with L140 and D142, which belong to the β2 strands and help to stabilize the dimer. Notably, L140 is one of the first residues to show chemical shift changes upon RNA interaction (Fig. 4). The subsequent interactions of poly(A) with residues of the RNP sites would further favor dimer dissociation. However, as the cross-linking assay indicates, an equilibrium between monomeric and dimeric CsPABPN1 bound to a poly(A) might also exist in the cell, and depending upon these domain topologies, other mRNA processing factors may be recruited to the poly(A) tail. Thus, monomeric or dimeric PABPNs bound to a poly(A) tail could play distinct roles in processes like alternative cleavage, polyadenylation or mRNA decay.

An intriguing aspect of the dimeric arrangement of the RRM domains of the citrus and *Xenopus* PABPNs is that they adopt an antiparallel orientation relative to the adjacent RRM domains of PABPCs, which bind poly(A) in an extended conformation [12]. Considering that three CsΔPABPN1 molecules bound the poly(A_{15}) probe (Fig. 2a), we hypothesized that single RRM PABPs could bind to a poly(A) tail in a similar fashion as the joined RRMs of PABPCs. The structural model of multiple CsΔPABPN1

molecules bound to poly(A_{50}) generated here shows how multiple single RRM domains could wrap around the RNA to create a super helical shield (Fig. 7a). In this configuration, the single CsΔPABPN1 RRM domains display almost the same binding topology as the clustered RRM domains of PABPCs, indicating that PABPNs and PABPCs have similar poly(A)-binding modes. According to this idea, it is worth noting that accumulation and nuclear localization of PABPCs compensated the depletion of human PABPN1 [38].

The model of multiple CsΔPABPN1 RRM domains covering the poly(A) RNA is also consistent with the filamentous structures of CsΔPABPN1 (Fig. 7d) and calf thymus PABP2 bound to poly(A) observed by transmission electron microscopy [36]. The binding of PABPNs along the poly(A) tail is thought to protect the mRNA and serve as a scaffold for the assembly of other mRNA processing factors. In this regard, our model offers structural insights into how this macromolecular structure is assemble and stabilized. The model also predicts that the negatively charged N-terminal and positively charged C-terminal, oriented to the outside of the superhelix, could interact with each other or act as anchoring points for additional protein-protein interactions, as experimental data indicate [5,8,39]. In line with this idea, it is interesting to note that CsHMG, an interactor of PthA4 and CsPABPN1, has a basic N-terminal and an acidic DE-rich C-terminal region [32] that could be responsible for CsPABPN1 interaction.

The acidic DEH-rich motif of CsPABPN1 is also of particular note, since it was not found in other PABPs (Fig. 1a). The fact that *C. sinensis* has a CsPABPN1 homolog that lacks the DEH-rich motif (Fig. 1a) suggests that this motif plays a particular role in mRNA binding or protein-protein interaction. Interestingly, the DEH-rich motif of CsPABPN1 is similar to a C-terminal region of human MTERF4, a RNA-binding protein that has the same structural topology of the TAL effectors [40-44]. The acidic C-terminal of MTERF4 is thought to increase the binding affinity to RNA and to mediate protein-protein interactions [45].

In addition to CsPABPN1, CsHMG and PthA4 also bound poly(U) in gel shift assays [32]. We do not know yet if the interaction of CsPABPN1, CsHMG or PthA4 with poly(U) is of any biological relevance. However, it is interesting to note that oligouridylation of the 3' end of mRNAs has recently been shown to

function as a mark for global mRNA decay in mammalian cells, whereas in plant cells, oligouridylation protected olygoadenylated mRNAs from 3' end degradation [46-47]. In addition, PABPC1 suppressed uridylation of polyadenylated mRNAs in mammalian cells [46]. Since CsPABPN1 interacted with CsPABPC1 and CsTRAX [32], which are implicated in RNA-mediated deadenylation [19,48,49], it is possible that CsPABPN1 could recognize U-tails in uridylated mRNAs, or bind adjacently to such U tracks. We have compared the RRMs of PABPs and poly(U)-binding proteins, such as those of the human HnRNP C protein [50], and found that the RRMs of poly(A) and poly(U)-binding proteins have virtually the same protein folding [Supplementary Fig. S4]. Moreover, we found that the RNA-binding mode of human HnRNP C and CsΔPABPN1 is partially conserved and involves interactions with key residues of the RNP1 and RNP2 sites located in equivalent positions in both structures [Supplementary Fig. S4]. Although these are structural evidences indicating that the RRMs of PABPs and poly(U)-binding proteins have similar RNA-binding modes, the binding specificity or affinity for adenine or uridine can also be determined by protein regions adjacent to the RRMs. Accordingly, in human HnRNP C, there is an additional site of poly(U) interaction outside the RRM-binding cleft [50]. Because the RRM of CsΔPABPN1 was not sufficient for the interaction with poly(U) in gel shift assays, it is likely that the N or the C-terminal region of CsPABPN1 helps to stabilize the poly(U) interaction. Further studies will be required to prove this concept.

Material and Methods

Protein expression and purification

The full-length CsPABPN1 and its deletion derivative comprising residues 57-188 (CsΔPABPN1), which spans the RRM domain (Fig. 1a), were subcloned into the *Nde*I/*Not*I sites of pET28a (Novagen). The recombinant 6xHis-tagged proteins were expressed in *Escherichia coli* BL21 (DE3) cells and purified by metal affinity and size-exclusion chromatography. Cells were grown at 37°C in LB medium containing kanamycin (50 $mg.mL^{-1}$) to OD600nm = 0.6, followed by induction with 0.4 mM isopropyl-thio-β-D-galactopyranoside (IPTG) for 3 h. After centrifugation, cells were suspended in binding buffer (20 mM

Tris-HCl pH 8.0, 200 mM NaCl, 15 mM imidazole) and incubated on ice with lysozyme (1 $mg.mL^{-1}$) for 30 min. Bacterial cells were disrupted by sonication and the soluble fractions were incubated with cobalt resin for 4 h at 4°C under slow agitation. The beads were washed four times with two column volumes of 20 mM Tris-HCl pH 8.0, 200 mM NaCl, 15 mM imidazole, and retained proteins were eluted with the same buffer containing 500 mM imidazole.

For protein NMR assignments, uniformly ^{15}N- and ^{13}C-^{15}N-labeled CsΔPABPN1 was prepared by growing the *E. coli* cells at 37°C in M9 minimal medium supplemented with ^{15}N-ammonium chloride and/or ^{13}C-glucose (Cambridge Isotopes, Inc) to OD600nm = 0.6, followed by induction with 0.4 mM IPTG for 4 h. Labeled CsΔPABPN1 was purified as described above and dialyzed against 20 mM sodium phosphate buffer, pH 6.4, with 50 mM NaCl.

Analytical gel filtration chromatography

After affinity purification, protein samples were incubated with thrombin at 25°C for 16 h, for 6xHis tag removal, and subsequently dialyzed against 20 mM Tris-HCl pH 8.0, 50 mM NaCl. Protein samples were concentrated and further purified on a Superdex 75 10/300 GL column (GE Healthcare) equilibrated with 20 mM Tris-HCl, pH 8.0, 50 mM NaCl. The estimated molecular weight of CsΔPABPN1 was determined by elution volume according to pre-run standards (bovine serum albumin - 68 kDa, ovalbumin - 44 kDa, ribonuclease A - 13.7 kDa and aprotinin - 6.5 kDa).

CD measurements

Circular dichroism (CD) spectroscopy experiments were conducted on a Jasco J-810 CD spectrophotometer equipped with a Peltier temperature control using 1-mm path quartz cuvettes. Spectra were acquired with 5 μM of CsΔPABPN1 in 20 mM sodium phosphate, pH 6.4. CD measurements were collected between 195 and 260 nm using a scanning rate of 50 $nm.min^{-1}$, with an average response time of

4 s. An average of 10 scans per measurement was made and the background spectrum of the buffer was subtracted from the curves.

Chemical cross-linking

Affinity-purified CsΔPABPN1 was dialyzed against 20 mM HEPES buffer, pH 7.5, and used in cross-linking reactions, as described by Song et al. [13]. CsΔPABPN1 (100 μM) alone or in combination with poly(A_{15}) at 25 μM to 1 mM was incubated on ice for 15 min. The crosslinker dithio-bis-succinimidylpropionate (DSP) at 100 μM was added to the reaction mixture, and after 30 min incubation at room temperature, the cross-linking reactions were analyzed by SDS-PAGE. Full-length CsPABPN1 was expressed as a GST-fusion protein [32]. The protein was bound to glutathione sepharose resin 4B (GE Healthcare) and washed four times with PBS buffer, pH 7.4, 1 mM DTT. CsPABPN1 was cleaved from GST with thrombin (2U) and eluted with two column volumes of PBS, pH 7.4. Twenty μM of CsPABPN1 was incubated with increasing amounts of poly(A) or poly(U) RNA, and the crosslinker DSP was added to the reaction mixtures as described above. The cross-linking reactions were separated on a 10% SDS gel and transferred onto a PVDF membrane. CsPABPN1 was detected by Western blot using a 1:3000 anti-rabbit CsPABPN1 serum raised against the purified CsΔPABPN1.

Electrophoretic mobility shift assays (EMSA)

The oligoribonucleotides A_{15}, C_{15}, G_{15} and U_{15} were labeled with 1 U of T4 polynucleotide kinase (Fermentas) and 20 μCi of [γ^{32}P]-ATP. Labeled probes were purified and incubated for 20 min with CsPABPN1 or CsΔPABPN1 (100, 250 or 500 ng) in 20 μL reactions in binding buffer containing 10 mM Tris-HCl, pH 7.5, 7 μM $MgCl_2$, 0.1 μM EDTA, 0.6 μM DTT [51]. RNA-protein complexes were resolved on non-denaturing 6% polyacrylamide gels and exposed to radiographic films for visualization.

NMR spectroscopy and structure calculations

Experiments were performed using an Agilent Inova NMR spectrometer at the Multiuser NMR facility at the Brazilian National Biosciences Laboratory (LNBio/CNPEM), operating at a ^{1}H Larmor frequency of 599.887 MHz, at 298 K. The spectrometer is equipped with a triple resonance cryogenic probe and a Z pulse-field gradient unit. Isotopically labeled CsΔPABPN1 (400 μM) was solubilized in 20 mM sodium phosphate buffer, pH 6.4 and 50 mM NaCl. Deuterated water (D_2O, Cambridge Isotopes, Inc) was used at 10% (v/v) or 100% (v/v) final concentrations. Standard NMR experiments including ^{1}H-^{15}N-HSQC, HNCACB, CBCACONH, HNCO, HNCACO, HCCH-TOCSY and hCCH-TOCSY [52] were acquired for sequential assignments. NOE-derived distance restraints were obtained from the ^{15}N-HSQC-NOESY and ^{13}C-HSQC-NOESY (separately optimized for aliphatics and aromatics). The data were processed and analyzed using the NMRPipe/NMRView software [53,54]. The structure of CsΔPABPN1 was calculated in a semi-automated iterative manner with the program CYANA version 2.1 [55], using 100 starting conformers. CYANA 2.1 protocol was applied to calibrate and assign NOE cross-peaks. After the first few rounds of automatic calculations, the NOESY spectra were analyzed again to identify additional cross-peaks consistent with the structural model and to correct misidentified NOEs. The structures obtained were further refined by restrained minimization and molecular dynamic studies using the CNS software [56]. The 20 structures with the lowest target function were selected to represent the ensemble of protein structures. The quality of the structures was analyzed with PROCHECK-NMR [57]. For poly(A_{15})-binding experiments, the 2D ^{1}H-^{15}N-HSQC spectra were acquired on uniformly ^{15}N-labeled samples with different molar equivalent ratios. The tau c value of CsΔPABPN1 was calculated based on 2D ^{1}H-^{15}N-HSQC longitudinal (T1) and transverse (T2) relaxation times, according to Kay et al. [58].

Small angle X-ray scattering data collection and analysis

SAXS data were collected at the D11A-SAXS beamline at the Brazilian Synchrotron Light Laboratory (LNLS/CNPEM). The radiation wavelength was set to 1.488 Å and a charge-coupled device area detector (MARCCD 165 mm) was used to record the scattering patterns. The sample-to-detector distance was set to a scattering vector ranging from 0.10 to 2.2 nm^{-1}. Protein samples at 4 and 6 $mg.mL^{-1}$ were analyzed in

20 mM Tris-HCl, pH 7.5, 7 µM $MgCl_2$, 0.1 µM EDTA and 0.6 µM DTT in the presence of 50 (low ionic strength) or 500 mM NaCl (high ionic strength). Protein samples were centrifuged for 30 min at 10,000 x *g* to eliminate any existing aggregates immediately before each measurement. The scattering curves of the protein solutions and buffers were collected in frames of 300 s. Two successive frames were collected to monitor radiation-induced damage in protein samples. The radii of gyration (R*g*) were evaluated using Guinier approximation as implemented in the program PRIMUS [59]. The indirect Fourier transform package GNOM was used to evaluate the pair-distance distribution functions *p*(r). The low resolution envelopes of the CsΔPABPN1 protein under low and high salt conditions were determined using *ab initio* modeling implemented in DAMMIN [60]. Averaged models were generated from several runs using DAMAVER suite programs [61]. The SAXS models and the NMR structure were superimposed with SUPCOMB [62].

Transmission electron microscopy

CsΔPABPN1 purified by affinity and gel-filtration chromatographies, at a 30 µM final concentration, was incubated on ice for 2 h in binding buffer containing 10 mM Tris-HCl, pH 7.5, 7 µM $MgCl_2$, 0.1 µM EDTA, 0.6 µM DTT and 20 U Ribolock (Thermo Scientific, USA), in the presence and absence of 1.0 µM poly(A_{50}) (Invitrogen, USA). CsPABPN1-poly(A_{50}), CsPABPN1 or poly(A_{50}) alone were applied to glow-discharged (15 mA, negative charge for 25 seconds) 400 mesh copper grids covered by a thin layer of continuous carbon film (Ted Pella Inc., USA). After 1 min, the excess sample was blotted using a filter paper and the grid washed three times using the binding buffer. Uranyl acetate solution (2%) was applied twice to the grids for 30 seconds and the excess solution was blotted. Images were collected at room temperature in a Jeol JEM-2100 microscope operated at 200 kV and recorded on a TemCam F-416 4k × 4k CMOS camera (Tietz Video and Image Processing Systems, Germany), using a nominal magnification of 60,000× and 100,000× in a defocus range of −1 to −3 µm.

Molecular modeling and docking

Modeling of the CsΔPABPN1 dimer and molecular docking studies of CsΔPABPN1 in complex with poly(A_9) RNA were carried out using the YASARA software [59]. The CsΔPABPN1 dimer model was generated using the solution structure of *Xenopus laevis* embryonic XlePABP2 (PDB code 2JWN) [13]. The poly(A_9) RNA was docked into the RNA recognition motif of CsΔPABPN1 based on the crystallographic structures of human PABPC bound to poly(A) RNA (PDB codes 1CVJ and 4F02) [12, 34].

The macromolecular structural model of contiguous CsΔPABPN1 RRM bond poly(A_{50}) was built based on the 3D structure of human PABPC which have two adjacent RRMs bound to a poly(A_9) RNA (PDB code 1CVJ) [12]. The PABPC chain A was edited to remove the RRM linker residues 87-97 and the regions 11-86 (RRM1) and 98-179 (RRM2) were referred to chain A and B, respectively. The structure of CsΔPABPN1 RRM was then superimposed on chains A and B to generated a PDB file of the CsΔPABPN1 RRM dimer bound to poly(A_9). Following energy minimization and simulated annealing, the CsΔPABPN1 RRM dimer was duplicated using a Python algorithm and the chain A of the duplicated model was superimposed on the chain B of the template model. In the duplicated model, chain A was removed and chain B became chain C of the seminal CsΔPABPN1 RRM concatemer. Next, the chain A of the template dimer was superimposed on the chain C of the concatemer, and the same process was repeated to create a CsΔPABPN1 RRM concatemer to cover the length of a poly(A_{50}) molecule (Supplementary Data S1). The poly(A_{50}) molecule was also built using the atomic coordinates of the poly(A_9) of 1CVJ [12], following the same principal used to generate the CsΔPABPN1 RRM concatemer. The final structure of the CsΔPABPN1 RRM concatemer bound to poly(A_{50}) was energy-minimized to eliminate possible steric hindrances. A steepest descent protocol, with backbone atoms fixed, was ran until atom speed reached 3,000 m/s, and a simulated annealing protocol was carried out to zero K as target temperature. All these tasks were performed with YASARA [63].

Site-directed mutagenesis

Key residues involved in poly(A)-binding were each replaced by alanine (Y111A, N114A, R136A and F150A) by site-directed mutagenesis using the QuickChange Site-Directed Mutagenesis kit (Stratagene). All mutated constructs were verified by DNA sequencing. The mutated proteins, subcloned into pET28a, were expressed as 6xHis-tag fusions and purified as described above.

Nuclear localization of CsPABPN1

The full-length CsPABPN1 cDNA and a deletion derivative lacking the C-terminal PY-NLS motif (Fig. 1a) were subcloned into the pENTR-3C and pENTR/D-TOPO vectors (Invitrogen), respectively, and moved into the binary vector pRFP1 for the production of RFP-CsPABPN1 fusion proteins. *Nicotiana benthamiana* leaves were transiently transformed with the respective constructs via *Agrobacterium tumefaciens* infiltrations and the subcellular localization of CsPABPN1 and its PY-NLS deletion derivative were monitored by confocal fluorescent microscopy as described previously [64].

Accession numbers

The NMR restraints have been deposited in the Biological Magnetic Resonance Bank (BMRB) and Protein Data Bank (PDB) with the accession numbers 19167 and 2M70, respectively.

Acknowledgments

This work was supported by São Paulo Research Foundation (FAPESP grant 2011/20468-1). MND, ASS and CEB received fellowships from FAPESP and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). We also thank the technical help of Jaqueline da Silva Rodrigues, the NMR facility of the National Laboratory of Biosciences (LNBio)), the Electron Microscopy facility of the Brazilian Nanotechnology National Laboratory and the D11A-SAXS beamline of the Synchrotron Light Laboratory (LNLS) at the Brazilian Nacional Centre for Energy and Material Science (CNPEM).

Author Contributions

MND performed protein purification, mutagenesis, CD measurements, chemical crosslinking and EMSA assays and helped with NMR experiments; MLS conducted NMR spectroscopy and determined the PABPN1 structure; ASS performed cloning, protein purification and crosslinking assays; JL and AS performed protein localization studies; ACMZ helped with NMR experiments and structure calculations; TACBS, MND and MTM performed SAXS experiments and data analysis; AC and RVP conducted electron microscopy experiments; PSLO performed molecular modeling and dockings; CEB supervised the study and wrote the article.

Competing Financial Interests

The authors declare no competing financial interests

References

1- Mangus, D.A., Evans, M.C. & Jacobson, A. (2003). Poly(A)-binding proteins: muntifunctional scaffolds for the post-transcriptional control of gene expression. Genome Biol. 4, 223.

2- Goss, D.J. & Kleiman, F.E. (2013). Poly(A) binding proteins: are they all created equal? WIREs RNA 4, 167-179.

3- Lemay, J.F., D`Amours, A., Lemieux, C., Lackner, D.H., St-Sauveur, V.G., Bahler, J., *et al*. (2010). The nuclear poly(A)-binding protein interacts with the exosome to promote synthesis of noncoding small nuclear RNAs. Mol. Cell 37, 34-45.

4- Banerjee, A., Apponi, L.H., Pavlath, G.K. & Corbett, A.H. (2013). PABPN1: molecular function and muscle disease. FEBS J. 280, 4230-4250.

5- Kühn, U. & Pieler T. (1996). *Xenopus* poly(A) binding protein: functional domains in RNA binding and protein-protein interaction. J. Mol. Biol. 256, 20-30.

6- Kozlov, G., Trempe, J., Khaleghpour, K., Kahvejian, A., Ekiel, I. & Gehring, K. (2001) Structure and function of the C-terminal PABC domain of human poly(A)-binding protein. Proc. Natl. Acad. Sci. USA 98, 5809-5813.

7- Gorgoni, B., Richardson, W.A., Burgess, H.M., Anderson, R.C., Wilkie, G.S., Gautier, P., *et al*. (2011). Poly(A)-binding proteins are functionally distinct and have essential roles during vertebrate development. Proc. Natl. Acad. Sci. USA. 108, 7858-7849.

8- Kühn, U., Nemeth, A., Meyer, S. & Wahle, E. (2003). The RNA binding domains of the nuclear poly(A)-binding protein. J. Biol. Chem. 278, 16916-16925.

9- Kühn, U. & Wahle, E. (2004). Structure and function of poly(A) binding proteins. Biochim. Biophys. Acta 25, 67-84.

10- Fronz, K., Güttinger, S., Burkert, K., Kühn, U., Stöhr, N., Schierhorn, A., *et al*. (2011). Arginine methylation of the nuclear poly(A) binding protein weakens the interaction with its nuclear import receptor, transportin. J. Biol. Chem. 286, 32986-32994.

11- Mallet, P.L. & Bachand, F. (2013). A proline-tyrosine nuclear localization signal (PY-NLS) is required for the nuclear import of fission yeast PAB2, but not of human PABPN1. Traffic 14, 282-294.

12- Deo, R.C., Bonanno, J.B., Sonenberg, N. & Burley, S.K. (1999). Recognition of polyadenylate RNA by the poly(A)-binding protein. Cell 98, 835-845.

13- Song. J, McGivern, J.V., Nichols, K.W., Markley, J.L. & Sheets, M.D. (2008). Structural basis for RNA recognition by a type II poly(A)-binding protein. Proc. Natl. Acad. Sci. USA 105, 15317-15322.

14- Ge, H., Zhou, D., Tong, S., Gao, Y., Teng, M. & Niu, L. (2008) Crystal structure and possible dimerization of the single RRM of human PABPN1. Proteins 71, 1539-1545.

15- Jenal, M., Elkon, R., Loayza-Puch, F., van Haaften, G., Kühn, U., Menzies, F.M., *et al*. (2012). The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. Cell 149, 538-553.

16- Kahvejian, A., Roy, G. & Sonenberg, N. (2001). The mRNA closed-loop model: the function of PABP and PABP-interacting proteins in mRNA translation. Cold Spring Harb. Symp. Quant. Biol. 66, 293-300.

17- Kühn, U., Gündel, M., Knoth, A., Kerwitz, Y., Rüdel, S. & Wahle, E. (2009). Poly(A) tail length is controlled by the nuclear poly(A)-binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor. J. Biol. Chem. 284, 22803-22814.

18- Apponi, L.H., Leung, S.W., Williams, K.R., Valentini, S.R., Corbett, A.H. & Pavlath, G.K. (2010). Loss of nuclear poly(A)-binding protein 1 causes defects in myogenesis and mRNA biogenesis. Hum. Mol. Genet. 19, 1058-1065.

19- Fabian, M.R., Mathonnet, G., Sundermeier, T., Mathys, H., Zipprich, J.T, Svitkin YV, *et al.* (2009). Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. Mol. Cell 35, 1-13.

20- Fabian, M.R. & Sonenberg, N. (2012). The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. Nat. Struct. Mol. Biol. 19, 586-593.

21- Moretti, F., Kaiser, C., Zdanowicz-Specht, A. & Hentze, M.W. (2012). PABP and the poly(A) tail augment microRNA repression by facilitated miRISC binding. Nat. Struct. Mol. Biol. 19, 603-608.

22- de Klerk, E., Venema, A., Anvar, S.Y., Goeman, J.J., Hu, O., Trollet, C., Dickson, G., *et al.* (2012). Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation. Nucleic Acids Res. 40, 9089-9101.

23- Grenier St-Sauveur, V., Soucek, S., Corbett, A.H. & Bachand, F. (2013). Poly(A) tail-mediated gene regulation by opposing roles of Nab2 and Pab2 nuclear poly(A)-binding proteins in pre-mRNA decay. Mol. Cell Biol. 33, 4718-4731.

24- Bresson, S.M. & Conrad, N.K. (2013). The human nuclear poly(A)-binging protein promotes RNA hyperadenylation and decay. PLOS Genetics 9, e1003893.

25- Belostotsky, D.A. & Meagher, R.B. (1993). Differential organ-specific expression of three poly(A)-binding-protein genes from *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA 90, 6686-6690.

26- Belostotsky, D.A. (2003). Unexpected complexity of poly(A)-binding protein gene families in flowering plants, three conserved lineages that are at least 200 millions years old and possible auto- and cross-regulation. Genetics 163, 311-319.

27- Palanivelu, R., Belostotsky, D.A. & Meagher, R.B. (2000). *Arabidopsis thaliana* poly(A)-binding protein 2 (PAB2) functions in yeast translational and mRNA decay processes. Plant J. 22, 187-198.

28- Le, H., Browning, K.S. & Gallie, D.R. (2000). The phosphorylation state of poly(A)-binding protein specifies its binding to poly(A) RNA and its interaction with eukaryotic initiation factor (eIF) 4F, eIFiso4F, and eIF4B. J. Biol. Chem. 275, 17452-17462.

29- Dufresne, P.J., Ubalijoro, E., Fortin, M.G. & Laliberté, J.F. (2008). *Arabidopsis thaliana* class II poly(A)-binding proteins are required for efficient multiplication of turnip mosaic virus. J. Gen. Virol. 89, 2339-2348.

30- Marom, L., Hen-Avivi, S., Pinchasi, D., Chekanova, J.A., Belostotsky, D.A. & Elroy-Stein, O. (2009). Diverse poly(A) binding proteins mediate internal translational initiation by a plant viral IRES. RNA Biol. 6, 446-454.

31- Iwakawa, H.O., Tajima, Y., Taniguchi, T., Kaido, M., Mise, K., Tomari, Y., *et al*. (2012). Poly(A)-binding protein facilitates translation of an uncapped/nonpolyadenylated viral RNA by binding to the 3' untranslated region. J. Virol. 86, 7836-7849.

32- de Souza, T.A., Soprano, A.S., Lira, N.P.V., Quaresma, A.J.C., Pauletti, B.A., Leme, A.F.P., *et al*. (2012). The TAL Effector PthA4 Interacts with nuclear factors involved in RNA-dependent processes including a HMG protein that selectively binds poly(U) RNA. PLoS One 7, e32305.

33- Wu, G.A., Prochnik, S., Jenkins, J., Salse, J., Hellsten, U., Murat, F., *et al*. (2014). Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. Nat. Biotechnol. 32, 656-662.

34- Safaee, N., Kozlov, G., Noronha, A.M., Xie, J., Wilds, C.J. & Gehring, K.B. (2012). Interdomain allostery promotes assembly of the polyA mRNA complex with PABP and eIF4G. *Mol. Cell* **48**, 375-386.

35- Birney, E., Kumar, S. & Krainer, A. (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.

36- Keller, R.W., Kühn, U., Aragón, M., Bornikova, L., Wahle, E. & Bear, D.G. (2000). The nuclear poly(A) binding protein, PABP2, forms an oligomeric particle covering the length of the poly(A) tail. J. Mol. Biol. 297, 569-583.

37- Fan, X., Dion, P., Laganiere, J., Brais, B. & Rouleau, G.A. (2001). Oligomerization of polyalanine expanded PABPN1 facilitates nuclear protein aggregation that is associated with cell death. Hum. Mol. Genet. 10, 2341-2351.

38- Bhattacharjee, R.B. & Bag, J. (2012). Depletion of nuclear poly(A) binding protein PABPN1 produces a compensatory response by cytoplasmic PABP4 and PABP5 in cultured human cells. PLoS One 7, e53036.

39- Kerwitz, Y., Kühn, U., Lilie, H., Knoth, A., Scheuermann, T., Friedrich, H., *et al*. (2003). Stimulation of poly(A) polymerase through a direct interaction with the nuclear poly(A) binding protein allosterically regulated by RNA. EMBO J. 22, 3705-3714.

40- Murakami, M.T., Sforça, M., Neves, J., Paiva, J., Domingues, M.N., Pereira, A.L.A., *et al*. (2010). The repeat domain of the type III effector protein PthA shows a TPR-like structure and undergoes conformational changes upon DNA interaction. Proteins 78, 3386–3395.

41- Cámara, Y., Asin-Cayuela, J., Park, C.B., Metodiev, M.D., Shi, Y., Ruzzenente B., *et al*. (2011). MTERF4 regulates translation by targeting the methyltransferase NSUN4 to the mammalian mitochondrial ribosome. Cell Metab. 13, 527-539.

42- Spåhr, H., Habermann, B., Gustafsson, C.M., Larsson, N.G. & Hallberg, B.M. (2012). Structure of the human MTERF4-NSUN4 protein complex that regulates mitochondrial ribosome biogenesis. Proc. Natl. Acad. Sci. USA 109, 15253-15258.

43- Mak, A.N., Bradley, P., Cernadas, R.A., Bogdanove, A.J. & Stoddard, B.L. (2012). The crystal structure of TAL effector PthXo1 bound to its DNA target. Science 335, 716-719.

44- Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.K., Shi, Y., *et al*. (2012). Structural basis for sequence-specific recognition of DNA by TAL effectors. Science 335, 720-723.

45- Yakubovskaya, E,, Guja, K.E., Mejia, E., Castano, S., Hambardjieva, E., Choi, W.S., *et al*. (2012). Structure of the essential MTERF4:NSUN4 protein complex reveals how an MTERF protein collaborates to facilitate rRNA modification. Structure 20, 1940-1947.

46- Lim, J., Ha, M., Chang, H., Kwon, S.C., Simanshu, D.K., Patel, D.J., *et al*. (2014). Uridylation by TUT4 and TUT7 marks mRNA for degradation. Cell 159, 1365-1376.

47- Sement, F,M,, Ferrier, E,, Zuber, H,, Merret, R,, Alioua, M,, Deragon, J,M,, *et al*. (2013) Uridylation prevents 3' trimming of oligoadenylated mRNAs. Nucleic Acids Res. 41, 7115-7127.

48- Liu, Y., Ye, X., Jiang, F., Liang, C., Chen, D., Peng, J., *et al*. (2009). C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. Science 325, 750-753.

49- Ye, X., Huang, N., Liu, Y., Paroo, Z., Huerta, C., Li, P., *et al*. (2011). Structure of C3PO and mechanism of human RISC activation. Nat. Struct. Mol. Biol. 18, 650-657.

50- Cieniková, Z., Damberger, F.F., Hall, J., Allain, F.H. & Maris, C. (2014). Structural and mechanistic insights into poly(uridine) tract recognition by the hnRNP C RNA recognition motif. J. Am. Chem. Soc. 136, 14536-14544.

51- Good, P.J., Abler, L., Herring, D. & Sheets, M.D. (2004). *Xenopus* embryonic poly(A) binding protein 2 (ePABP2) defines a new family of cytoplasmic poly(A) binding proteins expressed during the early stages of vertebrate development. Genesis 38,166-175.

52- Sattler, M., Schleucher, J. & Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog. Nucl. Magn. Reson. Spectrosc. 34, 93-158.

53 - Johnson, B.A. & Blevins, R.A. (1994). NMRView: a computer program for the visualization and analysis of NMR data. J. Biomol. NMR 4, 603-614.

54- Delaglio, F., Grzeisiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR 6, 277-293.

55- Güntert, P. (2004). Automated NMR structure calculation with CYANA. Methods Mol. Biol. 278, 353-378.

56- Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., *et al*. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D. Biol. Crystallogr. 54, 905-921.

57- Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R. & Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR 8, 477-486.

58- Kay, L.E., Torchia, D.A. & Bax, A. (1989) Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry 28: 8972-8979.

59- Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J. & Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J. Appl. Cryst. 36, 1277-1282.

60- Svergun, D.I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys J. 76, 2879-2886.

61- Volkov, V.V. & Svergun, D.I. (2003). Uniqueness of ab-initio shape determination in small-angle scattering. J. Appl. Cryst. 36, 860-864.

62- Kozin, M.B. & Svergun, D.I. (2001). Automated matching of high- and low-resolution structural models. J. Appl. Cryst. 34, 33-41.

63- Krieger, E. & Vriend, G. (2014). YASARA View - molecular graphics for all devices - from smartphones to workstations. Bioinformatics. 30, 2981-2982.

64- Conti, L., Price, G., O'Donnell, E., Schwessinger, B., Dominy, P. & Sadanandom, A. (2008). Small ubiquitin-like modifier proteases OVERLY TOLERANT TO SALT1 and -2 regulate salt stress responses in Arabidopsis. Plant Cell 20, 2894-2908.

Formatted: English (U.S.)

Formatted: English (U.S.)

Figure legends

Fig. 1. CsPABPN1 is homologous to type II PABPs and localizes to the nucleus of citrus protoplasts. (a) Amino acid sequence alignment of CsPABPN1 from sweet orange "Pera" cultivar with type II PABPs from *Citrus sinensis* (orange1.1g027515m and orange1.1g041633m), *Arabidopsis thaliana* (AtPABPN1), *Ricinus communis (*RcPABP*)*, *Lotus japonicus* (LjPABP), *Oryza sativa* (OsPABP), *Xenopus laevis* embryonic (XlePABP2) and *Homo sapiens* (HsPABPN1). Identical and conserved residues are highlighted in dark and light gray, respectively. The RRM domain of CsPABPN1 comprising the residues I110 to V180 is indicated. The secondary structural elements determined by NMR are shown above the alignment. Cylinders represent α helices and arrows β strands. The two conserved RNA binding sites, RNP2 and RNP1, are boxed. The putative nuclear localization signal (NLS-RX_6PY) is highlighted in the C-terminal end of CsPABPN1 and HsPABPN1. (b) Sub-cellular localization of RFP-CsPABPN1 (upper panel) and RFP-CsPABPN1-ΔNLS (middle panel) fusion proteins in *N. benthamiana* cells showing that CsPABPN1 is nucleoplasm located and that the deletion of its putative PY-NLS motif affects the protein nuclear targeting. RFP only control is shown on the bottom panel.

Fig. 2. Interaction of CsPABPN1 and CsΔPABPN1 with single-strand RNA probes. EMSA using ^{32}P-labeled single-strand RNA probes and increased amounts (100, 250 and 500 ng) of purified CsΔPABPN1 (a) or full-length CsPABPN1 (b). (a) Specific interaction of CsΔPABPN1 with poly(A_{15}) showing three major CsΔPABPN1:polyA_{15} complexes of distinct molecular weights (arrows), suggesting a 1:1, 2:1 and 3:1 CsΔPABPN1:polyA_{15} stoichiometry. The intensity of the shifted bands correlates with the amount of CsΔPABPN1 in the binding mix. (b) The full-length CsPABPN1 protein formed a single major complex of higher molecular weight with the poly(A_{15}) and poly(U_{15}) probes (arrows), suggesting an effect of the N- and/or C-terminal on the binding and specificity of CsPABPN1 to single-strand RNAs. Free probes are indicated by brackets.

Fig. 3. The solution structure of CsΔPABPN1. (a) Stereoscopic view of the backbone trace of the twenty lowest energy conformers representing the CsΔPABPN1 structure. The disordered N-terminal is not shown. (b) Superposition of CsΔPABPN1 (red) with the RRM domains of human PABPN1 (green) and *X. laevis* embryonic XlePABP2 (blue) proteins (PDB codes 3B4D and 2JWN, respectively). (c) Ribbon model of the CsPABPN1 RRM domain representing the secondary structure elements that were labeled taking into account the RRM domain only. The RNP2 and RNP1 are highlighted in blue and yellow, respectively. (d) Comparison of the RRM domains of CsPABPN1 (red), human PABPN1 (green) and *Xenopus* XlePABP2 (blue) with those of type I PABPC proteins (grey), including the PABPC RRM1/RRM2 (PDB code 1CVJ), HUD RRM1 (PDB code 1FXL), hnRNP A1 RRM1/RRM2 (PDB codes 1HA1 and 2UP1), SXL RRM1 (PDB codes 3SXL and 1BF7), P14 (PDB code 2F9D) and Daz-associated protein 1 RRM domains (PDB codes 2DGS and 2DH8). The variable loop3 (β2/β3 loop) adopts a different orientation in the PABPNs, relative to PABPCs.

Fig. 4. The molecular basis of CsPABPN1 poly(A) binding. (a) Superposition of ^{1}H-^{15}N HSQC spectra of CsΔPABPN1 in the absence (green) and presence (red) of poly(A_{15}) at a 1:20 poly(A) :CsPABPN1 ratio. The amino acid residues that underwent significant chemical shift changes upon RNA interaction are labeled. (b) Cartoon of the CsΔPABPN1 structure and its corresponding surface depicting the residues that showed chemical shift changes upon RNA interaction at 1:30 (blue), 1:20 (green) and 1:10 (magenta) poly(A):CsΔPABPN1 ratios. Most of these residues are located in the concave side of the β-sheet. (**c**) Electrostatic surface of CsΔPABPN1 showing that the concave side of the β-sheet is positively charged. (d) Structural model of CsΔPABPN1 bound to a poly(A_9) molecule (dark blue), showing the amino acid residues that likely contribute to RNA binding (magenta). (e) EMSA showing that the replacement of residues Y111, N114, R136 and F150, implicated in adenine binding, to alanines abrogated the binding of CsΔPABPN1 to poly(A_{15}).

Fig. 5. CsPABPN1 dimerizes in solution and yeast cells. (a) Analytical gel filtration showing that CsΔPABPN1 (red) has an estimated molecular weight of ~45 kDa, according to its elution time relative to the protein standards Ovalbumin (43 kDa), Ribonuclease A (13.7 kDa) and Aprotinin (6.5 kDa). (b) SAXS scattering and pair distance distribution function *p(r)* plots of CsΔPABPN1 analyzed under low (black curves) or high-salt (red curves) conditions. The corresponding d_{max} and R*g* values are indicated. (c) Cartoon depicting the superposition of the CsΔPABPN1 structure (green) fitted into the SAXS envelops determined under low (~~white~~top view) and high-salt conditions (~~magenta~~bottom view). (d) Yeast two-hybrid assays showing the strong self-interaction of full-length CsPABPN1, in comparison with CsΔPABPN1. pOBD-Protein + pOAD-Protein (1) and the respective controls pOBD-Protein + pOAD-empty (2) and pOBD-empty + pOAD-Protein (3) are indicated.

Fig. 6. CsΔPABPN1 undergoes a dimer to monomer transition in the presence of poly(A). (a) Cross-linking analysis of CsΔPABPN1 in the absence or presence of increased amounts of poly(A_{15}) showing a dimer (D) to monomer (M) conversion as the amount of poly(A_{15}) in the binding mix increases. (b) Dimeric model of CsΔPABPN1 (magenta and cyan) with secondary structure elements indicated. The dimer is stabilized by the antiparallel pairing of the β2 and β2' strands and by two salt bridges between R136 and D142, located in the ends of the opposing β2 strands. (c) Dimeric model of CsΔPABPN1 bound to poly(A_9), showing that the dimer arrangement does not preclude the interaction with the RNA molecule (dark blue). The electrostatic surface of the dimer in the absence of the ligand RNA, shown for comparison, depicts the positively charged cleft formed by the joined RRM domains. (d) Steric clashes of residues L140 and D142 of strand β2' (cyan) with two adjacent adenines (blue).

Fig. 7. Structural model of CsΔPABPN1 assembly on a poly(A) tail and formation of CsΔPABPN1 filaments in the presence of poly(A_{50}). (a) Proposed structural model for the assembly of multiple CsΔPABPN1 monomers (green and magenta) bound to a poly(A_{50}) strand (blue). The top view shows

fourteen CsΔPABPN1 monomers bound to poly(A_{50}) in a spiral arrangement, where seven monomers are sufficient to make one turn around the RNA molecule. The details of the model, including the electrostatic surface, are shown below, with side views depicting the spiral configuration of the CsΔPABPN1 polymer. (b) Structural alignment between two consecutive CsΔPABPN1 monomers with the human RRM1/2 domains of PABPC (cyan), showing that the CsΔPABPN1 monomers adopt almost the same conformation of PABPC in this macromolecular assembly. (c) Cartoon representation of the interaction interface between CsΔPABPN1 monomers depicting mainly the hydrogen bonds that stabilize the interaction. Four hydrogen bonds are observed between monomers A and B (R136-E169, Q125-E169, E122-K178, E122-T170), with two additional hydrogen bonds between monomers B and C (R136-V179, I139-Q176). (d) Transmission electron micrographs showing that CsΔPABPN1 forms filaments and particles in the presence of poly(A_{50}) (arrowheads). White bars correspond to 70 nm. The average diameter of the observed filaments (~4 nm) is consistent with the diameter of the structural model shown in panel (a).

Table legends

Table 1. NMR and refinement statistics for CsΔPABPN1

Structure and mechanism of dimer-monomer transition of a plant poly(A)-binding protein upon RNA interaction: insights into its poly(A) tail assembly

Mariane Noronha Domingues[1†], Mauricio Luis Sforça[1†], Adriana Santos Soprano[1], Jack Lee[2], Tatiana de Arruda Campos Brasil de Souza[3], Alexandre Cassago[4], Rodrigo Villares Portugal[4], Ana Carolina de Mattos Zeri[1], Mario Tyago Murakami[1], Ari Sadanandom[2], Paulo Sergio Lopes de Oliveira[1] and Celso Eduardo Benedetti[1*]

[1]Laboratório Nacional de Biociências, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, SP, CP6192, Brazil

[2]School of Biological and Biomedical Sciences, Durham University, Durham, DH1, UK

[3]Instituto Carlos Chagas, FIOCRUZ, Curitiba, PR, Brazil

[4]Laboratório Nacional de Nanotecnologia, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, SP, CP6192, Brazil

[†]Equal contributors

[*]Correspondence to: Celso E. Benedetti, Laboratório Nacional de Biociências, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, SP, CP6192, Brazil. Phone: +55 19 35121111; Fax: +55 19 35121006; Email: celso.benedetti@lnbio.cnpem.br.

Abstract

Poly(A)-binding proteins (PABPs) play crucial roles in mRNA biogenesis, stability, transport and translational control in most eukaryotic cells. Although animal PABPs are well-studied proteins, the biological role, three-dimensional structure and RNA-binding mode of plant PABPs remain largely uncharacterized. Here, we report the structural features and RNA-binding mode of a *Citrus sinensis* PABP (CsPABPN1). CsPABPN1 has a domain architecture of nuclear PABPs (PABPN) with a single RNA recognition motif (RRM) flanked by an acidic N-terminal and GRPF-rich C-terminal. The RRM domain of CsPABPN1 displays virtually the same three-dimensional structure and poly(A)-binding mode of animal PABPNs. However, while the CsPABPN1 RRM domain specifically binds poly(A), the full-length protein also binds poly(U). CsPABPN1 localizes to the nucleus of plant cells and undergoes a dimer-monomer transition upon poly(A) interaction. We show that poly(A) binding by CsPABPN1 begins with the recognition of the RNA-binding sites RNP1 and RNP2, followed by interactions with residues of the β2 strands, which stabilize the dimer, thus leading to dimer dissociation. Like human PABPN1, CsPABPN1 also seems to form filaments in the presence of poly(A). Based on these data, we propose a structural model in which contiguous CsPABPN1 RRM monomers wrap around the RNA molecule creating a super-helical structure that could not only shield the poly(A) tail, but also serve as a scaffold for the assembly of additional mRNA processing factors.

Keywords: CsPABPN1, poly(A)-binding protein, RRM domain, *Citrus sinensis*, dimer-monomer transition,

Introduction

Most eukaryotic mRNAs are post-transcriptionally modified by processes like capping, splicing, 3'-end cleavage, polyadenylation and deadenylation [1,2]. The process of polyadenylation involves the addition of a poly(A) tail to the 3'-end of mRNAs generating a scaffold for the binding of poly(A)-binding proteins (PABPs), the main regulatory proteins that interact with the poly(A) tail [1,2].

PABPs play multiple roles in post-transcriptional gene regulation, controlling not only the 3'-end processing, but also the stability, nuclear export, translational activity and decay of mRNAs [1-4]. Two evolutionary conserved PABPs exist in most eukaryotic cells: the cytoplasmic PABPCs and nuclear PABPNs [1,2].

PABPC1 is the most studied variant among the PABPC family members; it contains four copies of non-identical RNA recognition motifs (RRMs) in the N-terminal and a Pro-rich domain required for oligomerization and interaction with other proteins in the C-terminal [5-7]. In contrast, eukaryotic cells seem to have only one nuclear PABP (PABPN1), characterized by a coiled-coil N-terminal, a single internal RRM domain and a C-terminal region with a nuclear localization signal (NLS). The PABPN1 NLS is dependent on a Pro-Tyr dipeptide (PY-NLS) and a Gly-Arg-rich (GR) region that is also important for PABPN1 self-interaction [1,4,8-11].

Despite sharing structurally similar RRM domains, PABPC1 and PABPN1 have different domain architectures and play distinct roles in the life cycle or fate of mRNAs [1-4,12-15]. For instance, while PABPC1 has been implicated in the control of mRNA translation and miRNA-dependent deadenylation and gene silencing, PABPN1 regulates the interaction between Poly(A) polymerase and the cleavage and polyadenylation specific factor during poly(A) synthesis, controlling the length of the poly(A) tail and the processes of alternative cleavage [2,15-22]. More recently though, human PABPN1 and related yeast PAB2 have also been implicated in mRNA decay [23,24].

In contrast to animal cells, the biological roles played by plant PABPs are less clear. Flowering plants have multiple PABPs with four RRM domains, which are related to the mammalian PABPCs [25,26]. Some of them have been reported to interact with translation initiation factors, complement yeast

translational and mRNA decay processes, and to mediate translation of plant virus RNA [27-31]. However, plant PABPNs have not yet been functionally or structurally characterized.

In a previous protein-protein interaction study, we have identified two sweet orange (*Citrus sinensis*) PABPs, designated CsPABP1 and CsPABP2, as interacting partners of PthA4, the main transcriptional activator-like (TAL) effector of the citrus canker pathogen *Xanthomonas* citri [32]. CsPABP1 and CsPABP2, which are presumed to be the orthologs of mammalian PABPN1 and PABPC1, respectively, based on sequence similarities, also interacted with other PthA4-binding proteins implicated in chromatin remodeling, transcription regulation, mRNA processing and translational control [32]. In particular, both interacted with the citrus high mobility group protein CsHMG that selectively binds poly(U) RNA [32].

Here, to gain new insight into the three-dimensional structure and RNA-binding properties of CsPABP1, we solved the solution structure of a CsPABP1 deletion derivative carrying the entire RRM domain. We present evidence indicating that CsPABP1 is a dimer in solution and undergoes a dimer-to-monomer transition upon poly(A) binding. The identification of the key residues implicated in poly(A) interaction and dimer stabilization led us to propose a mechanism of dimer dissociation upon RNA binding. Our results also show that CsPABP1 localizes to the nucleus of plant cells. Collectively, our data reveal that CsPABP1, to our knowledge the first plant nuclear PABP to be characterized, indeed represents an authentic ortholog of the animal PABPNs and was thus renamed to CsPABPN1. Moreover, based on the three-dimensional structure of the CsPABPN1 RRM domain and electron microscopy images of the protein in the presence of poly(A), we propose a structural model where single RRM domains of CsPABPN1 arranged in tandem assemble into a super-helical structure that covers the poly(A) tail.

Results

CsPABPN1 is homologous to type II PABPs and localizes to the nucleus of plant cells

CsPABPN1 is a 234 amino acid protein bearing an acidic DE-rich N-terminal, a single central RRM domain and a basic GRPF-rich C-terminal region (Fig. 1a). This domain architecture is typical of type II or nuclear PABPs. CsPABPN1 is closely related to several uncharacterized plant PABPs belonging to

dicot species, and yet shares 54% and 43% identity with the *Xenopus* embryonic XlePABP2 and human PABPN1, respectively (Fig. 1a).

An interesting feature of CsPABPN1 is that it carries a DEH-rich insertion motif in the N-terminal region (27-DEHEHEHDQDHEHEHDADNENEED-50) that seems unique to citrus species, since this motif was not found in any of the PAPB sequences so far available in the Genbank (Fig. 1a). In the *Ricinus communis*, *Lotus japonicus* and *Arabidopsis thaliana* PABP orthologs, however, the DEH-rich motif is partially replaced by the SSRPEEEEEEVEEEYDE, DADADAEQQDHDFASNH and TEEYEEHGGEEGAAAGDEELE regions, respectively (Fig. 1a). In addition, in the *Citrus sinensis* genome [33], we found two genes encoding proteins related to the CsPABPN1 isolated from cultivar "Pera". The orange1.1g027515m gene encodes a protein that is nearly identical to CsPABPN1; nevertheless, it has a shorter C-terminal and a "DH" polymorphism within the DEH-rich motif. On the other hand, the second PAPB, encoded by the orange1.1g041633m gene, lacks the DEH-rich motif and shows a longer C-terminal (Fig. 1a).

Because CsPABPN1 has a domain architecture that is common to nuclear PABPs, and it carries a putative PY-NLS in the C-terminal end that is closely related to those of mammalian and yeast PABPNs [10,11] (Fig. 1a), we tested its subcellular localization in plant cells. We found that the full-length CsPABPN1 localizes to the nucleus of *Nicotiana benthamiana* cells, thus confirming that it belongs to the PABPN subfamily (Fig. 1b). In addition, CsPABPN1 without the PY-NLS motif was detected only in a dotted-like pattern in the cytosol of *N. benthamiana* cells (Fig. 1b), indicating that this motif, like in yeast [11], is required for the nuclear import of CsPABPN1.

The RRM domain of CsPABPN1 specifically binds poly(A), whereas full-length CsPABPN1 also binds poly(U)

RRM domains of PABPs are known to specifically bind poly(A) RNA [12,13,34]. To tested whether the CsPABPN1 RRM domain would show similar properties, a truncated version of CsPABPN1 (CsΔPABPN1), carrying the entire RRM domain (Fig. 1a), was used in RNA-binding assays with labeled

single-strand RNA probes. This truncation was designed base on secondary structure predictions, which indicated that the first 56 and last 46 residues flanking the CsPABPN1 RRM domain are unstructured. As expected, CsΔPABPN1 specifically bound the poly(A_{15}) probe in gel shift assays (Fig. 2a). Importantly, CsΔPABPN1 formed three major CsΔPABPN1:poly(A_{15}) complexes of increased molecular weight, which were more evident as the amount of CsΔPABPN1 in the binding mix increased (Fig. 2a). Given that single RRM domains bind up to five consecutive adenines [12,13,34], the three CsΔPABPN1:poly(A_{15}) complexes observed in Fig. 2a suggest a 1:1, 2:1 and 3:1 CsΔPABPN1:poly(A_{15}) stoichiometry.

In contrast, we found that, in addition to poly(A_{15}), full-length CsPABPN1 also bound poly(U_{15}) (Fig. 2b). Moreover, only a single major CsPABPN1:RNA complex of higher molecular weight was observed with the poly(A_{15}) and poly(U_{15}) probes, suggesting that the N- and/or C-terminal of CsPABPN1 affect the binding specificity of the protein to RNA molecules (Fig. 2b).

The RRM domain of CsPABPN1 adopts a canonical RRM fold

To investigate the structural features of CsPABPN1 and provide insights into how this protein binds poly(A), we solved the solution structure of CsΔPABPN1, since the full-length protein is fairly insoluble and could not be purified or concentrated in the amounts required for structural resolution. Circular dichroism (CD) measurements revealed that CsΔPABPN1 is structured in solution showing two major negative peaks at 208 and 222 nm, suggesting the presence of α-helices (Supplementary Fig. S1a). These data, and the fact that CsΔPABPN1 bound poly(A_{15}) (Fig. 2a), indicated that purified CsΔPABPN1 was suitable for structure resolution. Accordingly, ^{15}N-labeled CsΔPABPN1 showed excellent HSQC peak dispersion (Supplementary Fig. S1b).

A total of 2416 distance restraints, including 1023 medium and long NOEs, were used for structure calculations (Table 1). The twenty structures with the lowest energy were selected to represent the ensemble of protein structures (Fig. 3a), and the statistics show no significant violations of the molecular geometry parameters (Table 1). In addition, the Ramachandran plot analysis shows that 99.5% of *phi-psi*

angles are in the most favorable and additional allowed regions, with no angles in the disallowed region, whereas the root mean square deviations (RMSD) for the backbone and side chains are within expected values, indicating a good stereochemical quality of the ensemble (Table 1).

The first twenty N-terminal residues of CsΔPABPN1 were disordered, as revealed by their heteronuclear NOE values, chemical shifts and ^{15}N relaxation measurements, and thus were omitted from the structures represented in Fig. 3. In the structured N-terminal region, two extended α-helices precede the RRM domain of CsPABPN1 (Fig. 1a and 3b). These helices are not present or fully represented in the 3D structures of the RRM domains of human PABPN1 and *Xenopus* XlePABP2 [13,14], the only 3D structures of single-RRM domain PABPs reported to date (Fig. 3b).

On the other hand, as typically observed for RRM domains of PABPs [12-14], the CsPABPN1 RRM domain is composed of four-stranded antiparallel β-sheet spatially arranged and backed by two α-helices (Fig. 3c). The two conserved RNA-binding sites, known as RNP1 (K148 to F155) and RNP2 (I110 to V115) [12-14,34], correspond to the central β3 and β1 strands, respectively (Fig. 1a and 3c). A comparison of type I and type II PABP structures revealed that the RRM domain of CsPABPN1 shows the same topology of the RRM domains of PABPCs and PABPNs (Fig. 3d). However, as previously noticed for the human and *Xenopus* PABPNs [14], the so-called loop3 or variable loop [35], which joins strand β2 with β3, adopts a different orientation compared to those of PABPC RRMs (Fig. 3d).

The molecular basis of poly(A) binding in CsPABPN1

To investigate the residues implicated in CsΔPABPN1 binding to poly(A), we performed NMR titrations of ^{15}N-CsΔPABPN1 in the presence of poly(A_{15}). We found specific chemical shift changes in the ^{1}H-^{15}N-HSQC signals between poly(A)-bound and free protein (Fig. 4a). At a 1:30 poly(A):CsΔPABPN1 ratio, we observed significant changes in the chemical shifts of V112, N114, L140, G149, R183 and T184, followed by changes in I110, Y111, G113, C119, K148, Y152, V179, N185 and I186 at a 1:20 ratio (Fig. 4a and b). Not surprisingly, I110, Y111, V112, G113 and N114 comprise the RNP2 site, whereas K148,

G149 and Y152 belong to the RNP1 site (Fig. 1a). Increased amounts of poly(A) (1:10 ratio) led to additional changes in the chemical shifts of E103, V115, D116, Y117, T120, T133, R136, V137, T138, I139, T141, F150, E154, A160, L165, H173, K178, A181, and K182, thus confirming that poly(A) binding in CsΔPABPN1 is restricted to the concave side of the β-sheet (Fig. 4b), which has a large positively charged surface (Fig. 4c).

To provide further insights into how CsΔPABPN1 binds poly(A), we generated a structural model of CsΔPABPN1 in complex with poly(A_9), based on the 3D structures of the human PABPC RRM domains bound to poly(A) [12,34] and our NMR titration studies (Fig. 4d). According to this model, adenine recognition by CsΔPABPN1 is greatly mediated by Van der Waals contacts, hydrogen bonds and stacking interactions with the conserved residues of the RNP1 (F150, Y152, E154) and RNP2 (Y111, N114) sites, but also with K182 and residues of strand β2 (R136, T138) and helix α2 (Q99, K102) (Fig. 4d). Interestingly, Y111, N114, R136, F150, Y152 and E154 are among the residues that first underwent chemical shifts changes upon poly(A) interaction in our NMR titration experiments (Fig. 4a and b).

To verify these results, some of the residues implicated in adenine binding, including Y111, N114, R136 and F150, were replaced by alanines. As anticipated, we found that all the mutant proteins lost their ability to bind poly(A_{15}) in gel shift assays (Fig. 4e), thus corroborating our NMR studies and the structural model of the CsΔPABPN1-poly(A_9) complex.

Altogether, the results show that the poly(A) binding mode of the citrus CsΔPABPN1 is highly conserved with respect to that of *Xenopus* XlePABP2 and human PABPC RRM domains [12,13,34].

CsPABPN1 forms dimers in solution and yeast cells

CsΔPABPN1 was found as a monomer in the NMR experiments, as revealed by the tau-c value of 8.2 ± 0.5 ns (Supplementary Fig. S2). However, when the protein was analyzed by size-exclusion chromatography, it eluted as a single peak with an estimated molecular weight of approximately three times of its predicted size (Fig. 5a). Since XlePABP2 is a dimer in solution [13] and CsΔPABPN1 has an

extended and partially structured N-terminal region (Fig. 3a), we presumed that the CsΔPABPN1 elongated and partially unfolded N-terminal would likely affect its behavior in gel-filtration chromatography. To investigate this assumption, we analyzed the hydrodynamic behavior of CsΔPABPN1 by SAXS. Surprisingly, we found that under lower ionic strength conditions, as those of the NMR experiments, CsΔPABPN1 is a monomer in solution (Fig. 5b). In contrast, under higher salt conditions, the protein seems to dimerize (Fig. 5b). The change in size and shape of CsΔPABPN1 between the lower and higher ionic strength conditions can be observed by the corresponding scattering and pair distance distribution function *p(r)* curves (Fig. 5b). In addition, we found that while the CsΔPABPN1 NMR structure fitted remarkably well into the SAXS envelope model obtained under low-salt conditions, the SAXS envelope obtained under high-salt conditions showed a more elongated or extended shape consistent with a dimer (Fig. 5c).

To further investigate the dimeric nature of CsPABPN1, yeast two-hybrid assays were performed with CsΔPABPN1 and the full-length protein. We found that although the RRM domain of CsΔPABPN1 was not sufficient for a self-interaction, the full-length protein strongly self-interacted in yeast cells (Fig. 5d). This suggests that the negatively charged N-terminal or the C-terminal region, which carries the NLS and is required for bovine PABPN1 self-association [8], is important for CsPABPN1 self-interaction. The fact that full-length CsPABPN1 self-interacted in yeast is also in accord with the observation that it formed a single and higher molecular weight complex with the poly(A_{15}) compared with CsΔPABPN1 (Fig. 2). Taken together, the results show that CsPABPN1 can be found as monomer or dimer depending on the environmental conditions, thus suggesting a putative regulatory mechanism for poly(A) recognition and binding.

CsΔPABPN1 shows a dimer-monomer transition upon RNA binding

Given the overall structural similarities shared between CsΔPABPN1 and XlePABP2, we tested whether CsΔPABPN1 would also display a dimer-monomer transition upon RNA interaction [13]. Chemical cross-

linking was used to investigate the oligomeric state of CsΔPABPN1 in the presence and absence of poly(A_{15}). A cross-linking of the CsΔPABPN1 protein produced an approximately 32 kDa band observed in the SDS-PAGE gels, which is consistent with a CsΔPABPN1 dimer (Fig. 6a). However, in the presence of increased amounts of poly(A_{15}), a ~16 kDa band accumulated predominantly in the SDS-PAGE gels (Fig. 6a), showing that CsΔPABPN1 undergoes a dimer-monomer transition upon poly(A) binding. The full-length CsPABPN1 also undergoes a dimer-monomer transition upon poly(A) but not poly(U) interaction (Supplementary Fig. S3).

To understand how this transition might occur, we generated *in silico* models for dimeric CsΔPABPN1 free or bound to poly(A_9), based on the 3D structures of XlePABP2 (PDB code 2JWN) and human PABPC (PDB code 1CVJ), respectively [12,13]. The structural model of the dimer in the absence of RNA shows that, similar to XlePABP2, the CsΔPABPN1 dimer is stabilized by the antiparallel pairing of the β2 and β2' strands and by two salt bridges formed between R136 and D142, located in the ends of the opposing β2 strands (Fig. 6b). Consistent with the cross-linking assay showing that dimeric CsΔPABPN1 does not fully hinder poly(A) binding (Fig. 6a), the structural model of the CsΔPABPN1 dimer in complex with poly(A) suggests that the positively charged cleft created by the joined concave faces of the RRM domains would allow the binding of the poly(A) molecule (Fig. 6c). However, the dimer model also predicts steric clashes of two adjacent adenine bases with L140 and D142 of the β2' strand (Fig. 6d). Because the dimer is stabilized by the pairing of the β2 strands, and L140 is one of the first residues to show chemical shift changes upon poly(A) interaction (Fig. 4a), our results strongly indicate that poly(A) recognition by CsΔPABPN1 begins with the RNP sites, followed by interactions with the β2' residues (Fig. 4b and d), leading to dimer dissociation.

A proposed super-helical structure for CsPABPN1 bound to a poly(A) tail

As shown for XlePABP2 [13] and CsΔPABPN1 (Fig. 5d and 6b), dimerization of this group of PABPs is in part mediated by the antiparallel pairing of the β2 and β2' strands. However, in this dimeric

arrangement, the two RRM domains adopt an inverted orientation relative to the RRM domains of PABPCs, which are contiguously connected by a nine-residue linker [12]. In addition, in the crystal structure of the human PABPC (RRM1/2) bound to poly(A), the β-sheet surfaces of the RRM1 and 2 domains form a platform that binds poly(A) in an extended conformation of about four nucleotides per RRM domain [12]. Thus, considering that CsΔPABPN1 binds poly(A) preferentially as a monomer, and that the antiparallel dimer of CsΔPABPN1 is predicted to have steric clashes with adenine bases (Fig. 6d), it is expected that the RRM domains of multiple CsPABPN1 molecules would adopt a distinct arrangement when assembling into the poly(A) tail. Based on this assumption, and on the binding mode displayed by contiguous RRM domains of PABPCs, we generated a supra-macromolecular structural model for contiguous CsΔPABPN1 molecules bound to an extended poly(A_{50}) strand (Fig. 7a and Supplementary Data S1). According to this model, adjacent CsΔPABPN1 molecules adopt almost the same poly(A)-binding topology as the contiguous PABPC RRM1/2 domains [12], but with a small twist (Fig. 7b) that seems to favor the wrapping of CsΔPABPN1 molecules around the RNA. In fact, the model shows remarkably well that the tandem arranged CsΔPABPN1 monomers wrap around the RNA strand creating a superhelical shield (Fig. 7a and Supplementary Data S1). In addition, the model shows that seven CsΔPABPN1 molecules are sufficient to make one full turn around the poly(A) tail. This CsΔPABPN1 multimeric assembly is not only stabilized by interactions with the adenine bases but also by hydrogen bonds and van der Waals contacts between the CsΔPABPN1 monomers (Supplementary Data S1). Notably, the interaction interface between monomers A and B are similar to that of monomers B and C (Fig. 7c), indicating a symmetric arrangement of the CsΔPABPN1 monomers along the poly(A) tail. The model also suggests that this macromolecular assembly results from the polymerization of CsPABPN1 units upon poly(A) interaction and dimer disruption, allowing the formation of long filaments over the length of the poly(A) tail (Fig. 7a).

Because calf thymus PABP2 forms both linear filaments and oligomeric particles in the presence of poly(A) [36], we decide to investigate if CsΔPABPN1 could also form filaments or particles with

poly(A). Similar to calf thymus PABP2, CsΔPABPN1 formed both filaments and particles associated with filaments in the presence of poly(A_{50}) (Fig. 7d). These structures were not observed with either the protein or poly(A_{50}) alone, indicating that they are products of the CsΔPABPN1-poly(A_{50}) interaction. The average diameter of the observed filaments and particles are approximately 4 and 11 nm, respectively. The overall shape and dimensions of these structures are thus reminiscent of the 7 nm filaments and 21 nm particles of full-length PABP2, which is twice as big as CsΔPABPN1 [36]. Furthermore, the diameter of the observed CsΔPABPN1-poly(A_{50}) filaments is consistent with that of the structural model (Fig. 7a).

Discussion

While PABPCs have been extensively studied and shown to play roles in translation initiation and mRNA decay in several plant species, and to mediate virus RNA translation in host-virus interactions [27-31], plant PABPNs have not yet been functionally or structurally characterized.

Here, we describe the first three-dimensional structure of the RNA-binding domain of a plant PABPN family member. The solution structure of the citrus CsPABPN1 RRM domain shows that the protein adopts a canonical RRM fold, with a poly(A)-binding mode that is also conserved among animal PABPNs. Our NMR titration studies revealed that poly(A) recognition by CsΔPABPN1 starts with residues of the RNP2 (β1 strand) and RNP1 (β3 strand) sites, followed by interactions with residues of the β2 strand, implicated in protein dimerization. Indeed, we show that CsΔPABPN1 can be found as monomer or dimer in solution, and that similar to the *Xenopus* XlePABP2 [13], undergoes a dimer-monomer transition upon poly(A) interaction. Our structural model of dimeric CsΔPABPN1 strongly suggests that dimerization of CsΔPABPN1 is, to some extent, mediated by the antiparallel pairing of the β2 strands, as demonstrated for the XlePABP2 protein [13]. However, yeast two-hybrid assays also indicate an important role of the N- and/or C-terminal domains in CsPABPN1 self-interaction, since the RRM domain alone was not sufficient for an interaction in yeast cells (Fig. 5d). This result is consistent with the observation that dimerization and nuclear localization of mammalian PABPN1 depend on the

arginine-rich C-terminal region and its methylated status, and that the nuclear import of yeast PAB2 requires a PY-NLS motif located in the C-terminal end [8,10,11,37]. Given that the CsPABPN1 C-terminal region is also rich in arginines and carries a putative PY-NLS motif (Fig. 1a), it is likely that this region is important not only for CsPABPN1 dimerization but also for its nuclear import. In fact, deletion of the CsPABPN1 PY-NLS motif caused the protein to accumulate in the cytosol of *N. benthamiana* cells (Fig. 1b), suggesting that the mechanism for the nuclear targeting of CsPABPN1 is similar to that of yeast PAB2 [11].

The fact that CsPABPN1 undergoes a dimer-to-monomer transition upon poly(A) interaction prompted us to investigate the molecular basis of such transition. Based on our NMR titration and cross-linking experiments combined with the molecular modeling studies of monomeric and dimeric CsΔPABPN1, free and bound to poly(A), we propose a mechanistic model for the dimer-monomer transition. The well-formed concave and positively charged cavity of the joined RRM domains in the dimeric arrangement (Fig. 6) serves as a platform for the poly(A) binding. However, as the RNA docks into the dimer cleft, it experiences some clashes with L140 and D142, which belong to the β2 strands and help to stabilize the dimer. Notably, L140 is one of the first residues to show chemical shift changes upon RNA interaction (Fig. 4). The subsequent interactions of poly(A) with residues of the RNP sites would further favor dimer dissociation. However, as the cross-linking assay indicates, an equilibrium between monomeric and dimeric CsPABPN1 bound to a poly(A) might also exist in the cell, and depending upon these domain topologies, other mRNA processing factors may be recruited to the poly(A) tail. Thus, monomeric or dimeric PABPNs bound to a poly(A) tail could play distinct roles in processes like alternative cleavage, polyadenylation or mRNA decay.

An intriguing aspect of the dimeric arrangement of the RRM domains of the citrus and *Xenopus* PABPNs is that they adopt an antiparallel orientation relative to the adjacent RRM domains of PABPCs, which bind poly(A) in an extended conformation [12]. Considering that three CsΔPABPN1 molecules bound the poly(A_{15}) probe (Fig. 2a), we hypothesized that single RRM PABPs could bind to a poly(A) tail in a similar fashion as the joined RRMs of PABPCs. The structural model of multiple CsΔPABPN1

molecules bound to poly(A_{50}) generated here shows how multiple single RRM domains could wrap around the RNA to create a super helical shield (Fig. 7a). In this configuration, the single CsΔPABPN1 RRM domains display almost the same binding topology as the clustered RRM domains of PABPCs, indicating that PABPNs and PABPCs have similar poly(A)-binding modes. According to this idea, it is worth noting that accumulation and nuclear localization of PABPCs compensated the depletion of human PABPN1 [38].

The model of multiple CsΔPABPN1 RRM domains covering the poly(A) RNA is also consistent with the filamentous structures of CsΔPABPN1 (Fig. 7d) and calf thymus PABP2 bound to poly(A) observed by transmission electron microscopy [36]. The binding of PABPNs along the poly(A) tail is thought to protect the mRNA and serve as a scaffold for the assembly of other mRNA processing factors. In this regard, our model offers structural insights into how this macromolecular structure is assemble and stabilized. The model also predicts that the negatively charged N-terminal and positively charged C-terminal, oriented to the outside of the superhelix, could interact with each other or act as anchoring points for additional protein-protein interactions, as experimental data indicate [5,8,39]. In line with this idea, it is interesting to note that CsHMG, an interactor of PthA4 and CsPABPN1, has a basic N-terminal and an acidic DE-rich C-terminal region [32] that could be responsible for CsPABPN1 interaction.

The acidic DEH-rich motif of CsPABPN1 is also of particular note, since it was not found in other PABPs (Fig. 1a). The fact that *C. sinensis* has a CsPABPN1 homolog that lacks the DEH-rich motif (Fig. 1a) suggests that this motif plays a particular role in mRNA binding or protein-protein interaction. Interestingly, the DEH-rich motif of CsPABPN1 is similar to a C-terminal region of human MTERF4, a RNA-binding protein that has the same structural topology of the TAL effectors [40-44]. The acidic C-terminal of MTERF4 is thought to increase the binding affinity to RNA and to mediate protein-protein interactions [45].

In addition to CsPABPN1, CsHMG and PthA4 also bound poly(U) in gel shift assays [32]. We do not know yet if the interaction of CsPABPN1, CsHMG or PthA4 with poly(U) is of any biological relevance. However, it is interesting to note that oligouridylation of the 3’ end of mRNAs has recently been shown to

function as a mark for global mRNA decay in mammalian cells, whereas in plant cells, oligouridylation protected olygoadenylated mRNAs from 3' end degradation [46-47]. In addition, PABPC1 suppressed uridylation of polyadenylated mRNAs in mammalian cells [46]. Since CsPABPN1 interacted with CsPABPC1 and CsTRAX [32], which are implicated in RNA-mediated deadenylation [19,48,49], it is possible that CsPABPN1 could recognize U-tails in uridylated mRNAs, or bind adjacently to such U tracks. We have compared the RRMs of PABPs and poly(U)-binding proteins, such as those of the human HnRNP C protein [50], and found that the RRMs of poly(A) and poly(U)-binding proteins have virtually the same protein folding [Supplementary Fig. S4]. Moreover, we found that the RNA-binding mode of human HnRNP C and CsΔPABPN1 is partially conserved and involves interactions with key residues of the RNP1 and RNP2 sites located in equivalent positions in both structures [Supplementary Fig. S4]. Although these are structural evidences indicating that the RRMs of PABPs and poly(U)-binding proteins have similar RNA-binding modes, the binding specificity or affinity for adenine or uridine can also be determined by protein regions adjacent to the RRMs. Accordingly, in human HnRNP C, there is an additional site of poly(U) interaction outside the RRM-binding cleft [50]. Because the RRM of CsΔPABPN1 was not sufficient for the interaction with poly(U) in gel shift assays, it is likely that the N or the C-terminal region of CsPABPN1 helps to stabilize the poly(U) interaction. Further studies will be required to prove this concept.

Material and Methods

Protein expression and purification

The full-length CsPABPN1 and its deletion derivative comprising residues 57-188 (CsΔPABPN1), which spans the RRM domain (Fig. 1a), were subcloned into the *Ndel*I/*Not*I sites of pET28a (Novagen). The recombinant 6xHis-tagged proteins were expressed in *Escherichia coli* BL21 (DE3) cells and purified by metal affinity and size-exclusion chromatography. Cells were grown at 37°C in LB medium containing kanamycin (50 $mg.mL^{-1}$) to OD600nm = 0.6, followed by induction with 0.4 mM isopropyl-thio-β-D-galactopyranoside (IPTG) for 3 h. After centrifugation, cells were suspended in binding buffer (20 mM

Tris-HCl pH 8.0, 200 mM NaCl, 15 mM imidazole) and incubated on ice with lysozyme (1 $mg.mL^{-1}$) for 30 min. Bacterial cells were disrupted by sonication and the soluble fractions were incubated with cobalt resin for 4 h at 4°C under slow agitation. The beads were washed four times with two column volumes of 20 mM Tris-HCl pH 8.0, 200 mM NaCl, 15 mM imidazole, and retained proteins were eluted with the same buffer containing 500 mM imidazole.

For protein NMR assignments, uniformly ^{15}N- and ^{13}C-^{15}N-labeled CsΔPABPN1 was prepared by growing the *E. coli* cells at 37°C in M9 minimal medium supplemented with ^{15}N-ammonium chloride and/or ^{13}C-glucose (Cambridge Isotopes, Inc) to OD600nm = 0.6, followed by induction with 0.4 mM IPTG for 4 h. Labeled CsΔPABPN1 was purified as described above and dialyzed against 20 mM sodium phosphate buffer, pH 6.4, with 50 mM NaCl.

Analytical gel filtration chromatography

After affinity purification, protein samples were incubated with thrombin at 25°C for 16 h, for 6xHis tag removal, and subsequently dialyzed against 20 mM Tris-HCl pH 8.0, 50 mM NaCl. Protein samples were concentrated and further purified on a Superdex 75 10/300 GL column (GE Healthcare) equilibrated with 20 mM Tris-HCl, pH 8.0, 50 mM NaCl. The estimated molecular weight of CsΔPABPN1 was determined by elution volume according to pre-run standards (bovine serum albumin - 68 kDa, ovalbumin - 44 kDa, ribonuclease A - 13.7 kDa and aprotinin - 6.5 kDa).

CD measurements

Circular dichroism (CD) spectroscopy experiments were conducted on a Jasco J-810 CD spectrophotometer equipped with a Peltier temperature control using 1-mm path quartz cuvettes. Spectra were acquired with 5 μM of CsΔPABPN1 in 20 mM sodium phosphate, pH 6.4. CD measurements were collected between 195 and 260 nm using a scanning rate of 50 $nm.min^{-1}$, with an average response time of

4 s. An average of 10 scans per measurement was made and the background spectrum of the buffer was subtracted from the curves.

Chemical cross-linking

Affinity-purified CsΔPABPN1 was dialyzed against 20 mM HEPES buffer, pH 7.5, and used in cross-linking reactions, as described by Song et al. [13]. CsΔPABPN1 (100 μM) alone or in combination with poly(A_{15}) at 25 μM to 1 mM was incubated on ice for 15 min. The crosslinker dithio-bis-succinimidylpropionate (DSP) at 100 μM was added to the reaction mixture, and after 30 min incubation at room temperature, the cross-linking reactions were analyzed by SDS-PAGE. Full-length CsPABPN1 was expressed as a GST-fusion protein [32]. The protein was bound to glutathione sepharose resin 4B (GE Healthcare) and washed four times with PBS buffer, pH 7.4, 1 mM DTT. CsPABPN1 was cleaved from GST with thrombin (2U) and eluted with two column volumes of PBS, pH 7.4. Twenty μM of CsPABPN1 was incubated with increasing amounts of poly(A) or poly(U) RNA, and the crosslinker DSP was added to the reaction mixtures as described above. The cross-linking reactions were separated on a 10% SDS gel and transferred onto a PVDF membrane. CsPABPN1 was detected by Western blot using a 1:3000 anti-rabbit CsPABPN1 serum raised against the purified CsΔPABPN1.

Electrophoretic mobility shift assays (EMSA)

The oligoribonucleotides A_{15}, C_{15}, G_{15} and U_{15} were labeled with 1 U of T4 polynucleotide kinase (Fermentas) and 20 μCi of [γ^{32}P]-ATP. Labeled probes were purified and incubated for 20 min with CsPABPN1 or CsΔPABPN1 (100, 250 or 500 ng) in 20 μL reactions in binding buffer containing 10 mM Tris-HCl, pH 7.5, 7 μM $MgCl_2$, 0.1 μM EDTA, 0.6 μM DTT [51]. RNA-protein complexes were resolved on non-denaturing 6% polyacrylamide gels and exposed to radiographic films for visualization.

NMR spectroscopy and structure calculations

Experiments were performed using an Agilent Inova NMR spectrometer at the Multiuser NMR facility at the Brazilian National Biosciences Laboratory (LNBio/CNPEM), operating at a ^{1}H Larmor frequency of 599.887 MHz, at 298 K. The spectrometer is equipped with a triple resonance cryogenic probe and a Z pulse-field gradient unit. Isotopically labeled CsΔPABPN1 (400 μM) was solubilized in 20 mM sodium phosphate buffer, pH 6.4 and 50 mM NaCl. Deuterated water (D_2O, Cambridge Isotopes, Inc) was used at 10% (v/v) or 100% (v/v) final concentrations. Standard NMR experiments including ^{1}H-^{15}N-HSQC, HNCACB, CBCACONH, HNCO, HNCACO, HCCH-TOCSY and hCCH-TOCSY [52] were acquired for sequential assignments. NOE-derived distance restraints were obtained from the ^{15}N-HSQC-NOESY and ^{13}C-HSQC-NOESY (separately optimized for aliphatics and aromatics). The data were processed and analyzed using the NMRPipe/NMRView software [53,54]. The structure of CsΔPABPN1 was calculated in a semi-automated iterative manner with the program CYANA version 2.1 [55], using 100 starting conformers. CYANA 2.1 protocol was applied to calibrate and assign NOE cross-peaks. After the first few rounds of automatic calculations, the NOESY spectra were analyzed again to identify additional cross-peaks consistent with the structural model and to correct misidentified NOEs. The structures obtained were further refined by restrained minimization and molecular dynamic studies using the CNS software [56]. The 20 structures with the lowest target function were selected to represent the ensemble of protein structures. The quality of the structures was analyzed with PROCHECK-NMR [57]. For poly(A_{15})-binding experiments, the 2D ^{1}H-^{15}N-HSQC spectra were acquired on uniformly ^{15}N-labeled samples with different molar equivalent ratios. The tau c value of CsΔPABPN1 was calculated based on 2D ^{1}H-^{15}N-HSQC longitudinal (T1) and transverse (T2) relaxation times, according to Kay et al. [58].

Small angle X-ray scattering data collection and analysis

SAXS data were collected at the D11A-SAXS beamline at the Brazilian Synchrotron Light Laboratory (LNLS/CNPEM). The radiation wavelength was set to 1.488 Å and a charge-coupled device area detector (MARCCD 165 mm) was used to record the scattering patterns. The sample-to-detector distance was set to a scattering vector ranging from 0.10 to 2.2 nm^{-1}. Protein samples at 4 and 6 $mg.mL^{-1}$ were analyzed in

20 mM Tris-HCl, pH 7.5, 7 μM $MgCl_2$, 0.1 μM EDTA and 0.6 μM DTT in the presence of 50 (low ionic strength) or 500 mM NaCl (high ionic strength). Protein samples were centrifuged for 30 min at 10,000 x *g* to eliminate any existing aggregates immediately before each measurement. The scattering curves of the protein solutions and buffers were collected in frames of 300 s. Two successive frames were collected to monitor radiation-induced damage in protein samples. The radii of gyration (R*g*) were evaluated using Guinier approximation as implemented in the program PRIMUS [59]. The indirect Fourier transform package GNOM was used to evaluate the pair-distance distribution functions *p*(r). The low resolution envelopes of the CsΔPABPN1 protein under low and high salt conditions were determined using *ab initio* modeling implemented in DAMMIN [60]. Averaged models were generated from several runs using DAMAVER suite programs [61]. The SAXS models and the NMR structure were superimposed with SUPCOMB [62].

Transmission electron microscopy

CsΔPABPN1 purified by affinity and gel-filtration chromatographies, at a 30 μM final concentration, was incubated on ice for 2 h in binding buffer containing 10 mM Tris-HCl, pH 7.5, 7 μM $MgCl_2$, 0.1 μM EDTA, 0.6 μM DTT and 20 U Ribolock (Thermo Scientific, USA), in the presence and absence of 1.0 μM poly(A_{50}) (Invitrogen, USA). CsPABPN1-poly(A_{50}), CsPABPN1 or poly(A_{50}) alone were applied to glow-discharged (15 mA, negative charge for 25 seconds) 400 mesh copper grids covered by a thin layer of continuous carbon film (Ted Pella Inc., USA). After 1 min, the excess sample was blotted using a filter paper and the grid washed three times using the binding buffer. Uranyl acetate solution (2%) was applied twice to the grids for 30 seconds and the excess solution was blotted. Images were collected at room temperature in a Jeol JEM-2100 microscope operated at 200 kV and recorded on a TemCam F-416 4k × 4k CMOS camera (Tietz Video and Image Processing Systems, Germany), using a nominal magnification of 60,000× and 100,000× in a defocus range of −1 to −3 μm.

Molecular modeling and docking

Modeling of the CsΔPABPN1 dimer and molecular docking studies of CsΔPABPN1 in complex with poly(A_9) RNA were carried out using the YASARA software [59]. The CsΔPABPN1 dimer model was generated using the solution structure of *Xenopus laevis* embryonic XlePABP2 (PDB code 2JWN) [13]. The poly(A_9) RNA was docked into the RNA recognition motif of CsΔPABPN1 based on the crystallographic structures of human PABPC bound to poly(A) RNA (PDB codes 1CVJ and 4F02) [12, 34].

The macromolecular structural model of contiguous CsΔPABPN1 RRM bond poly(A_{50}) was built based on the 3D structure of human PABPC which have two adjacent RRMs bound to a poly(A_9) RNA (PDB code 1CVJ) [12]. The PABPC chain A was edited to remove the RRM linker residues 87-97 and the regions 11-86 (RRM1) and 98-179 (RRM2) were referred to chain A and B, respectively. The structure of CsΔPABPN1 RRM was then superimposed on chains A and B to generated a PDB file of the CsΔPABPN1 RRM dimer bound to poly(A_9). Following energy minimization and simulated annealing, the CsΔPABPN1 RRM dimer was duplicated using a Python algorithm and the chain A of the duplicated model was superimposed on the chain B of the template model. In the duplicated model, chain A was removed and chain B became chain C of the seminal CsΔPABPN1 RRM concatemer. Next, the chain A of the template dimer was superimposed on the chain C of the concatemer, and the same process was repeated to create a CsΔPABPN1 RRM concatemer to cover the length of a poly(A_{50}) molecule (Supplementary Data S1). The poly(A_{50}) molecule was also built using the atomic coordinates of the poly(A_9) of 1CVJ [12], following the same principal used to generate the CsΔPABPN1 RRM concatemer. The final structure of the CsΔPABPN1 RRM concatemer bound to poly(A_{50}) was energy-minimized to eliminate possible steric hindrances. A steepest descent protocol, with backbone atoms fixed, was ran until atom speed reached 3,000 m/s, and a simulated annealing protocol was carried out to zero K as target temperature. All these tasks were performed with YASARA [63].

Site-directed mutagenesis

Key residues involved in poly(A)-binding were each replaced by alanine (Y111A, N114A, R136A and F150A) by site-directed mutagenesis using the QuickChange Site-Directed Mutagenesis kit (Stratagene). All mutated constructs were verified by DNA sequencing. The mutated proteins, subcloned into pET28a, were expressed as 6xHis-tag fusions and purified as described above.

Nuclear localization of CsPABPN1

The full-length CsPABPN1 cDNA and a deletion derivative lacking the C-terminal PY-NLS motif (Fig. 1a) were subcloned into the pENTR-3C and pENTR/D-TOPO vectors (Invitrogen), respectively, and moved into the binary vector pRFP1 for the production of RFP-CsPABPN1 fusion proteins. *Nicotiana benthamiana* leaves were transiently transformed with the respective constructs via *Agrobacterium tumefaciens* infiltrations and the subcellular localization of CsPABPN1 and its PY-NLS deletion derivative were monitored by confocal fluorescent microscopy as described previously [64].

Accession numbers

The NMR restraints have been deposited in the Biological Magnetic Resonance Bank (BMRB) and Protein Data Bank (PDB) with the accession numbers 19167 and 2M70, respectively.

Acknowledgments

This work was supported by São Paulo Research Foundation (FAPESP grant 2011/20468-1). MND, ASS and CEB received fellowships from FAPESP and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). We also thank the technical help of Jaqueline da Silva Rodrigues, the NMR facility of the National Laboratory of Biosciences (LNBio)), the Electron Microscopy facility of the Brazilian Nanotechnology National Laboratory and the D11A-SAXS beamline of the Synchrotron Light Laboratory (LNLS) at the Brazilian Nacional Centre for Energy and Material Science (CNPEM).

Author Contributions

MND performed protein purification, mutagenesis, CD measurements, chemical crosslinking and EMSA assays and helped with NMR experiments; MLS conducted NMR spectroscopy and determined the PABPN1 structure; ASS performed cloning, protein purification and crosslinking assays; JL and AS performed protein localization studies; ACMZ helped with NMR experiments and structure calculations; TACBS, MND and MTM performed SAXS experiments and data analysis; AC and RVP conducted electron microscopy experiments; PSLO performed molecular modeling and dockings; CEB supervised the study and wrote the article.

Competing Financial Interests

The authors declare no competing financial interests

References

1- Mangus, D.A., Evans, M.C. & Jacobson, A. (2003). Poly(A)-binding proteins: muntifunctional scaffolds for the post-transcriptional control of gene expression. Genome Biol. 4, 223.

2- Goss, D.J. & Kleiman, F.E. (2013). Poly(A) binding proteins: are they all created equal? WIREs RNA 4, 167-179.

3- Lemay, J.F., D`Amours, A., Lemieux, C., Lackner, D.H., St-Sauveur, V.G., Bahler, J., *et al*. (2010). The nuclear poly(A)-binding protein interacts with the exosome to promote synthesis of noncoding small nuclear RNAs. Mol. Cell 37, 34-45.

4- Banerjee, A., Apponi, L.H., Pavlath, G.K. & Corbett, A.H. (2013). PABPN1: molecular function and muscle disease. FEBS J. 280, 4230-4250.

5- Kühn, U. & Pieler T. (1996). *Xenopus* poly(A) binding protein: functional domains in RNA binding and protein-protein interaction. J. Mol. Biol. 256, 20-30.

6- Kozlov, G., Trempe, J., Khaleghpour, K., Kahvejian, A., Ekiel, I. & Gehring, K. (2001) Structure and function of the C-terminal PABC domain of human poly(A)-binding protein. Proc. Natl. Acad. Sci. USA 98, 5809-5813.

7- Gorgoni, B., Richardson, W.A., Burgess, H.M., Anderson, R.C., Wilkie, G.S., Gautier, P., *et al*. (2011). Poly(A)-binding proteins are functionally distinct and have essential roles during vertebrate development. Proc. Natl. Acad. Sci. USA. 108, 7858-7849.

8- Kühn, U., Nemeth, A., Meyer, S. & Wahle, E. (2003). The RNA binding domains of the nuclear poly(A)-binding protein. J. Biol. Chem. 278, 16916-16925.

9- Kühn, U. & Wahle, E. (2004). Structure and function of poly(A) binding proteins. Biochim. Biophys. Acta 25, 67-84.

10- Fronz, K., Güttinger, S., Burkert, K., Kühn, U., Stöhr, N., Schierhorn, A., *et al*. (2011). Arginine methylation of the nuclear poly(A) binding protein weakens the interaction with its nuclear import receptor, transportin. J. Biol. Chem. 286, 32986-32994.

11- Mallet, P.L. & Bachand, F. (2013). A proline-tyrosine nuclear localization signal (PY-NLS) is required for the nuclear import of fission yeast PAB2, but not of human PABPN1. Traffic 14, 282-294.

12- Deo, R.C., Bonanno, J.B., Sonenberg, N. & Burley, S.K. (1999). Recognition of polyadenylate RNA by the poly(A)-binding protein. Cell 98, 835-845.

13- Song. J, McGivern, J.V., Nichols, K.W., Markley, J.L. & Sheets, M.D. (2008). Structural basis for RNA recognition by a type II poly(A)-binding protein. Proc. Natl. Acad. Sci. USA 105, 15317-15322.

14- Ge, H., Zhou, D., Tong, S., Gao, Y., Teng, M. & Niu, L. (2008) Crystal structure and possible dimerization of the single RRM of human PABPN1. Proteins 71, 1539-1545.

15- Jenal, M., Elkon, R., Loayza-Puch, F., van Haaften, G., Kühn, U., Menzies, F.M., *et al*. (2012). The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. Cell 149, 538-553.

16- Kahvejian, A., Roy, G. & Sonenberg, N. (2001). The mRNA closed-loop model: the function of PABP and PABP-interacting proteins in mRNA translation. Cold Spring Harb. Symp. Quant. Biol. 66, 293-300.

17- Kühn, U., Gündel, M., Knoth, A., Kerwitz, Y., Rüdel, S. & Wahle, E. (2009). Poly(A) tail length is controlled by the nuclear poly(A)-binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor. J. Biol. Chem. 284, 22803-22814.

18- Apponi, L.H., Leung, S.W., Williams, K.R., Valentini, S.R., Corbett, A.H. & Pavlath, G.K. (2010). Loss of nuclear poly(A)-binding protein 1 causes defects in myogenesis and mRNA biogenesis. Hum. Mol. Genet. 19, 1058-1065.

19- Fabian, M.R., Mathonnet, G., Sundermeier, T., Mathys, H., Zipprich, J.T, Svitkin YV, *et al*. (2009). Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. Mol. Cell 35, 1-13.

20- Fabian, M.R. & Sonenberg, N. (2012). The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. Nat. Struct. Mol. Biol. 19, 586-593.

21- Moretti, F., Kaiser, C., Zdanowicz-Specht, A. & Hentze, M.W. (2012). PABP and the poly(A) tail augment microRNA repression by facilitated miRISC binding. Nat. Struct. Mol. Biol. 19, 603-608.

22- de Klerk, E., Venema, A., Anvar, S.Y., Goeman, J.J., Hu, O., Trollet, C., Dickson, G., *et al*. (2012). Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation. Nucleic Acids Res. 40, 9089-9101.

23- Grenier St-Sauveur, V., Soucek, S., Corbett, A.H. & Bachand, F. (2013). Poly(A) tail-mediated gene regulation by opposing roles of Nab2 and Pab2 nuclear poly(A)-binding proteins in pre-mRNA decay. Mol. Cell Biol. 33, 4718-4731.

24- Bresson, S.M. & Conrad, N.K. (2013). The human nuclear poly(A)-binging protein promotes RNA hyperadenylation and decay. PLOS Genetics 9, e1003893.

25- Belostotsky, D.A. & Meagher, R.B. (1993). Differential organ-specific expression of three poly(A)-binding-protein genes from *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA 90, 6686-6690.

26- Belostotsky, D.A. (2003). Unexpected complexity of poly(A)-binding protein gene families in flowering plants, three conserved lineages that are at least 200 millions years old and possible auto- and cross-regulation. Genetics 163, 311-319.

27- Palanivelu, R., Belostotsky, D.A. & Meagher, R.B. (2000). *Arabidopsis thaliana* poly(A)-binding protein 2 (PAB2) functions in yeast translational and mRNA decay processes. Plant J. 22, 187-198.

28- Le, H., Browning, K.S. & Gallie, D.R. (2000). The phosphorylation state of poly(A)-binding protein specifies its binding to poly(A) RNA and its interaction with eukaryotic initiation factor (eIF) 4F, eIFiso4F, and eIF4B. J. Biol. Chem. 275, 17452-17462.

29- Dufresne, P.J., Ubalijoro, E., Fortin, M.G. & Laliberté, J.F. (2008). *Arabidopsis thaliana* class II poly(A)-binding proteins are required for efficient multiplication of turnip mosaic virus. J. Gen. Virol. 89, 2339-2348.

30- Marom, L., Hen-Avivi, S., Pinchasi, D., Chekanova, J.A., Belostotsky, D.A. & Elroy-Stein, O. (2009). Diverse poly(A) binding proteins mediate internal translational initiation by a plant viral IRES. RNA Biol. 6, 446-454.

31- Iwakawa, H.O., Tajima, Y., Taniguchi, T., Kaido, M., Mise, K., Tomari, Y., *et al*. (2012). Poly(A)-binding protein facilitates translation of an uncapped/nonpolyadenylated viral RNA by binding to the 3' untranslated region. J. Virol. 86, 7836-7849.

32- de Souza, T.A., Soprano, A.S., Lira, N.P.V., Quaresma, A.J.C., Pauletti, B.A., Leme, A.F.P., *et al*. (2012). The TAL Effector PthA4 Interacts with nuclear factors involved in RNA-dependent processes including a HMG protein that selectively binds poly(U) RNA. PLoS One 7, e32305.

33- Wu, G.A., Prochnik, S., Jenkins, J., Salse, J., Hellsten, U., Murat, F., *et al*. (2014). Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. Nat. Biotechnol. 32, 656-662.

34- Safaee, N., Kozlov, G., Noronha, A.M., Xie, J., Wilds, C.J. & Gehring, K.B. (2012). Interdomain allostery promotes assembly of the polyA mRNA complex with PABP and eIF4G. *Mol. Cell* **48**, 375-386.

35- Birney, E., Kumar, S. & Krainer, A. (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.

36- Keller, R.W., Kühn, U., Aragón, M., Bornikova, L., Wahle, E. & Bear, D.G. (2000). The nuclear poly(A) binding protein, PABP2, forms an oligomeric particle covering the length of the poly(A) tail. J. Mol. Biol. 297, 569-583.

37- Fan, X., Dion, P., Laganiere, J., Brais, B. & Rouleau, G.A. (2001). Oligomerization of polyalanine expanded PABPN1 facilitates nuclear protein aggregation that is associated with cell death. Hum. Mol. Genet. 10, 2341-2351.

38- Bhattacharjee, R.B. & Bag, J. (2012). Depletion of nuclear poly(A) binding protein PABPN1 produces a compensatory response by cytoplasmic PABP4 and PABP5 in cultured human cells. PLoS One 7, e53036.

39- Kerwitz, Y., Kühn, U., Lilie, H., Knoth, A., Scheuermann, T., Friedrich, H., *et al*. (2003). Stimulation of poly(A) polymerase through a direct interaction with the nuclear poly(A) binding protein allosterically regulated by RNA. EMBO J. 22, 3705-3714.

40- Murakami, M.T., Sforça, M., Neves, J., Paiva, J., Domingues, M.N., Pereira, A.L.A., *et al*. (2010). The repeat domain of the type III effector protein PthA shows a TPR-like structure and undergoes conformational changes upon DNA interaction. Proteins 78, 3386–3395.

41- Cámara, Y., Asin-Cayuela, J., Park, C.B., Metodiev, M.D., Shi, Y., Ruzzenente B., *et al*. (2011). MTERF4 regulates translation by targeting the methyltransferase NSUN4 to the mammalian mitochondrial ribosome. Cell Metab. 13, 527-539.

42- Spåhr, H., Habermann, B., Gustafsson, C.M., Larsson, N.G. & Hallberg, B.M. (2012). Structure of the human MTERF4-NSUN4 protein complex that regulates mitochondrial ribosome biogenesis. Proc. Natl. Acad. Sci. USA 109, 15253-15258.

43- Mak, A.N., Bradley, P., Cernadas, R.A., Bogdanove, A.J. & Stoddard, B.L. (2012). The crystal structure of TAL effector PthXo1 bound to its DNA target. Science 335, 716-719.

44- Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.K., Shi, Y., *et al*. (2012). Structural basis for sequence-specific recognition of DNA by TAL effectors. Science 335, 720-723.

45- Yakubovskaya, E,, Guja, K.E., Mejia, E., Castano, S., Hambardjieva, E., Choi, W.S., *et al*. (2012). Structure of the essential MTERF4:NSUN4 protein complex reveals how an MTERF protein collaborates to facilitate rRNA modification. Structure 20, 1940-1947.

46- Lim, J., Ha, M., Chang, H., Kwon, S.C., Simanshu, D.K., Patel, D.J., *et al*. (2014). Uridylation by TUT4 and TUT7 marks mRNA for degradation. Cell 159, 1365-1376.

47- Sement, F,M,, Ferrier, E,, Zuber, H,, Merret, R,, Alioua, M,, Deragon, J,M,, *et al*. (2013) Uridylation prevents 3' trimming of oligoadenylated mRNAs. Nucleic Acids Res. 41, 7115-7127.

48- Liu, Y., Ye, X., Jiang, F., Liang, C., Chen, D., Peng, J., *et al*. (2009). C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. Science 325, 750-753.

49- Ye, X., Huang, N., Liu, Y., Paroo, Z., Huerta, C., Li, P., *et al*. (2011). Structure of C3PO and mechanism of human RISC activation. Nat. Struct. Mol. Biol. 18, 650-657.

50- Cieniková, Z., Damberger, F.F., Hall, J., Allain, F.H. & Maris, C. (2014). Structural and mechanistic insights into poly(uridine) tract recognition by the hnRNP C RNA recognition motif. J. Am. Chem. Soc. 136, 14536-14544.

51- Good, P.J., Abler, L., Herring, D. & Sheets, M.D. (2004). *Xenopus* embryonic poly(A) binding protein 2 (ePABP2) defines a new family of cytoplasmic poly(A) binding proteins expressed during the early stages of vertebrate development. Genesis 38,166-175.

52- Sattler, M., Schleucher, J. & Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog. Nucl. Magn. Reson. Spectrosc. 34, 93-158.

53 - Johnson, B.A. & Blevins, R.A. (1994). NMRView: a computer program for the visualization and analysis of NMR data. J. Biomol. NMR 4, 603-614.

54- Delaglio, F., Grzeisiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR 6, 277-293.

55- Güntert, P. (2004). Automated NMR structure calculation with CYANA. Methods Mol. Biol. 278, 353-378.

56- Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., *et al*. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D. Biol. Crystallogr. 54, 905-921.

57- Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R. & Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR 8, 477-486.

58- Kay, L.E., Torchia, D.A. & Bax, A. (1989) Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry 28: 8972-8979.

59- Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J. & Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J. Appl. Cryst. 36, 1277-1282.

60- Svergun, D.I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys J. 76, 2879-2886.

61- Volkov, V.V. & Svergun, D.I. (2003). Uniqueness of ab-initio shape determination in small-angle scattering. J. Appl. Cryst. 36, 860-864.

62- Kozin, M.B. & Svergun, D.I. (2001). Automated matching of high- and low-resolution structural models. J. Appl. Cryst. 34, 33-41.

63- Krieger, E. & Vriend, G. (2014). YASARA View - molecular graphics for all devices - from smartphones to workstations. Bioinformatics. 30, 2981-2982.

64- Conti, L., Price, G., O'Donnell, E., Schwessinger, B., Dominy, P. & Sadanandom, A. (2008). Small ubiquitin-like modifier proteases OVERLY TOLERANT TO SALT1 and -2 regulate salt stress responses in Arabidopsis. Plant Cell 20, 2894-2908.

Figure legends

Fig. 1. CsPABPN1 is homologous to type II PABPs and localizes to the nucleus of citrus protoplasts. (a) Amino acid sequence alignment of CsPABPN1 from sweet orange "Pera" cultivar with type II PABPs from *Citrus sinensis* (orange1.1g027515m and orange1.1g041633m), *Arabidopsis thaliana* (AtPABPN1), *Ricinus communis (*RcPABP*)*, *Lotus japonicus* (LjPABP), *Oryza sativa* (OsPABP), *Xenopus laevis* embryonic (XlePABP2) and *Homo sapiens* (HsPABPN1). Identical and conserved residues are highlighted in dark and light gray, respectively. The RRM domain of CsPABPN1 comprising the residues I110 to V180 is indicated. The secondary structural elements determined by NMR are shown above the alignment. Cylinders represent α helices and arrows β strands. The two conserved RNA binding sites, RNP2 and RNP1, are boxed. The putative nuclear localization signal (NLS-RX_6PY) is highlighted in the C-terminal end of CsPABPN1 and HsPABPN1. (b) Sub-cellular localization of RFP-CsPABPN1 (upper panel) and RFP-CsPABPN1-ΔNLS (middle panel) fusion proteins in *N. benthamiana* cells showing that CsPABPN1 is nucleoplasm located and that the deletion of its putative PY-NLS motif affects the protein nuclear targeting. RFP only control is shown on the bottom panel.

Fig. 2. Interaction of CsPABPN1 and CsΔPABPN1 with single-strand RNA probes. EMSA using ^{32}P-labeled single-strand RNA probes and increased amounts (100, 250 and 500 ng) of purified CsΔPABPN1 (a) or full-length CsPABPN1 (b). (a) Specific interaction of CsΔPABPN1 with poly(A_{15}) showing three major CsΔPABPN1:polyA_{15} complexes of distinct molecular weights (arrows), suggesting a 1:1, 2:1 and 3:1 CsΔPABPN1:polyA_{15} stoichiometry. The intensity of the shifted bands correlates with the amount of CsΔPABPN1 in the binding mix. (b) The full-length CsPABPN1 protein formed a single major complex of higher molecular weight with the poly(A_{15}) and poly(U_{15}) probes (arrows), suggesting an effect of the N- and/or C-terminal on the binding and specificity of CsPABPN1 to single-strand RNAs. Free probes are indicated by brackets.

Fig. 3. The solution structure of CsΔPABPN1. (a) Stereoscopic view of the backbone trace of the twenty lowest energy conformers representing the CsΔPABPN1 structure. The disordered N-terminal is not shown. (b) Superposition of CsΔPABPN1 (red) with the RRM domains of human PABPN1 (green) and *X. laevis* embryonic XlePABP2 (blue) proteins (PDB codes 3B4D and 2JWN, respectively). (c) Ribbon model of the CsPABPN1 RRM domain representing the secondary structure elements that were labeled taking into account the RRM domain only. The RNP2 and RNP1 are highlighted in blue and yellow, respectively. (d) Comparison of the RRM domains of CsPABPN1 (red), human PABPN1 (green) and *Xenopus* XlePABP2 (blue) with those of type I PABPC proteins (grey), including the PABPC RRM1/RRM2 (PDB code 1CVJ), HUD RRM1 (PDB code 1FXL), hnRNP A1 RRM1/RRM2 (PDB codes 1HA1 and 2UP1), SXL RRM1 (PDB codes 3SXL and 1BF7), P14 (PDB code 2F9D) and Daz-associated protein 1 RRM domains (PDB codes 2DGS and 2DH8). The variable loop3 (β2/β3 loop) adopts a different orientation in the PABPNs, relative to PABPCs.

Fig. 4. The molecular basis of CsPABPN1 poly(A) binding. (a) Superposition of ^{1}H-^{15}N HSQC spectra of CsΔPABPN1 in the absence (green) and presence (red) of poly(A_{15}) at a 1:20 poly(A) :CsPABPN1 ratio. The amino acid residues that underwent significant chemical shift changes upon RNA interaction are labeled. (b) Cartoon of the CsΔPABPN1 structure and its corresponding surface depicting the residues that showed chemical shift changes upon RNA interaction at 1:30 (blue), 1:20 (green) and 1:10 (magenta) poly(A):CsΔPABPN1 ratios. Most of these residues are located in the concave side of the β-sheet. (**c**) Electrostatic surface of CsΔPABPN1 showing that the concave side of the β-sheet is positively charged. (d) Structural model of CsΔPABPN1 bound to a poly(A_9) molecule (dark blue), showing the amino acid residues that likely contribute to RNA binding (magenta). (e) EMSA showing that the replacement of residues Y111, N114, R136 and F150, implicated in adenine binding, to alanines abrogated the binding of CsΔPABPN1 to poly(A_{15}).

Fig. 5. CsPABPN1 dimerizes in solution and yeast cells. (a) Analytical gel filtration showing that CsΔPABPN1 (red) has an estimated molecular weight of ~45 kDa, according to its elution time relative to the protein standards Ovalbumin (43 kDa), Ribonuclease A (13.7 kDa) and Aprotinin (6.5 kDa). (b) SAXS scattering and pair distance distribution function *p(r)* plots of CsΔPABPN1 analyzed under low (black curves) or high-salt (red curves) conditions. The corresponding d_{max} and R*g* values are indicated. (c) Cartoon depicting the superposition of the CsΔPABPN1 structure (green) fitted into the SAXS envelops determined under low (top view) and high-salt conditions (bottom view). (d) Yeast two-hybrid assays showing the strong self-interaction of full-length CsPABPN1, in comparison with CsΔPABPN1. pOBD-Protein + pOAD-Protein (1) and the respective controls pOBD-Protein + pOAD-empty (2) and pOBD-empty + pOAD-Protein (3) are indicated.

Fig. 6. CsΔPABPN1 undergoes a dimer to monomer transition in the presence of poly(A). (a) Cross-linking analysis of CsΔPABPN1 in the absence or presence of increased amounts of poly(A_{15}) showing a dimer (D) to monomer (M) conversion as the amount of poly(A_{15}) in the binding mix increases. (b) Dimeric model of CsΔPABPN1 (magenta and cyan) with secondary structure elements indicated. The dimer is stabilized by the antiparallel pairing of the β2 and β2' strands and by two salt bridges between R136 and D142, located in the ends of the opposing β2 strands. (c) Dimeric model of CsΔPABPN1 bound to poly(A_9), showing that the dimer arrangement does not preclude the interaction with the RNA molecule (dark blue). The electrostatic surface of the dimer in the absence of the ligand RNA, shown for comparison, depicts the positively charged cleft formed by the joined RRM domains. (d) Steric clashes of residues L140 and D142 of strand β2' (cyan) with two adjacent adenines (blue).

Fig. 7. Structural model of CsΔPABPN1 assembly on a poly(A) tail and formation of CsΔPABPN1 filaments in the presence of poly(A_{50}). (a) Proposed structural model for the assembly of multiple CsΔPABPN1 monomers (green and magenta) bound to a poly(A_{50}) strand (blue). The top view shows

fourteen CsΔPABPN1 monomers bound to poly(A_{50}) in a spiral arrangement, where seven monomers are sufficient to make one turn around the RNA molecule. The details of the model, including the electrostatic surface, are shown below, with side views depicting the spiral configuration of the CsΔPABPN1 polymer. (b) Structural alignment between two consecutive CsΔPABPN1 monomers with the human RRM1/2 domains of PABPC (cyan), showing that the CsΔPABPN1 monomers adopt almost the same conformation of PABPC in this macromolecular assembly. (c) Cartoon representation of the interaction interface between CsΔPABPN1 monomers depicting mainly the hydrogen bonds that stabilize the interaction. Four hydrogen bonds are observed between monomers A and B (R136-E169, Q125-E169, E122-K178, E122-T170), with two additional hydrogen bonds between monomers B and C (R136-V179, I139-Q176). (d) Transmission electron micrographs showing that CsΔPABPN1 forms filaments and particles in the presence of poly(A_{50}) (arrowheads). White bars correspond to 70 nm. The average diameter of the observed filaments (~4 nm) is consistent with the diameter of the structural model shown in panel (a).

Table legends

Table 1. NMR and refinement statistics for CsΔPABPN1

Table 1. NMR and refinement statistics for CsΔPABPN1

Structural Statistics	
NOE	2416
Intra-residues (i,i)	708
Sequential (i, i+1)	685
Medium range (i, i+2 to i+4)	524
Long range (i, i≥5)	499
Rmsd from the average structure	
Backbone (residues 40-150)	0.83 ± 0.44
Heavy atoms (residues 40-150)	1.06 ± 0.35
Ramachandran plot	
Most favored regions (%)	79.5
Additional allowed regions (%)	20.0
Generously allowed regions (%)	0.5
Disallowed regions (%)	0.0
CYANA	
Target function (Å)	1.19 ± 0.12
Distance violation >0.20	0
Dihedral angles >5Å	0

Figure 1

Click here to download high resolution image

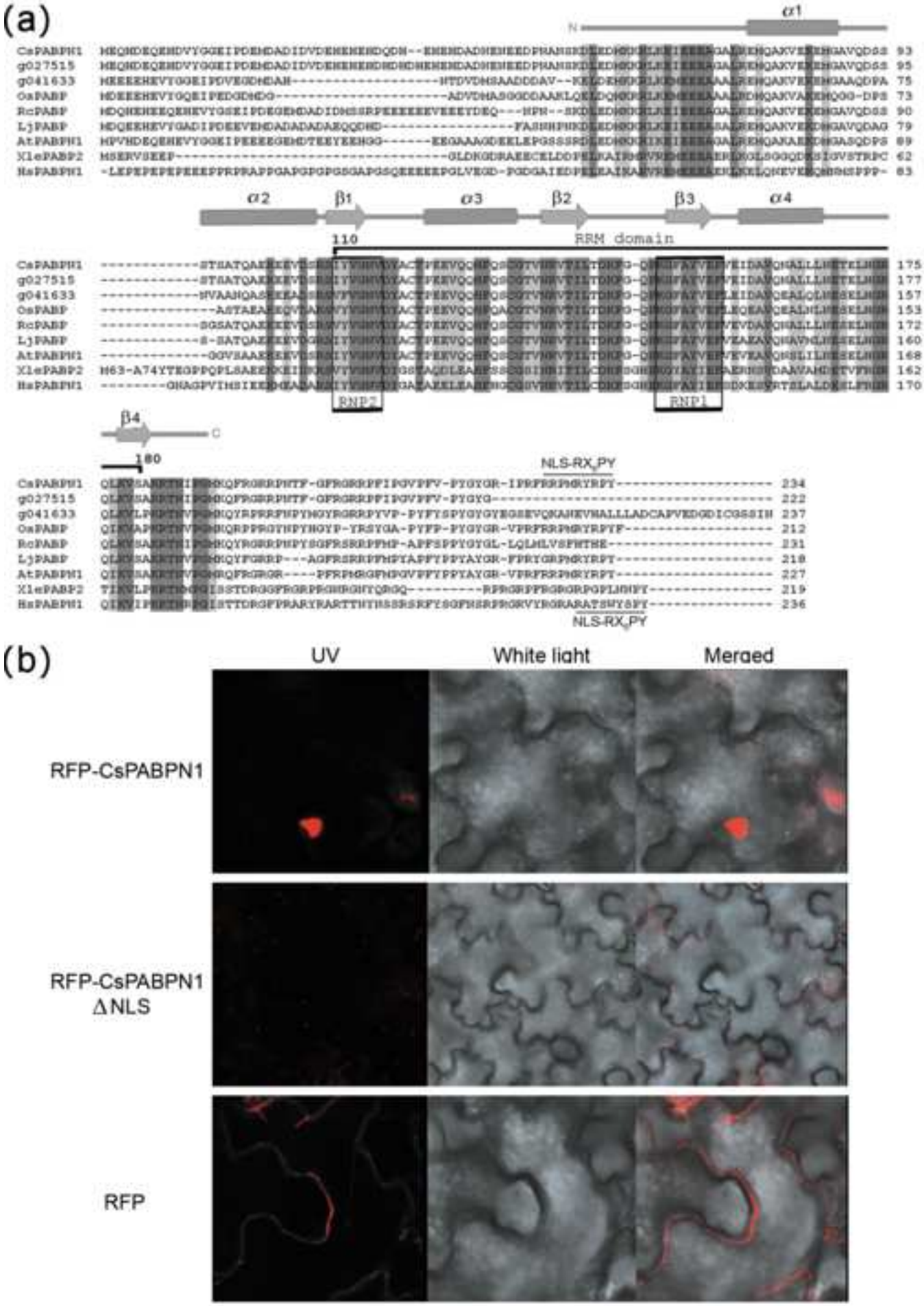


Figure 2

Click here to download high resolution image

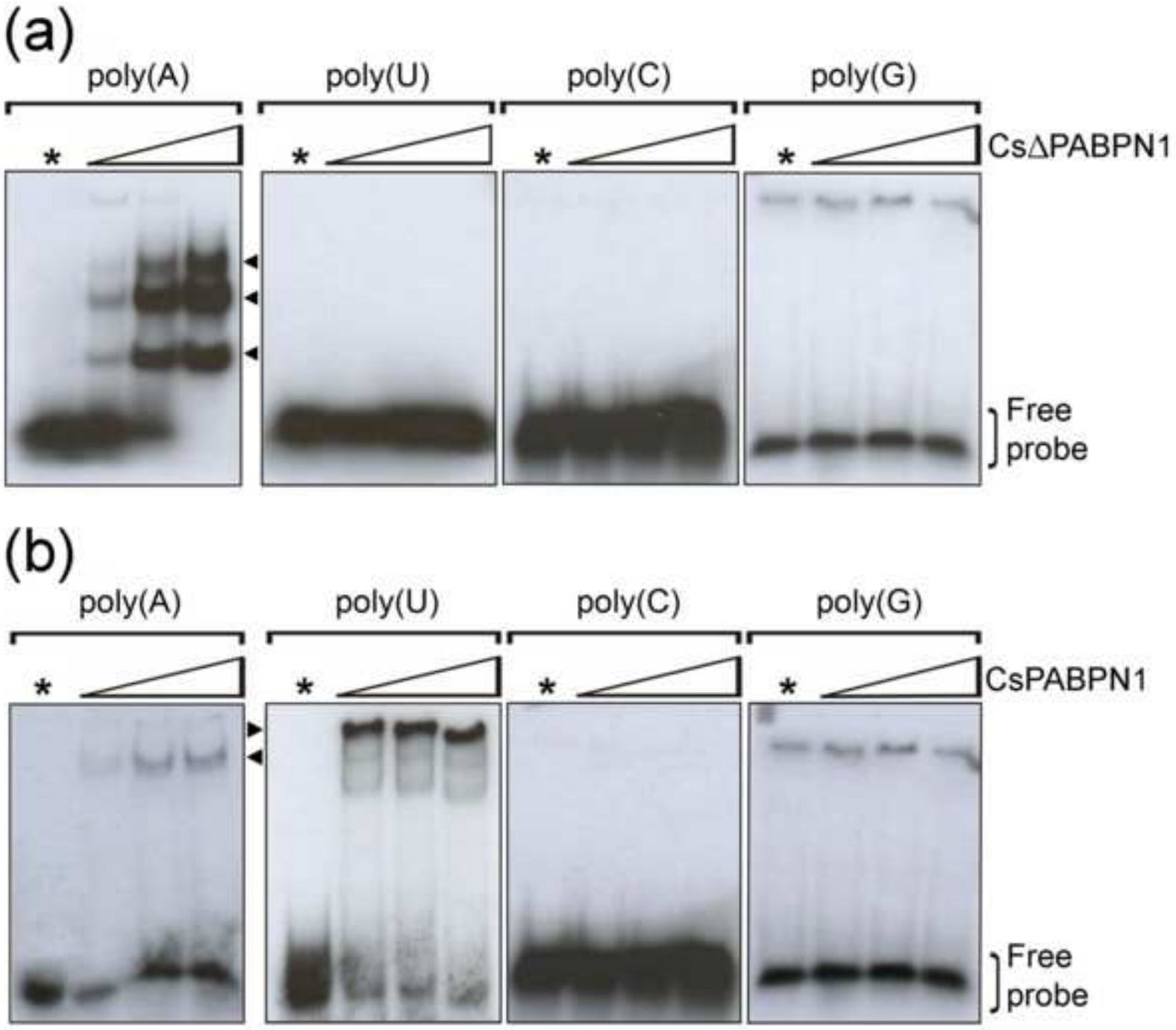


Figure 3

Click here to download high resolution image

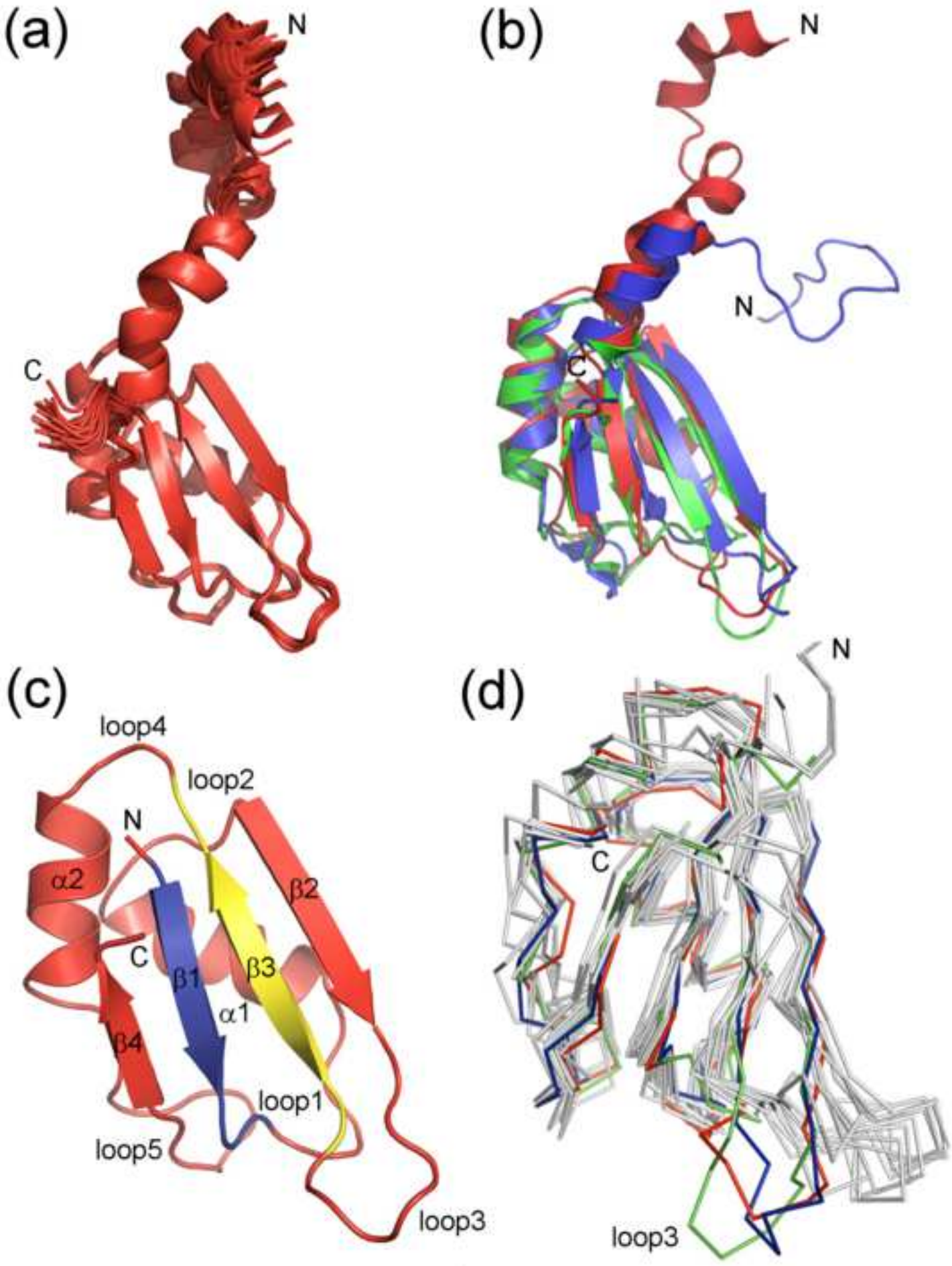


Figure 4

Click here to download high resolution image

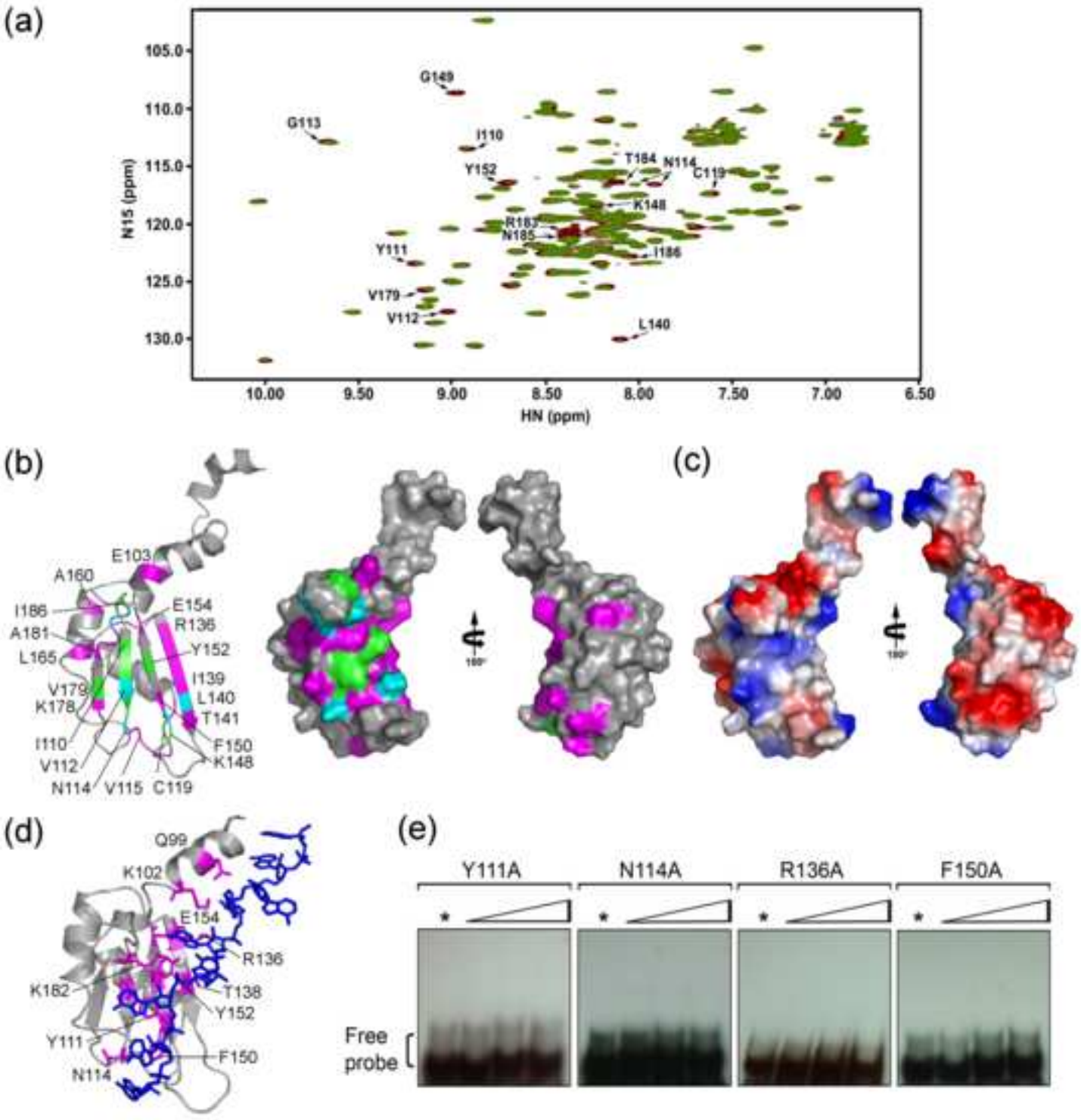


Figure 5

Click here to download high resolution image

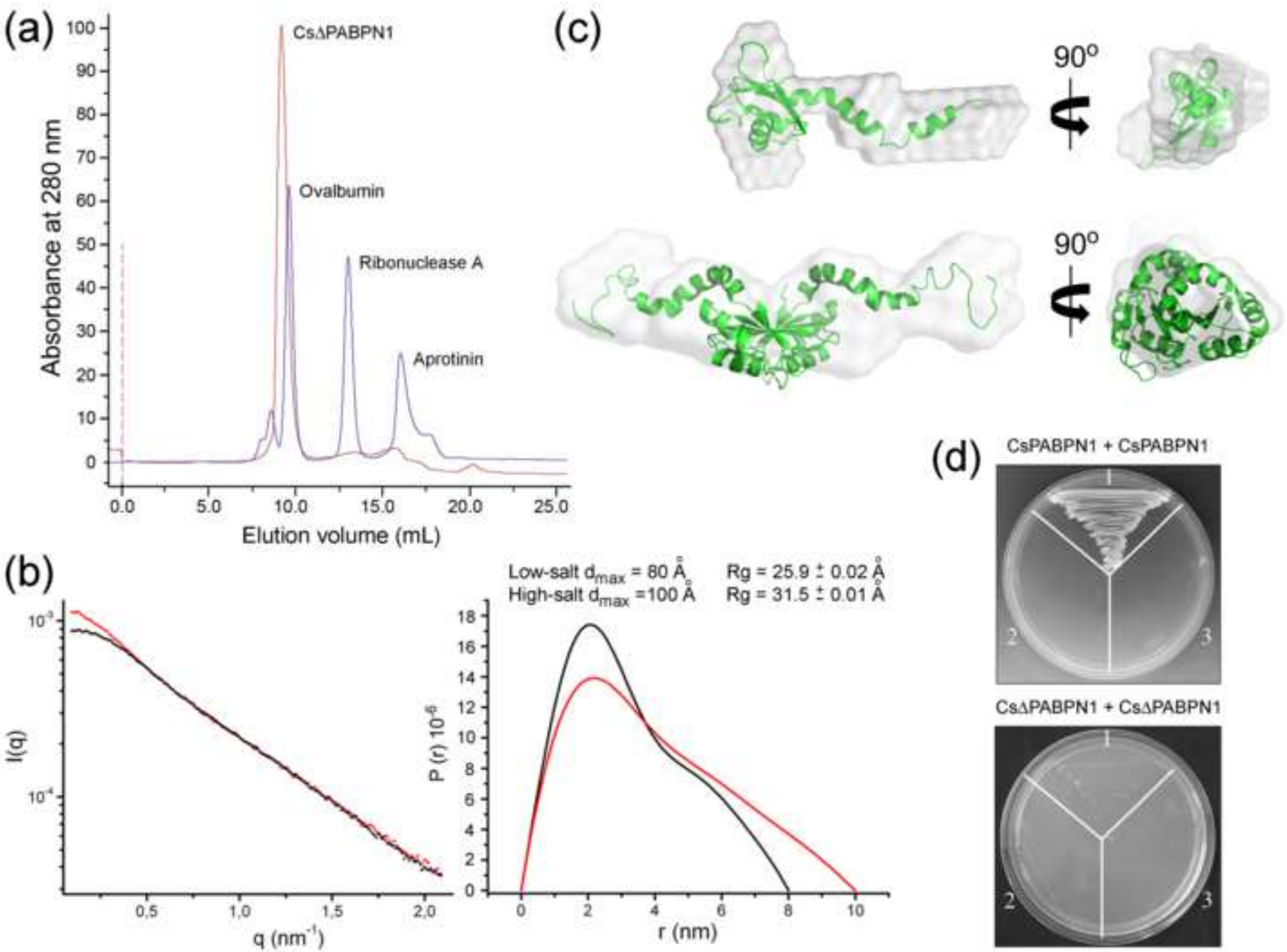


Figure 6

Click here to download high resolution image

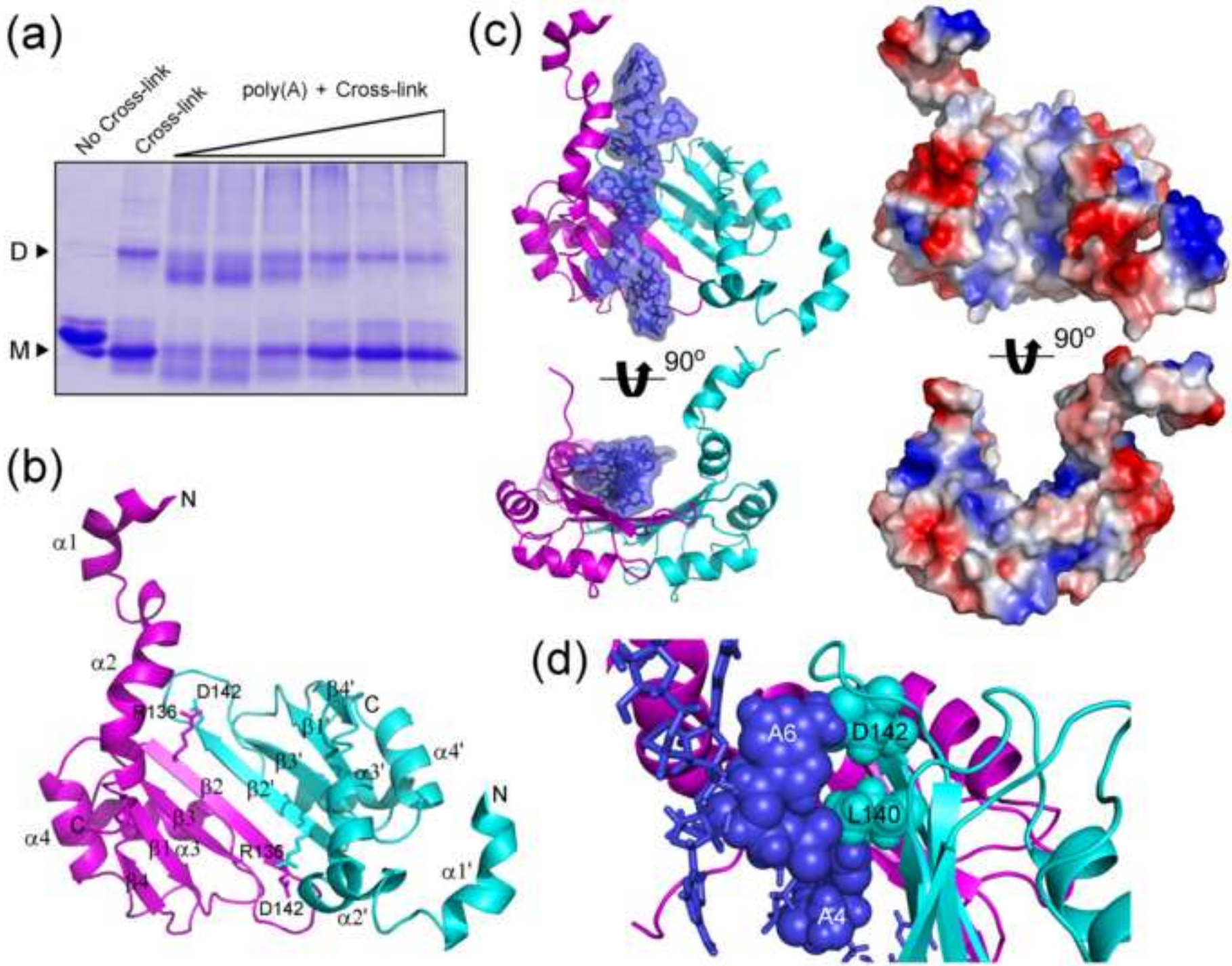
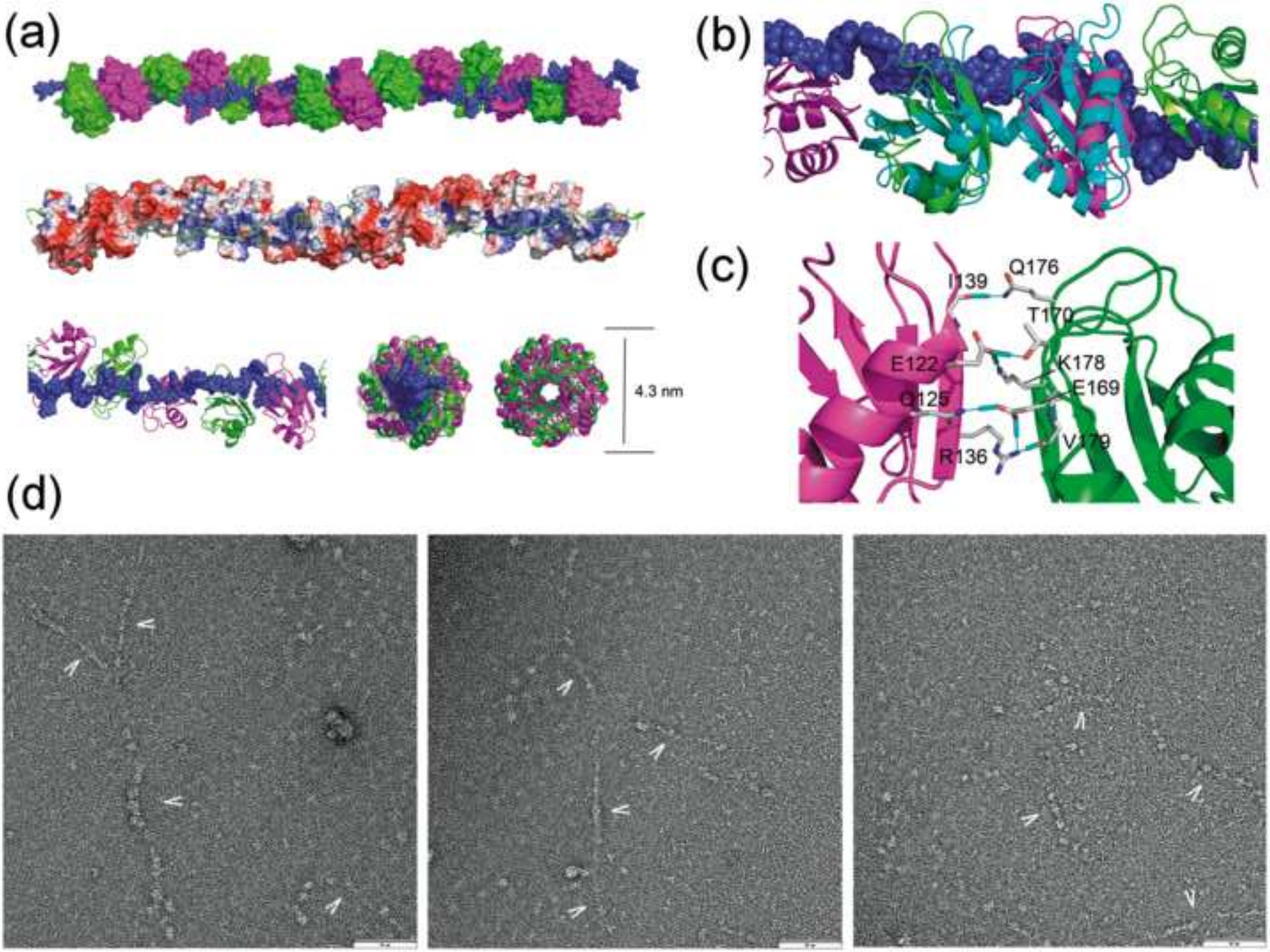


Figure 7

Click here to download high resolution image



Supplementary Fig. S1

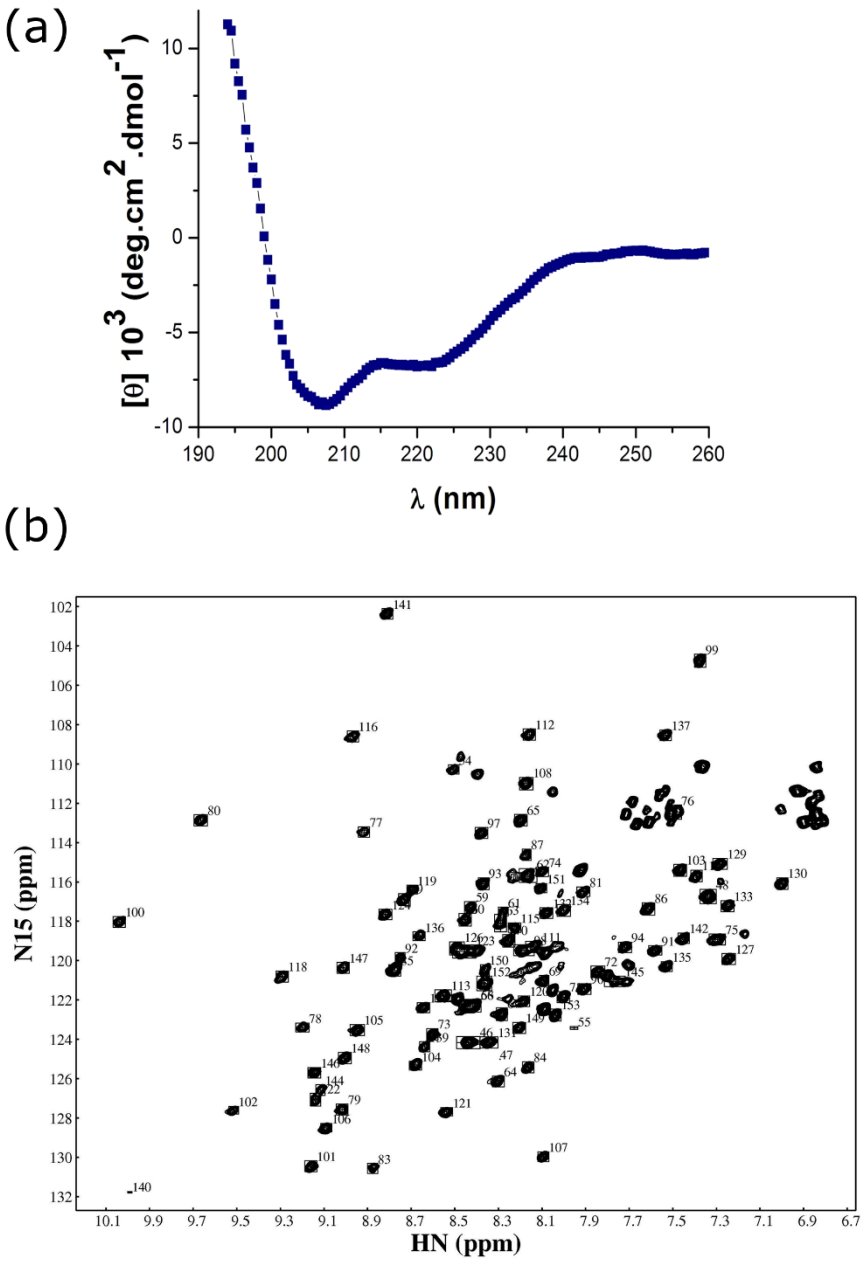


Fig. S1. CsΔPABPN1 is structured in solution. (a) Circular dichroism curve of CsΔPABPN1 showing two major negative peaks at 208 and 222 nm, indicating the presence of α-helices. (b) ^{15}NHSQC of CsΔPABPN1 showing excellent peak dispersion.

Supplementary Fig. S2

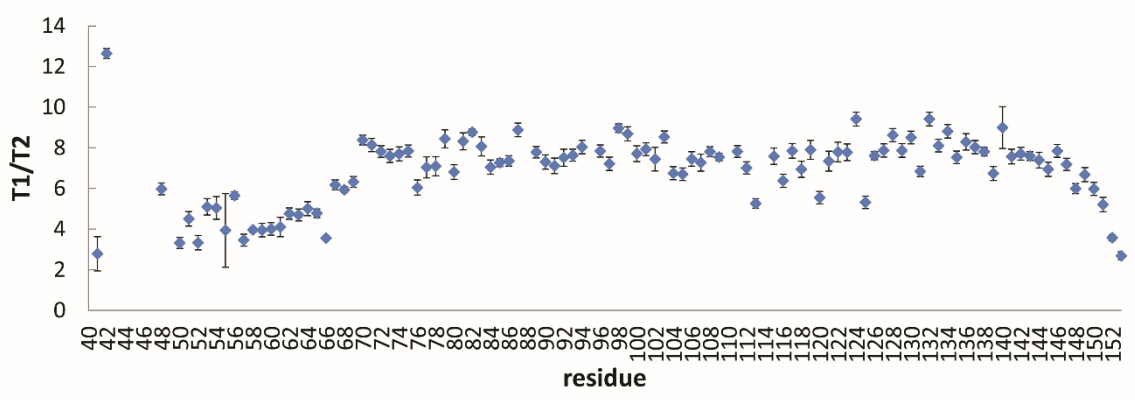


Fig. S2. Tau C value of CsΔPABPN1 determined by 2D ^{15}N-^{1}H-HSQC longitudinal relaxation time (T1) and the transverse relaxation time (T2) ratio (T1/T2) versus residue number. The calculated tau c value of 8.2 ± 0.5 ns is consistent with a protein of approximately 16 kDa, which indicates that CsΔPABPN1 is a monomer in the buffering conditions used in the NMR experiments.

Supplementary Fig. S3

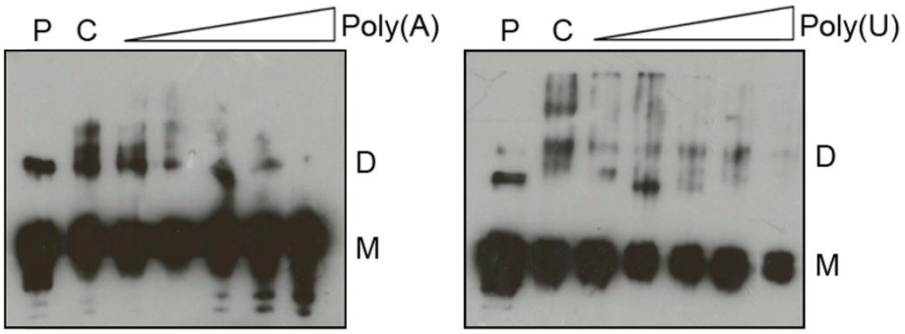
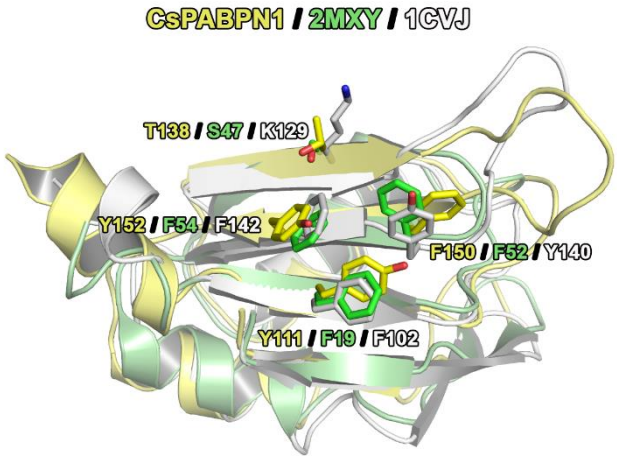


Fig. S3. Cross-linking analysis showing that full-length CsPABPN1 undergoes a dimer to monomer transition in the presence of poly(A), but not poly(U). The full-length CsPABPN1 protein was incubated in the absence (P) or presence of the crosslinker DSP (C), or in the presence of DSP and increased amounts of poly(A) or poly(U) RNA. CsPABPN1 forms monomers (M) and dimers (D) in the absence or presence of the crosslinker and shows a dimer to monomer conversion as the amount of poly(A) increases in the binding mix. The dimer/monomer proportion of CsPABPN1 does not seem to change substantially when the protein is titrated with increasing amounts of poly(U).

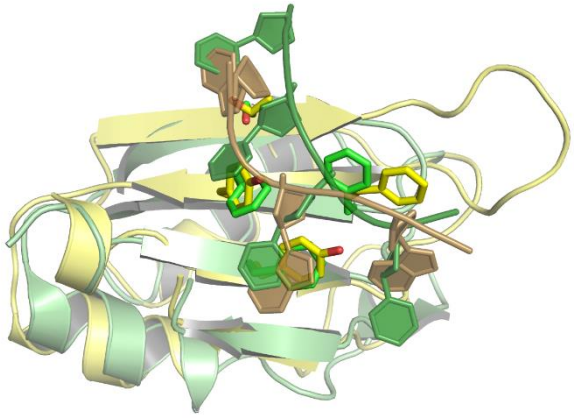
Supplementary Fig. S4

(a)

```
                RNP2                                                RNP1
CsPABPN1  SRSIYVGNVDYACTPEE-VQQHFQSCGTVNRVTILTDKFGQPKGFAYVEFVE 157
2MXY      NSRVFIGNLNTLVVKKSDVEAIFSKYGKIVGCSVH-------KGFAFVQYVN 59
1CVJ      VGNIFIKNLDKSIDNKA-LYDTFSAFGNILSCKVVCDENG-SKGYGFVHFET 147
             ::: *::     :  :   *.  *.:   .:        **:.:*.:
```



(b)



(c)

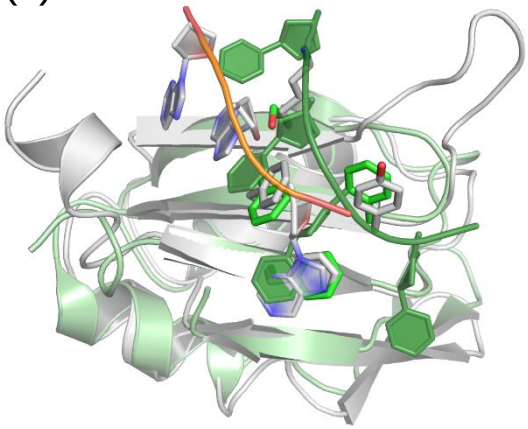


Fig. S4. The CsPABPN1, PABPC and HnRNP C RRMs show similar 3D structures and conserved RNA-binding modes. (a) Structural alignments of the CsPABPN1, human PABPC (1CVJ) and human poly(U)-binding protein HnRNP C (2MXY) RRMs. Despite the overall low amino acid sequence identity, the RRMs display conserved RNP sites and similar structural folding. Three aromatic residues of the RNP sites (blue) that are critical for the interaction with poly(A) in CsΔPABPN1 (yellow) and PABPC (white), and with poly(U) in HnRNP C (green), are structurally conserved. (b) In the structure of HnRNP C bound to poly(U), F19, S47, F52 and F54 make contacts with uridine bases (green), including stacking interactions with F19 and F54 [50]. Similarly, in our CsΔPABPN1-poly(A) modeled structure, Y111 and Y152, which respectively replaces F19 and F54 in HnRNP C, also make stacking interactions with adenines (sand). Moreover, T138, which corresponds to S47 in HnRNP C (panel 'a'), also interacts with an adenine base. (c) Structural alignment between PABPC and HnRNP C bound to poly(A) and poly(U), respectively. Although the RNA-binding mode of PAPBC and HnRNP C is similar, K129, which is in the equivalent position of S47, does not interact with the adenine base.